CSCI E-11

Class 4
February 17, 2026
Data and Network Risks

Harvard
Extension School
HARVARD DIVISION OF
CONTINUING EDUCATION

News story background: L3Harris is a major defense contractor.

L3Harris Technologies, Inc.

L3HARRIS™

Company type	Public
Traded as	NYSE: LHX S&P 500 component
Industry	Aerospace and defense
Predecessor	L3 Technologies Harris Corporation
Founded	June 29, 2019; 6 years ago
Headquarters	Melbourne, Florida, United States
Key people	Chris Kubasik (chair and CEO)
Revenue	▲ US$21.3 billion (2024)
Operating income	▲ US$1.92 billion (2024)
Net income	▲ US$1.50 billion (2024)
Total assets	▲ US$42.0 billion (2024)
Total equity	▲ US$19.5 billion (2024)
Number of employees	47,000 (2025)
Website	l3harris.com

Footnotes / references
[1][2]

https://en.wikipedia.org/wiki/L3Harris

TRENCHANT

An L3Harris Technologies Company

Supporting national security operations with end-point intelligence solutions, Trenchant is a world authority on cyber capabilities, operating in the fields of computer network operations and vulnerability research.

Our end-to-end solutions include:

- Vulnerability and exploit research
- APIs for intelligence operations
- Device and access capabilities
- Automated product testing
- First-class CNO products

https://www.l3harris.com/all-capabilities/trenchant

Feb. 11, 2026: Trenchant manager pled guilty to selling hacking tools.

TechCrunch

Lorenzo Franceschi-Bicchierai 10:41 AM PST · February 11, 2026

In October, Australian national Peter Williams, 39, pleaded guilty to selling eight hacking tools that he stole from his employer Trenchant, a division of the U.S. defense contractor L3Harris, which sells its surveillance-enabling tools to the U.S. government and its closest allies. Williams admitted to making more than $1.3 million in crypto from the sales between 2022 and 2025, per the Justice Department.

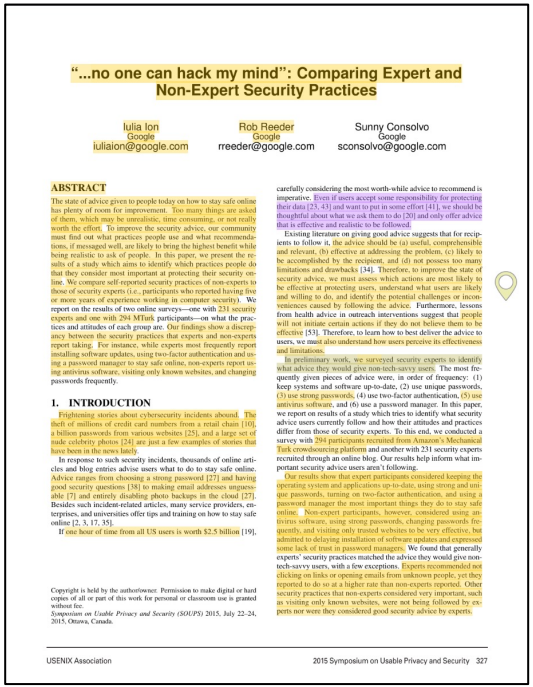
The Russian broker is likely Operation Zero, which offers up to $20 million for tools to hack into Android devices and iPhones. The company explicitly says it only sells to the Russian government and local organizations.

https://techcrunch.com/2026/02/11/doj-says-trenchant-boss-sold-exploits-to-russian-broker-capable-of-accessing-millions-of-computers-and-devices/

Class Updates/News

- Reading quiz late policy: no change.
  - But we will look at extra credit opportunities at the break.
- You have the do the labs in order!
- In the labs, you **must** run `e11 grade [lab#]` to get a grade
  - `e11 check [lab#]` tells you what you would get but it doesn't submit the grade.

Discussion: Readings



"...no one can hack my mind": Comparing Expert and Non-Expert Security Practices

Iulia Ion
Google
iuliaion@google.com

Rob Reeder
Google
rreeder@google.com

Sunny Consolvo
Google
sconsolvo@google.com

ABSTRACT

The state of advice given to people today on how to stay safe online has plenty of room for improvement. Too many things are asked of them, which may be unrealistic, time consuming, or not really worth the effort. To improve the security advice, our community must find out what practices people use and what recommendations, if messaged well, are likely to bring the highest benefit while being realistic to ask of people. In this paper, we present the results of a study which aims to identify which practices people do that they consider most important at protecting their security online. We compare self-reported security practices of non-experts to those of security experts (i.e., participants who reported having five or more years of experience working in computer security). We report on the results of two online surveys—one with 231 security experts and one with 294 MTurk participants—on what the practices and attitudes of each group are. Our findings show a discrepancy between the security practices that experts and non-experts report taking. For instance, while experts most frequently report installing software updates, using two-factor authentication and using a password manager to stay safe online, non-experts report using antivirus software, visiting only known websites, and changing passwords frequently.

1. INTRODUCTION

Frightening stories about cybersecurity incidents abound. The theft of millions of credit card numbers from a retail chain [10], a billion passwords from various websites [25], and a large set of nude celebrity photos [24] are just a few examples of stories that have been in the news lately.

In response to such security incidents, thousands of online articles and blog entries advise users what to do to stay safe online. Advice ranges from choosing a strong password [27] and having good security questions [38] to making email addresses unguessable [7] and entirely disabling photo backups in the cloud [27]. Besides such incident-related articles, many service providers, enterprises, and universities offer tips and training on how to stay safe online [2, 3, 17, 35].

If one hour of time from all US users is worth $2.5 billion [19], carefully considering the most worth-while advice to recommend is imperative. Even if users accept some responsibility for protecting their data [23, 43] and want to put in some effort [41], we should be thoughtful about what we ask them to do [20] and only offer advice that is effective and realistic to be followed.

Existing literature on giving good advice suggests that for recipients to follow it, the advice should be (a) useful, comprehensible and relevant, (b) effective at addressing the problem, (c) likely to be accomplished by the recipient, and (d) not possess too many limitations and drawbacks [34]. Therefore, to improve the state of security advice, we must assess which actions are most likely to be effective at protecting users, understand what users are likely and willing to do, and identify the potential challenges or inconveniences caused by following the advice. Furthermore, lessons from health advice in outreach interventions suggest that people will not initiate certain actions if they do not believe them to be effective [53]. Therefore, to learn how to best deliver the advice to users, we must also understand how users perceive its effectiveness and limitations.

In preliminary work, we surveyed security experts to identify what advice they would give non-tech-savvy users. The most frequently given pieces of advice were, in order of frequency: (1) keep systems and software up-to-date, (2) use unique passwords, (3) use strong passwords, (4) use two-factor authentication, (5) use antivirus software, and (6) use a password manager. In this paper, we report on results of a study which tries to identify what security advice users currently follow and how their attitudes and practices differ from those of security experts. To this end, we conducted a survey with 294 participants recruited from Amazon's Mechanical Turk crowdsourcing platform and another with 231 security experts recruited through an online blog. Our results help inform what important security advice users aren't following.

Our results show that expert participants considered keeping the operating system and applications up-to-date, using strong and unique passwords, turning on two-factor authentication, and using a password manager the most important things they do to stay safe online. Non-expert participants, however, considered using antivirus software, using strong passwords, changing passwords frequently, and visiting only trusted websites to be very effective, but admitted to delaying installation of software updates and expressed some lack of trust in password managers. We found that generally experts' security practices matched the advice they would give non-tech-savvy users, with a few exceptions. Experts recommended not clicking on links or opening emails from unknown people, yet they reported to do so at a higher rate than non-experts reported. Other security practices that non-experts considered very important, such as visiting only known websites, were not being followed by experts nor were they considered good security advice by experts.

Symposium on Usable Privacy and Security (SOUPS) 2015, July 22–24, 2015, Ottawa, Canada.

Replication: No One Can Hack My Mind
Revisiting a Study on Expert and Non-Expert Security Practices and Advice

Karoline Busse
University of Bonn
busse@cs.uni-bonn.de

Julia Schäfer
Univeristy of Bonn
sfjuscha@gmail.com

Matthew Smith
University of Bonn / Fraunhofer FKIE
smith@cs.uni-bonn.de

Abstract

A 2015 study by Julia Ion, Rob Reeder, and Sunny Consolvo examined the self-reported security behavior of security experts and non-experts. They also analyzed what kind of security advice experts gave to non-experts and how realistic and effective they think typical advice is.

Now, roughly four years later, we aimed to replicate and extend this study with a similar set of non-experts and a different set of experts. For the non-experts, we recruited 288 MTurk participants, just as Ion et al. did. We also recruited 75 mostly European security experts, in contrast to the mostly US sample from Ion et al. Our findings show that despite the different samples and the four years that have passed, the most common pieces of expert advice are mostly unchanged, with one notable exception. In addition, we did see a fair amount of fluctuation in the long tail of advice. Non-expert self-reported behavior, however, is unchanged, meaning that the gap between experts and non-experts seen in Ion et al.'s work is still just as prominent in our study. To extend the work, we also conducted an A/B study to get a better understanding of one of the key questions concerning experts' recommendations, and we identified types of advice where research by the usable security community is most sorely needed.

1 Introduction

Whenever the media picks up on the latest data breach, various sources seize the opportunity to give advice such as "Do not use the same passwords for all systems" [9] or "Antivirus software is crucial to protecting your computer." [23] Under this barrage of different advice, selecting and following "good" advice is a difficult task for users [10]. Factors such as socioeconomic status, consumer habits, or conveniences also play a role in the decision-making process [24, 26]. Even when advice is regarded as "good" by a user, it is not necessarily a given that they know how to apply it in their own individual context. We must not overlook the limits of users' capability taking into account the complexity of any advice we give [5].

In 2015, Ion, Reeder, and Consolvo explored the opinions and beliefs of expert and non-expert users in a survey study and found that users neglect three vital security practices that experts strongly advise: installing software updates, using two-factor authentication, and using a password manager. On the other side, non-experts regarded antivirus software as a very important security practice, unlike the experts, who were not convinced by it. Almost four years have passed since that study, which is a long period of time in terms of technological innovation and security practices. Security and privacy continue to gain more widespread recognition, so we were interested to see what, if anything, had changed with respect to expert advice and non-expert self-reported behavior.

We thus conducted two online surveys, one for experts and one for non-experts, and compared the results to the previous study by Ion et al. Many of the past security topics and advice covered in the original work are still relevant today. We also discovered that some of the topics relevant to users in the past have been replaced by newer topics, for example, the spread of blocking extensions for web browsers, which are able to manage cookies. Where in the past, users were concerned with regularly deleting cookies, they now rely on blocking extensions.

Apart from seeing if and how our sample differed from the original, we wanted to explore a methodological issue in the original study. One of the central parts of the original study concerned how effective *and* realistic particular types of advice are. This information from experts was gathered using compound questions, and the advice was ranked and compared on that basis. Compound questions can be problematic because it is not clear how participants combine the separate

USENIX Symposium on Usable Privacy and Security (SOUPS) 2019.
August 11–13, 2019, Santa Clara, CA, USA.

"...No one Can Hack My Mind": Comparing Expert and Non-Expert Security Practices" (SOUPS 2015)
https://www.usenix.org/conference/soups2015/proceedings/presentation/ion

(Optional) Replication: No One Can Hack My Mind Revisiting a Study on Expert and Non-Expert Security Practices and Advice, Karoline Busse and Julia Schäfer
https://www.usenix.org/conference/soups2019/presentation/busse

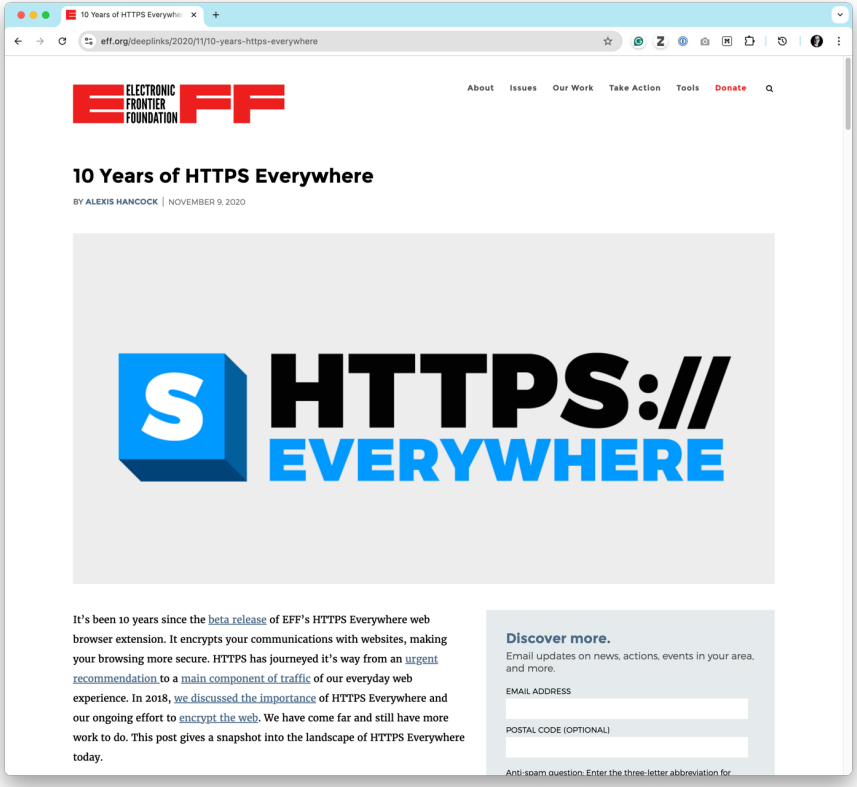
HTTPS Everywhere

2010 - EFF releases “https everywhere” browser extension

2012 - Mozilla, EFF and UMich start Let’s Encrypt

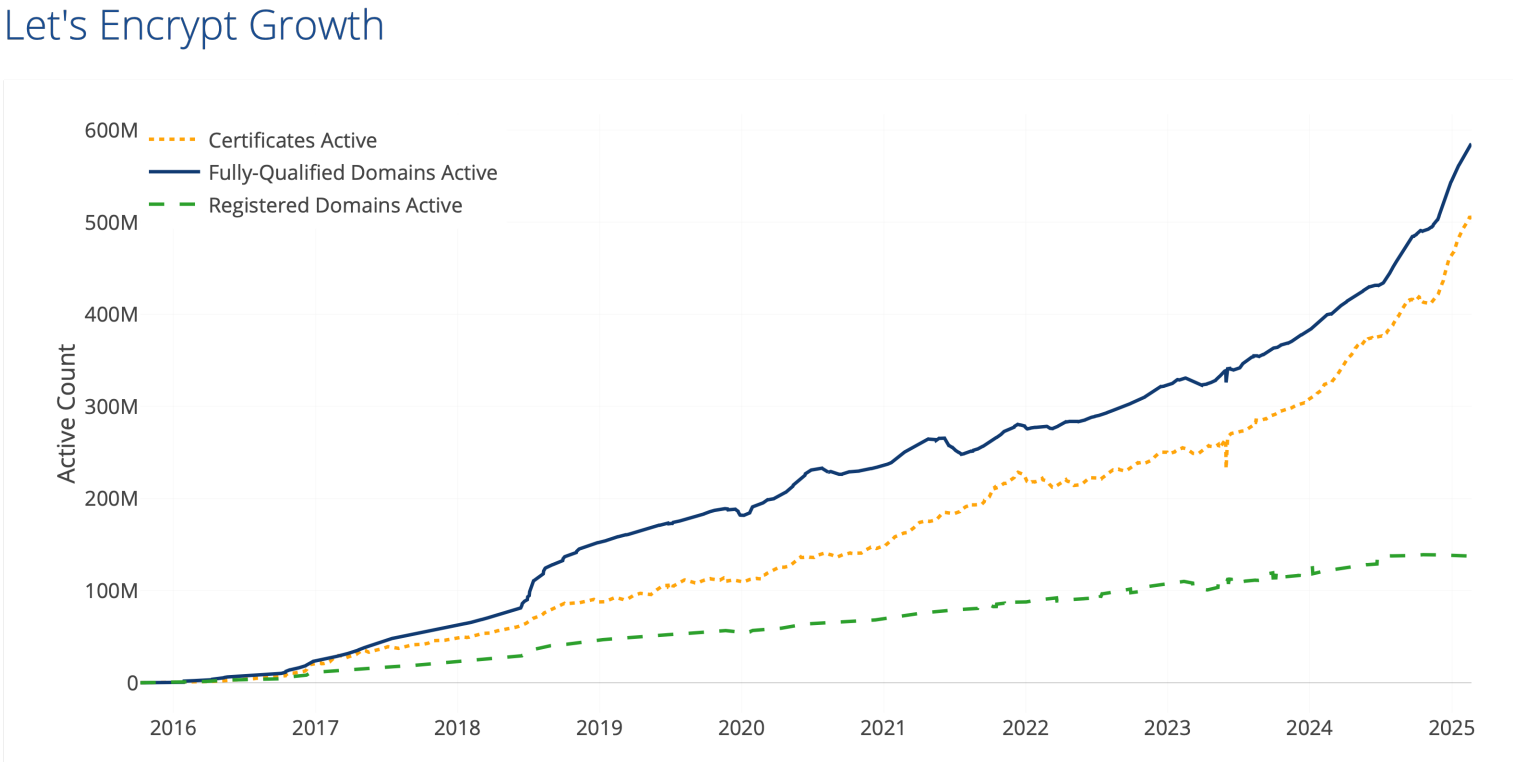
2015 - Let’s Encrypt issues first certificate

2020 - Let’s Encrypt issued a billion certificates



Let’s Encrypt has had a huge impact!

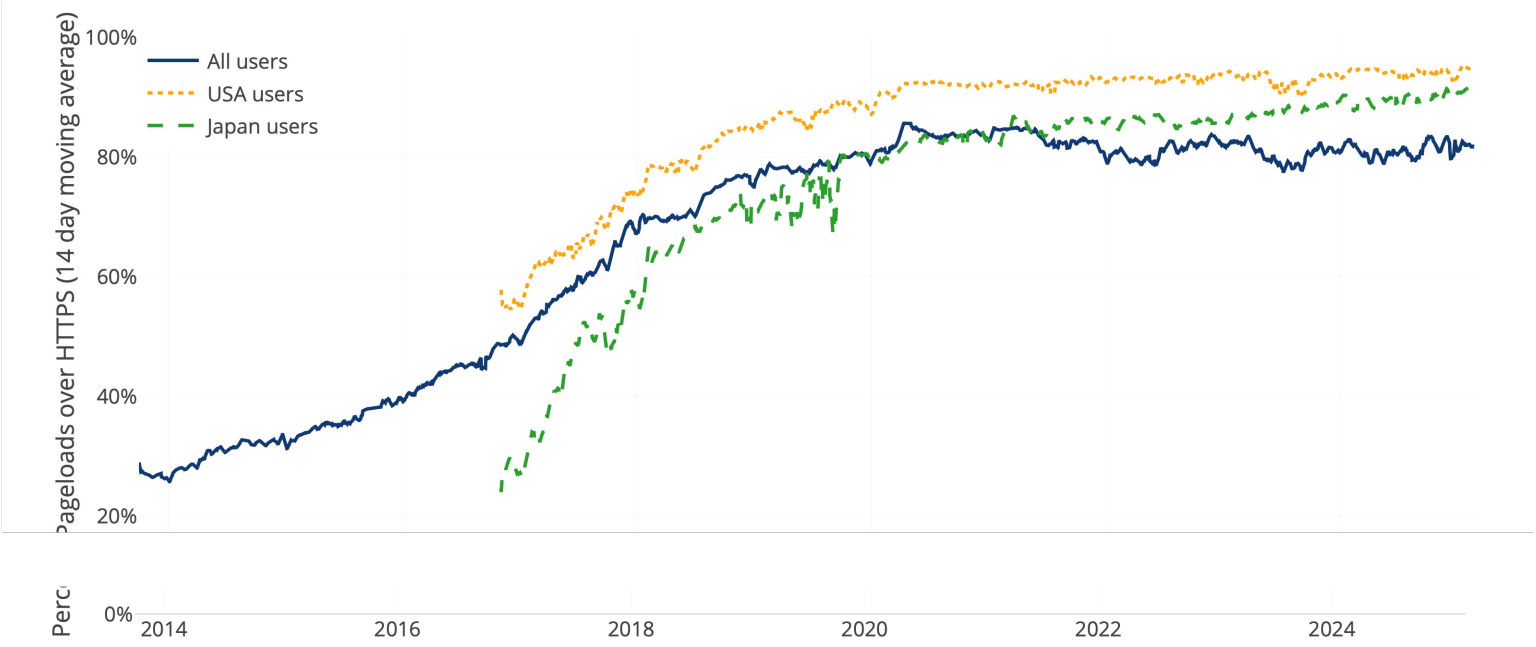
https://letsencrypt.org/stats/



HTTPS market penetration (of users using Firefox).

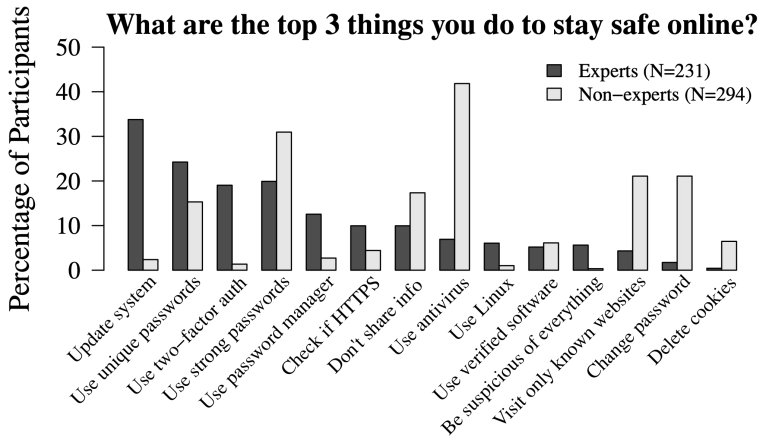
Percentage of Web Pages Loaded by Firefox Using HTTPS

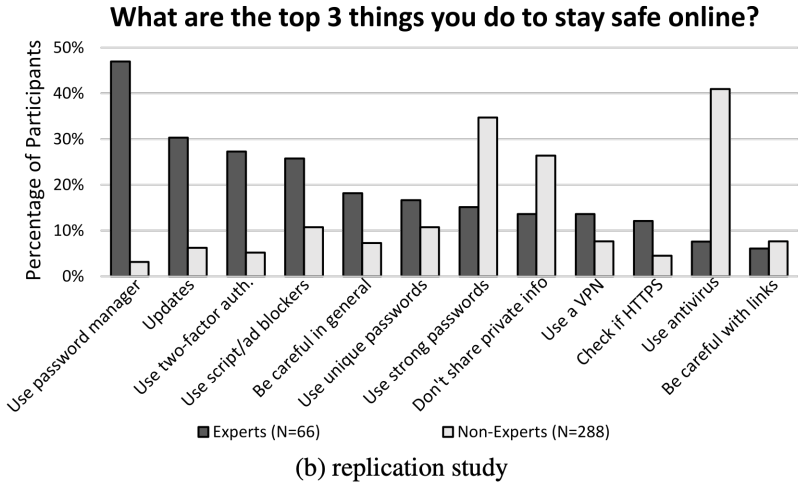
(14-day moving average, source: Firefox Telemetry)



Replication Study

Comparison of original and replication study for “top 3.”





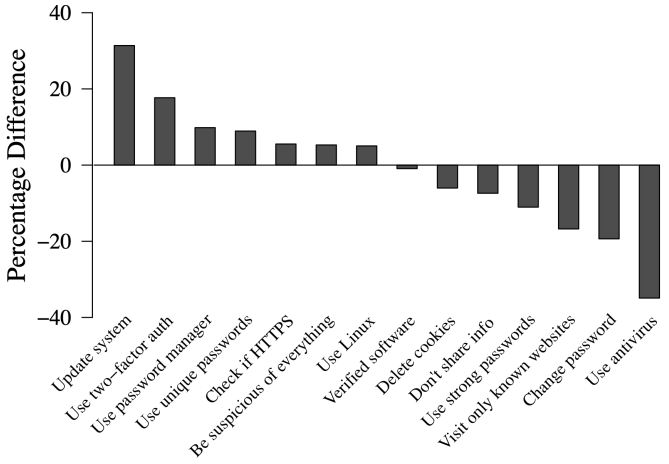
(b) replication study

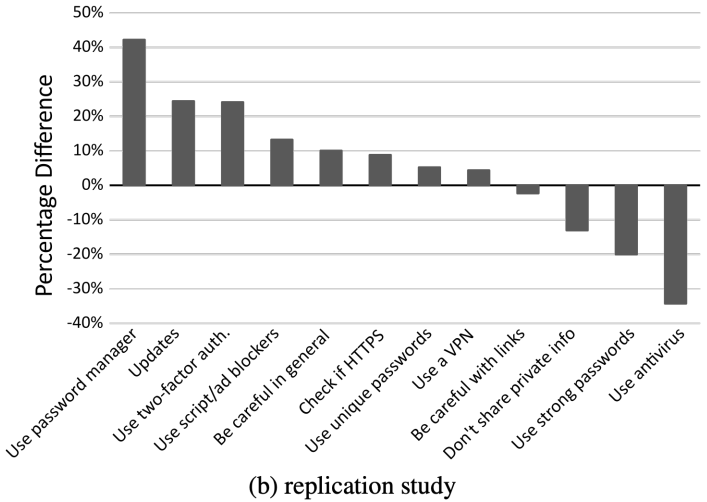
Figure 1: Security measures mentioned by at least 5% of each group. While most experts said they keep their system updated and use two-factor authentication to stay safe online, non-experts emphasized using antivirus software and using strong passwords.

2015

2019

Comparison of original and replication study for differences.





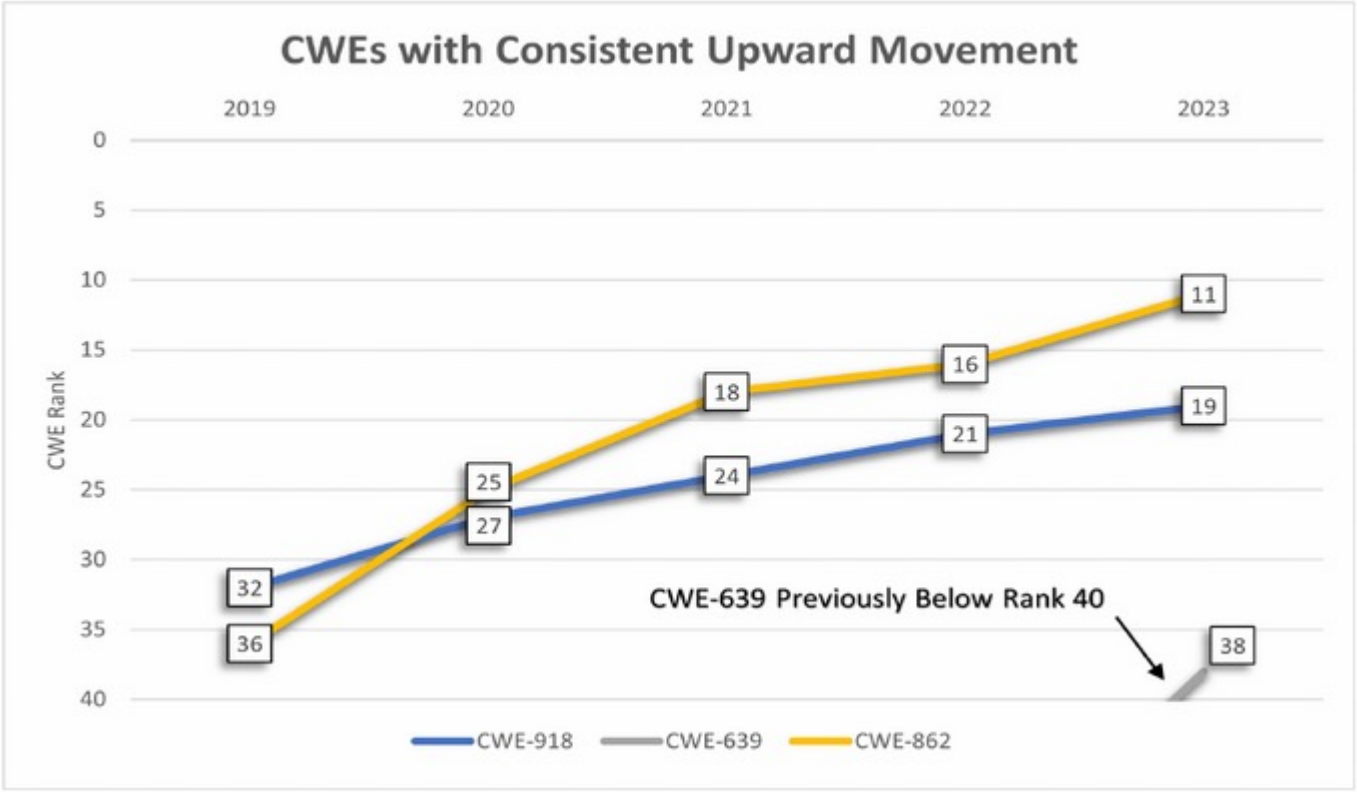
(b) replication study

Figure 2: Percentage difference of experts and non-experts mentioning these security practices when asked what are the top three things they do to stay safe online. Security measures with a positive percentage difference were mentioned more by experts than non-experts; those with a negative percentage difference were mentioned more by non-experts.

Replication study identified “effective but unrealistic practices” for improvement.

- Password security
- 2FA
- Securely handling links and attachments from emails
- Centralizing application updates

CWE Top 25 Most Dangerous Software Weaknesses



2025

CWE-862: 4
CWE-918: 22
CWE-639: 24

CWE-862

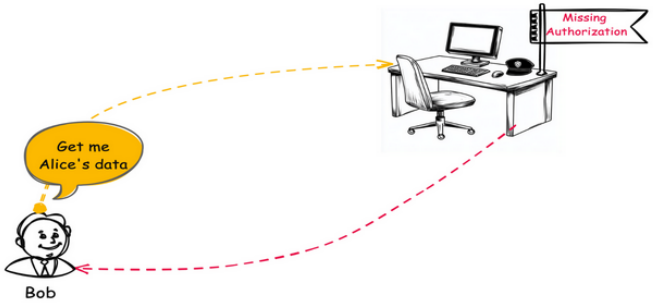
CWE-862: Missing Authorization

Weakness ID: 862
Vulnerability Mapping: ALLOWED (with careful review of mapping notes)
Abstraction: Class

View customized information: Conceptual | Operational | Mapping Friendly | **Complete** | Custom

Description

The product does not perform an authorization check when an actor attempts to access a resource or perform an action.



CWE-918

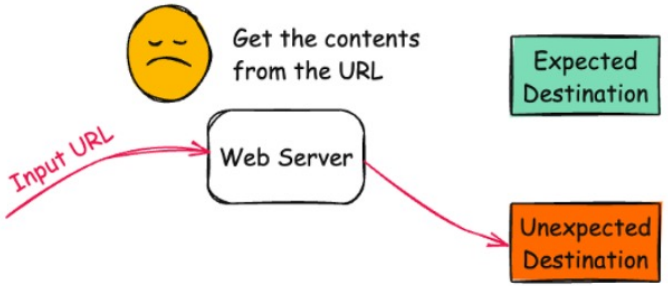
CWE-918: Server-Side Request Forgery (SSRF)

Weakness ID: 918
Vulnerability Mapping: ALLOWED
Abstraction: Base

View customized information: Conceptual | Operational | Mapping Friendly | **Complete** | Custom

Description

The web server receives a URL or similar request from an upstream component and retrieves the contents of this URL, but it does not sufficiently ensure that the request is being sent to the expected destination.



CWE-639

CWE-639: Authorization Bypass Through User-Controlled Key

Weakness ID: 639
Vulnerability Mapping: ALLOWED
Abstraction: Base

View customized information: Conceptual | Operational | Mapping Friendly | **Complete** | Custom

Description

The system's authorization functionality does not prevent one user from gaining access to another user's data or record by modifying the key value identifying the data.

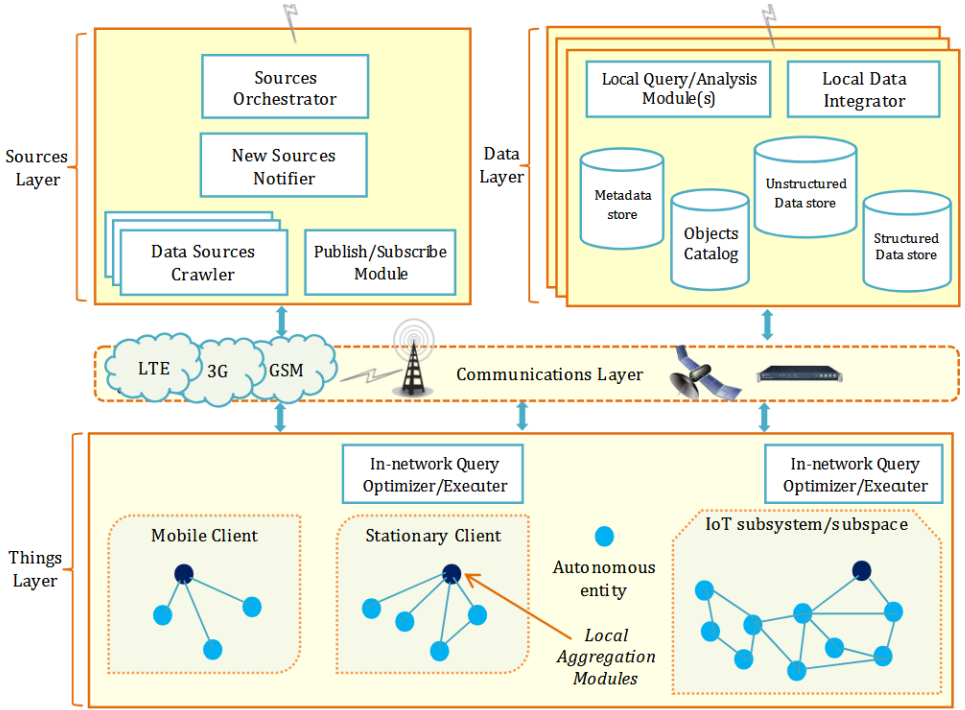
Extended Description

Retrieval of a user record occurs in the system based on some key value that is under user control. The key would typically identify a user-related record stored in the system and would be used to lookup that record for presentation to the user. It is likely that an attacker would have to be an authenticated user in the system. However, the authorization process would not properly check the data access operation to ensure that the authenticated user performing the operation has sufficient entitlements to perform the requested data access, hence bypassing any other authorization checks present in the system.

For example, attackers can look at places where user specific data is retrieved (e.g. search screens) and determine whether the key for the item being looked up is controllable externally. The key may be a hidden field in the HTML form field, might be passed as a URL parameter or as an unencrypted cookie variable, then in each of these cases it will be possible to tamper with the key value.

One manifestation of this weakness is when a system uses sequential or otherwise easily-guessable session IDs that would allow one user to easily switch to another user's session and read/modify their data.

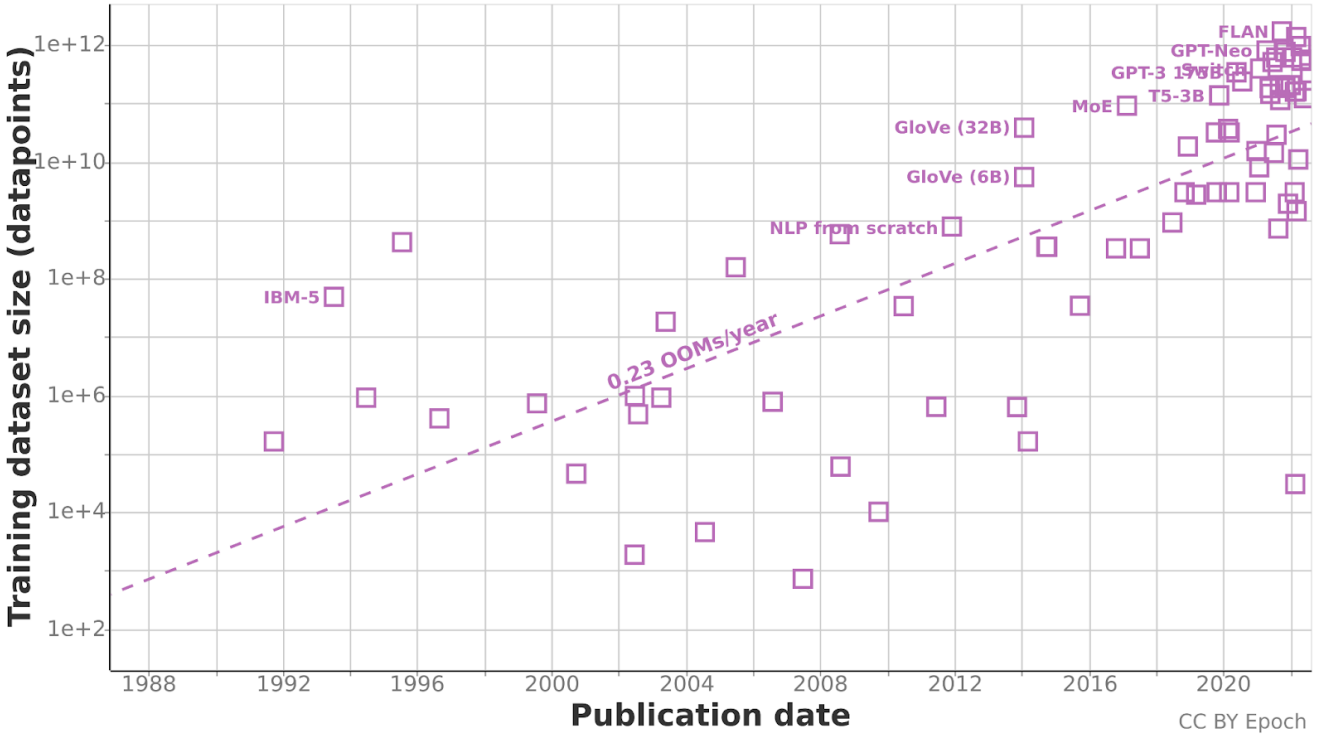
The Internet of Things needs to move a lot of data through the network.



Data Management for the Internet of Things: Design Primitives and Solution

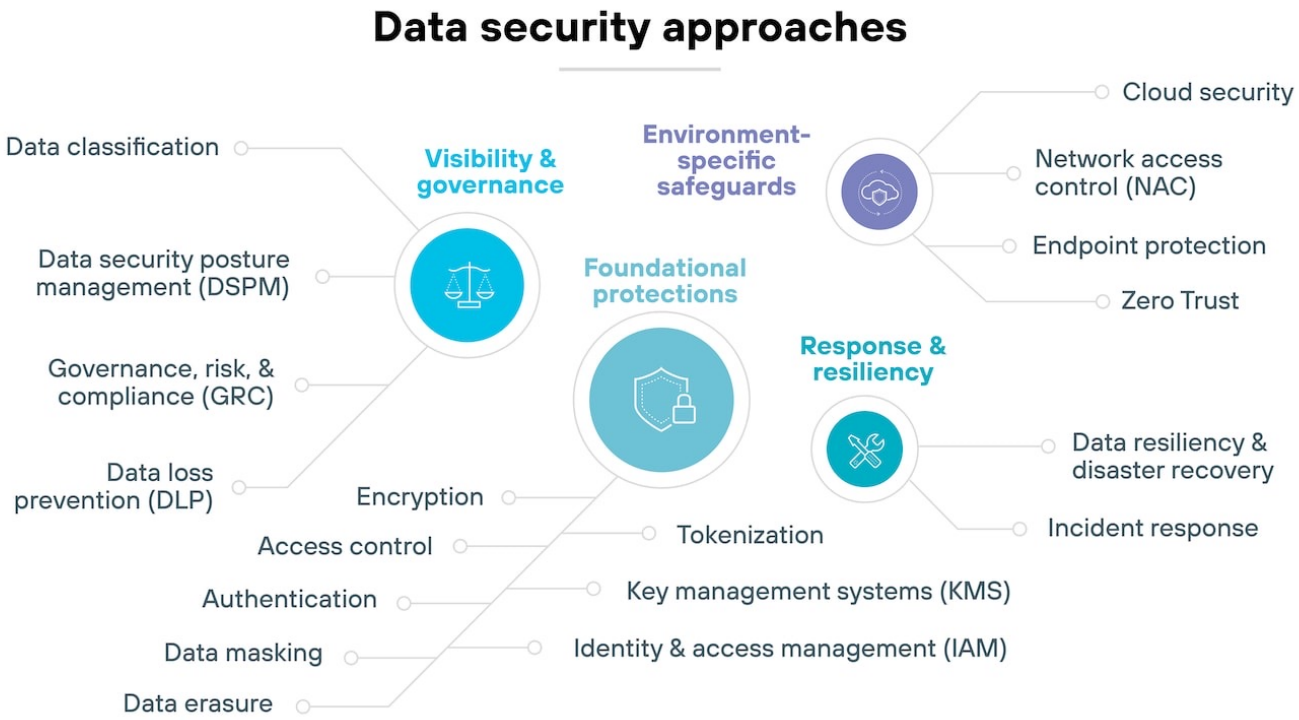
Mervat Abu-Elkheir [1,*], Mohammad Hayajneh [2] and Najah Abu Ali [2]

Modern approaches to AI consume huge amounts of data.



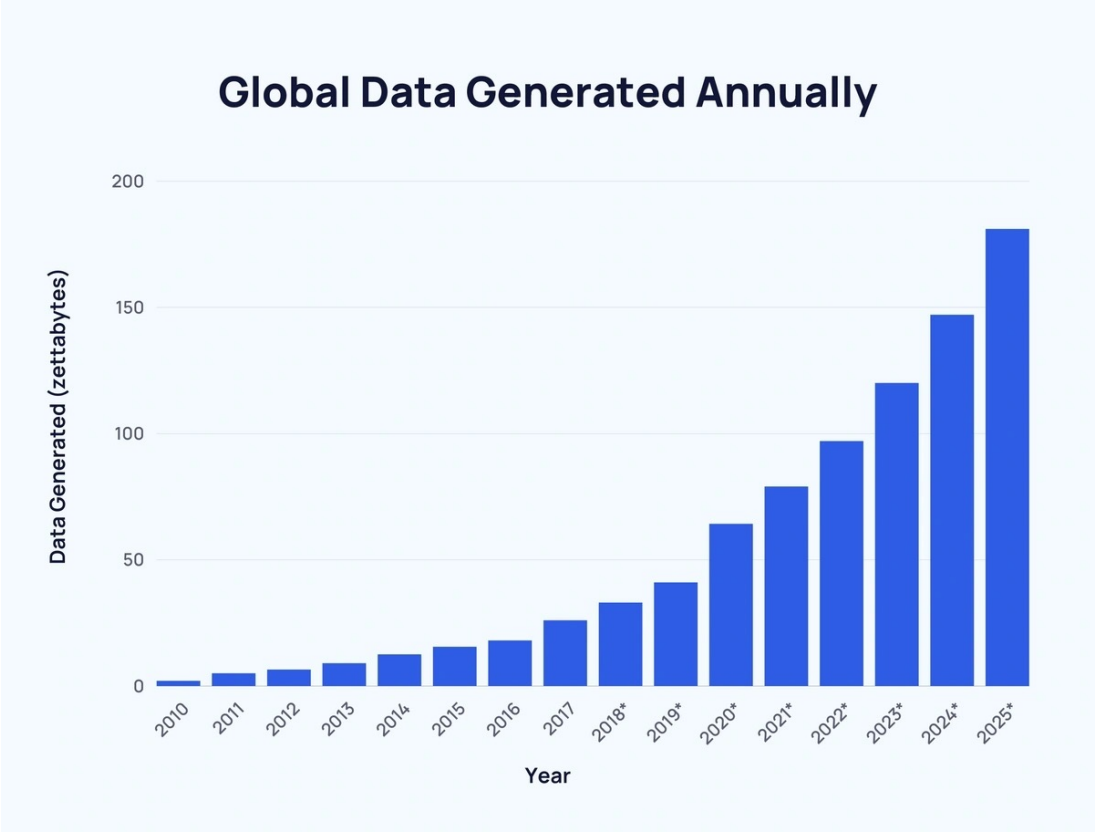
https://epoch.ai/publications/trends-in-training-dataset-sizes

Protecting confidentiality & integrity of data in transit and at rest is core to security.

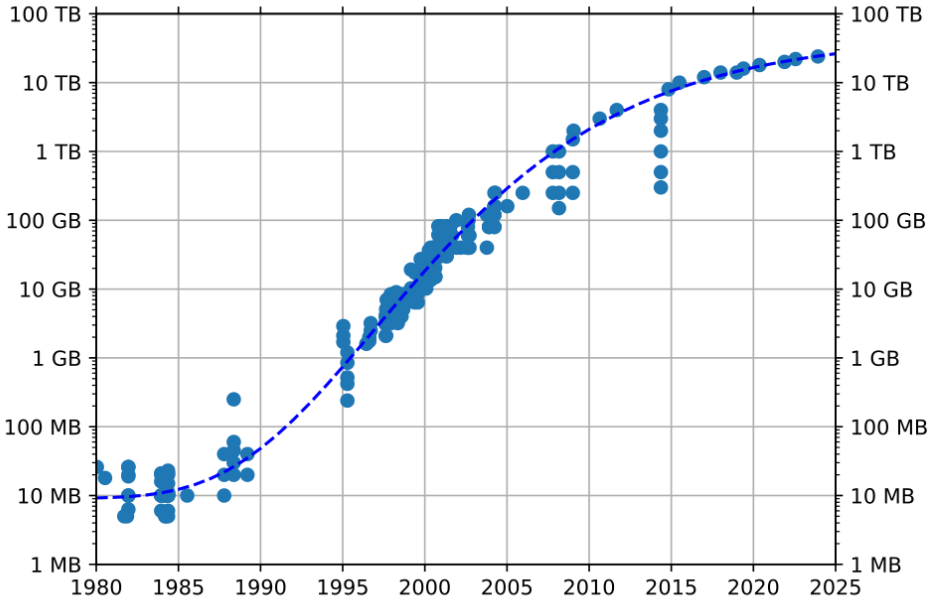


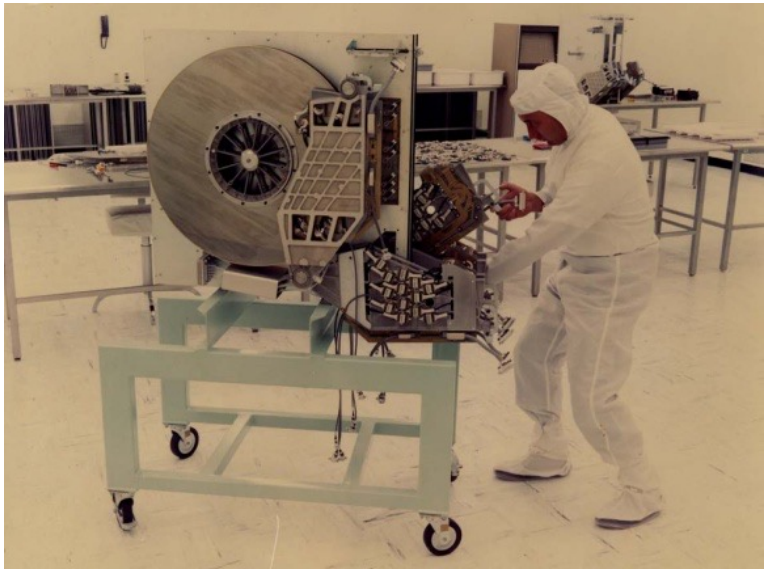
https://www.paloaltonetworks.com/cyberpedia/what-is-data-security

There is an unfathomable amount of data created each day, increasing rapidly.



At the same time, the capacity to store data has increased exponentially.





Retail hard drive capacity over time, 1980-2024

https://commons.wikimedia.org/wiki/File:Hard_drive_capacity_over_time.svg

Structured data stores records of data with defined fields

CUSTOMER

CUSTOMER_ID	LAST_NAME	FIRST_NAME	STREET	CITY	ZIP_CODE	COUNTRY
10302	Boucher	Leo	54, rue Royale	Nantes	44000	France
11244	Smith	Laurent	8489 Strong St	Las Vegas	83030	USA
11405	Han	James	636 St Kilda Road	Sydney	3004	Australia
11993	Mueller	Tomas	Berliner Weg 15	Tamm	71732	Germany
12111	Carter	Nataly	5 Tomahawk	Los Angeles	90006	USA
14121	Cortez	Nola	Av. Grande, 86	Madrid	28034	Spain
14400	Brown	Frank	165 S 7th St	Chester	33134	USA
14578	Wilson	Sarah	Seestreet #6101	Emory	1734	USA
14622	Jones	John	71 San Diego Ave	Arlington	69004	USA

A very brief introduction to databases.

Q: Do you know what this is?

Relational databases store information in tables.

Row (or Record)

Casey Arias	tharris@nielsen.net	S36055	3.37	1502
Cameron Kidd	brandonbaker@yahoo.com	S95137	3.45	1479
Elizabeth Taylor	richard93@hotmail.com	S71424	2.38	577
Ashley Smith	crichardson@miranda-lewis.com	S86340	3.02	945
Nancy Case	christopherdavis@gmail.com	S77339	3.59	467

Column (or Field)

The fields of a table are defined in a “Schema”

Casey Arias	tharris@nielsen.net	S36055	3.37	1502
Cameron Kidd	brandonbaker@yahoo.com	S95137	3.45	1479
Elizabeth Taylor	richard93@hotmail.com	S71424	2.38	577
Ashley Smith	crichardson@miranda-lewis.com	S86340	3.02	945
Nancy Case	christopherdavis@gmail.com	S77339	3.59	467

```
CREATE TABLE students (name, email, student_id, gpa, sat_score)
```

* Note: I left out the data types for those fields for simplicity. This statement won’t work as is!

SQL is a language for searching and updating relational databases.

```
SELECT * from students where student_id = "S71424"
```

Casey Arias	tharris@nielsen.net	S36055	3.37	1502
Cameron Kidd	brandonbaker@yahoo.com	S95137	3.45	1479
Elizabeth Taylor	richard93@hotmail.com	S71424	2.38	577
Ashley Smith	crichardson@miranda-lewis.com	S86340	3.02	945
Nancy Case	christopherdavis@gmail.com	S77339	3.59	467

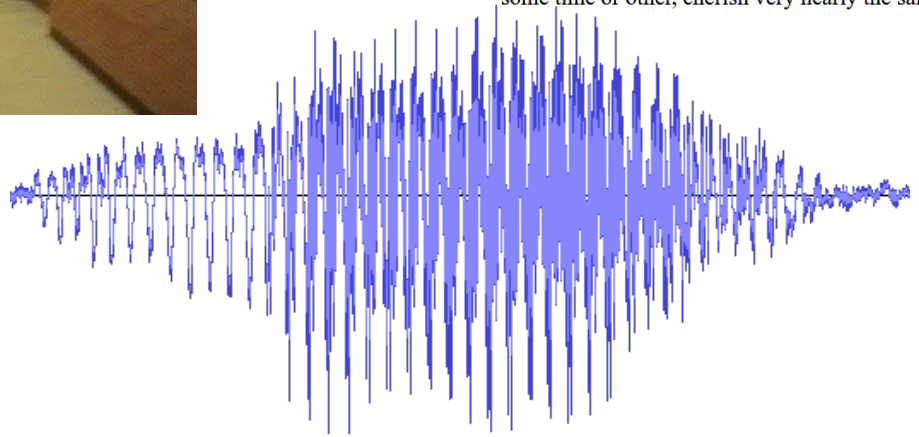
Semi-structured data has some degree of overall structure but not every element is fully defined

```
–<breakfast_menu>
    <script/>
   –<food>
        <name>Belgian Waffles</name>
        <price>$5.95</price>
      –<description>
            Two of our famous Belgian Waffles with plenty of real maple syrup
        </description>
        <calories>650</calories>
    </food>
   –<food>
        <name>Strawberry Belgian Waffles</name>
        <price>$7.95</price>
      –<description>
            Light Belgian waffles covered with strawberries and whipped cream
        </description>
        <calories>900</calories>
    </food>
```

Unstructured data contains raw unlabeled information

CHAPTER 1. Loomings.

Call me Ishmael. Some years ago—never mind how long precisely—having little or no money in my purse, and nothing particular to interest me on shore, I thought I would sail about a little and see the watery part of the world. It is a way I have of driving off the spleen and regulating the circulation. Whenever I find myself growing grim about the mouth; whenever it is a damp, drizzly November in my soul; whenever I find myself involuntarily pausing before coffin warehouses, and bringing up the rear of every funeral I meet; and especially whenever my hypos get such an upper hand of me, that it requires a strong moral principle to prevent me from deliberately stepping into the street, and methodically knocking people's hats off—then, I account it high time to get to sea as soon as I can. This is my substitute for pistol and ball. With a philosophical flourish Cato throws himself upon his sword; I quietly take to the ship. There is nothing surprising in this. If they but knew it, almost all men in their degree, some time or other, cherish very nearly the same feelings towards the ocean with me.



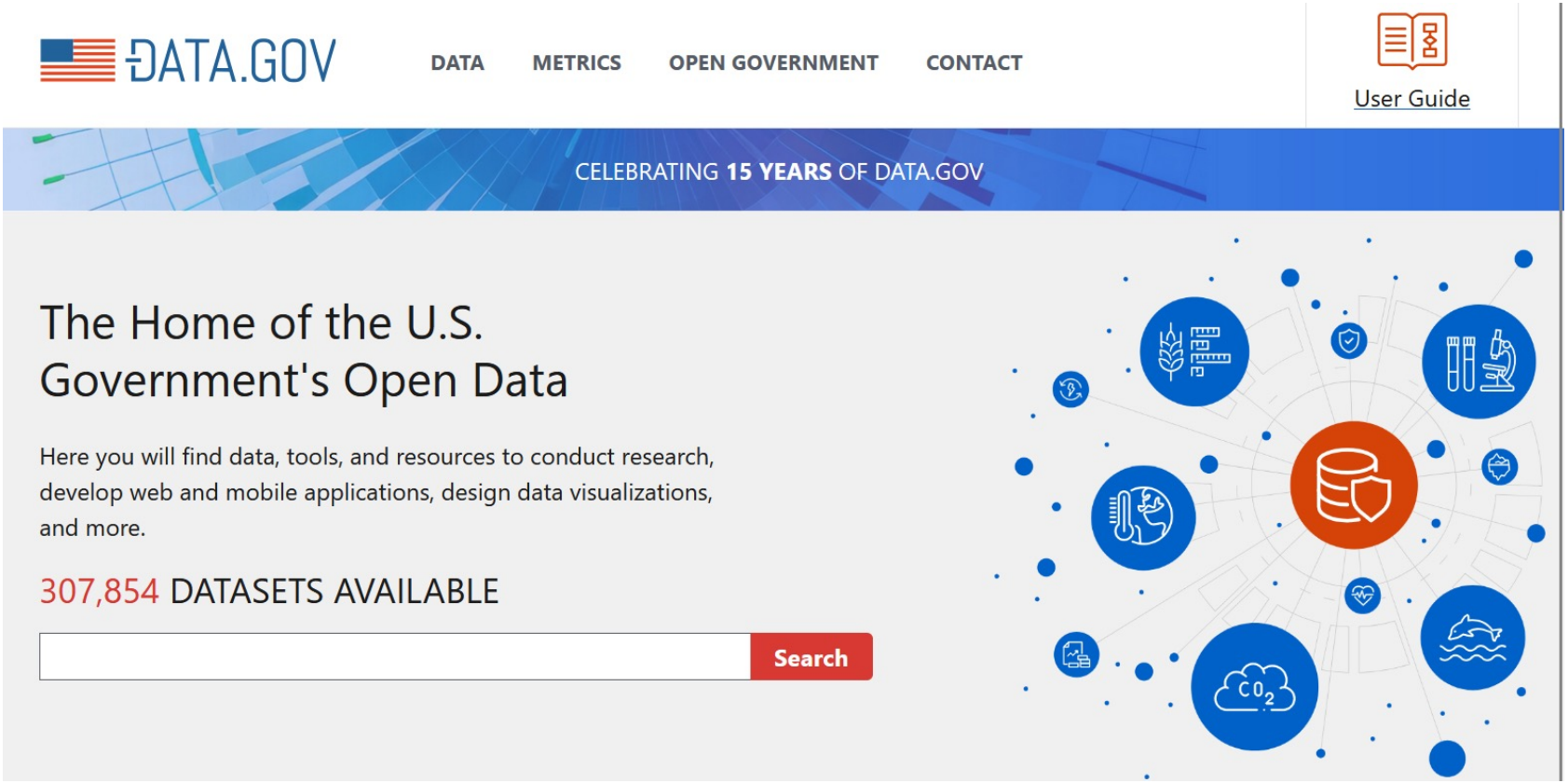
Metadata is data about data.

Author	Crowley, John, 1942-
Series	Aegypt novels **volume 1** Bantam spectra book
Published	Toronto ; New York : Bantam Books, 1987.
ISBN	0553051946
Physical Desc	390 pages ; 24 cm.
Status	On Shelf **Newton - Adult** FICTION CROWLEY, J. 1 available Where is it?

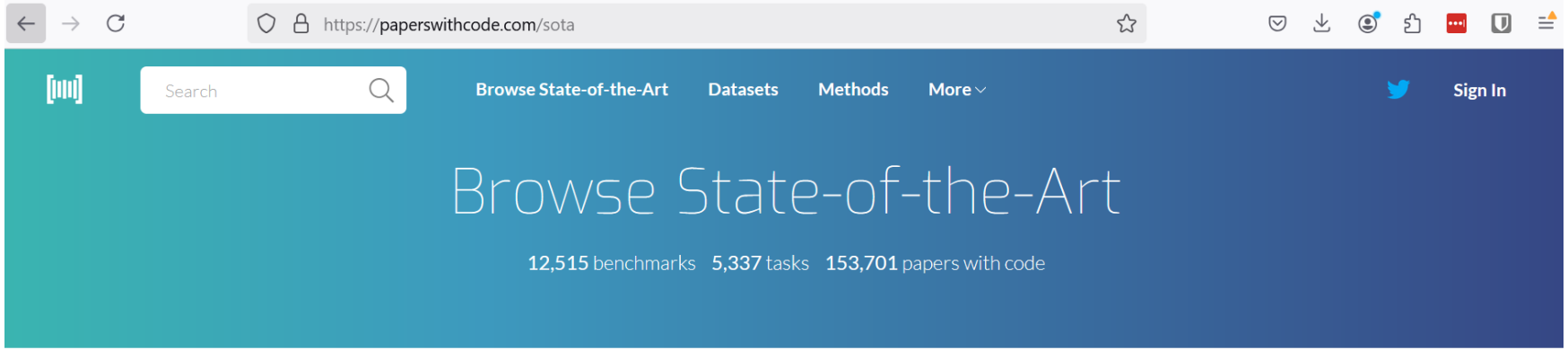
Questions to Ask About Data

- Where does it come from?
- How was it generated/created?
- Who owns it?
- How is it acquired?
- Does it contain sensitive data?
- How is it stored and retrieved efficiently?
- Is it labeled? Is it accurate? How do we know?

Free and open availability of data supports research and reproducibility.



paperswithcode.com



Computer Vision

Semantic Segmentation

345 benchmarks

6215 papers with code

Object Detection

388 benchmarks

4427 papers with code

Image Classification

489 benchmarks

4401 papers with code

Representation Learning

16 benchmarks

4369 papers with code

Contrastive Learning

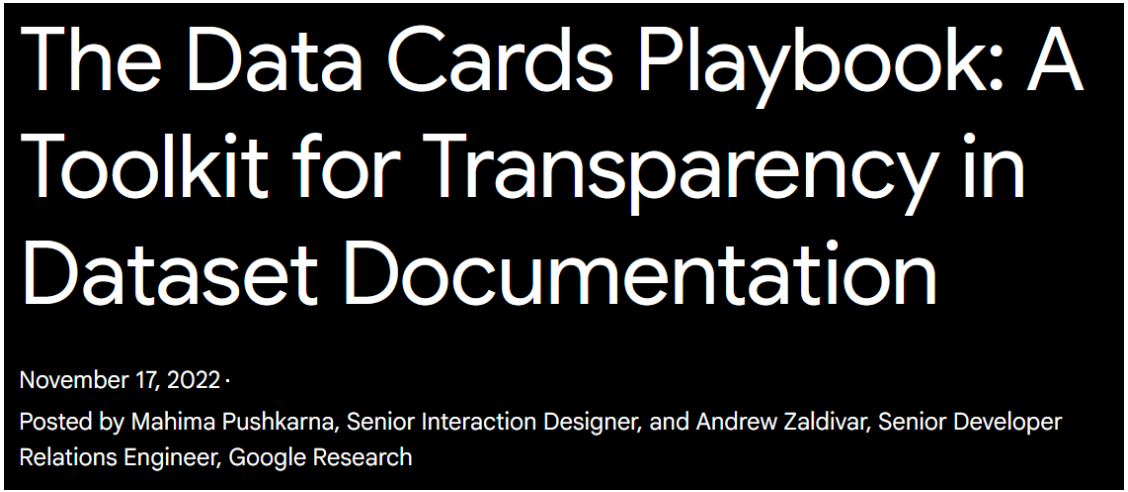
4 benchmarks

2792 papers with code

Data Cards

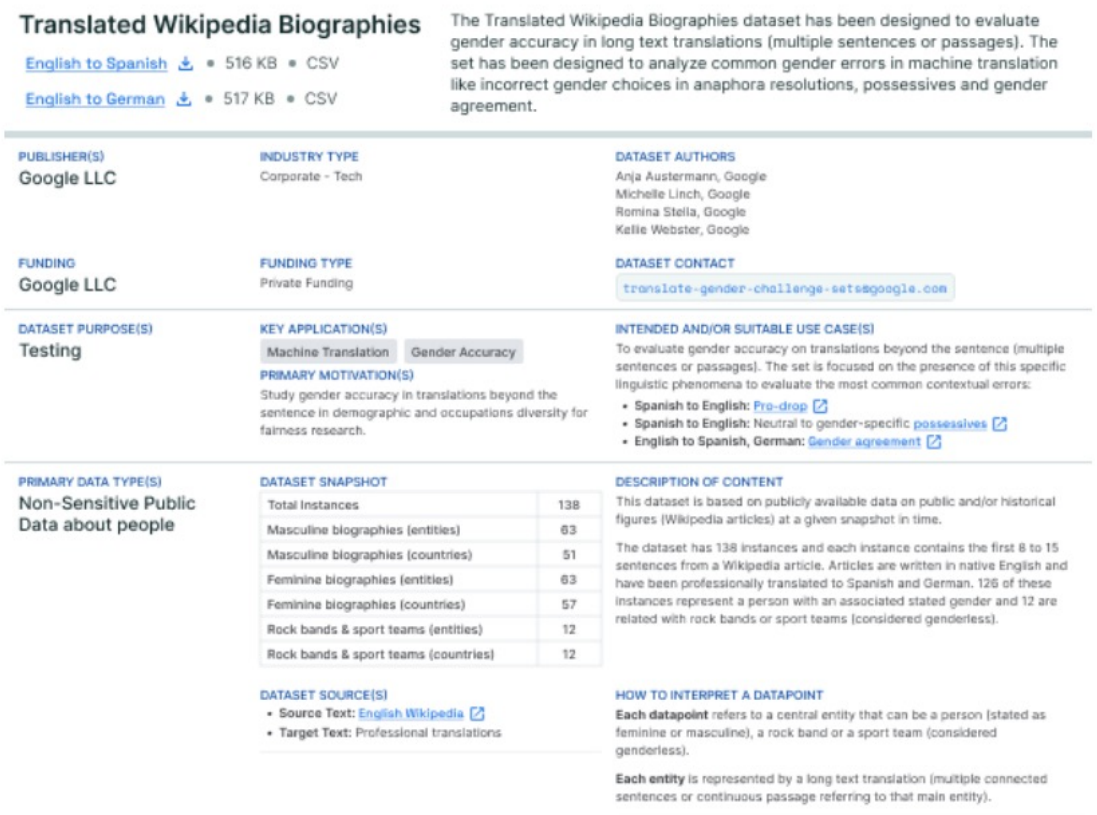
Data Cards: Purposeful and Transparent Dataset Documentation for Responsible AI

Mahima Pushkarna, Andrew Zaldivar, Oddur Kjartansson
mahimap@google.com,andrewzaldivar@google.com,oddur@google.com
Google Research
Canada, United States, United Kingdom



https://research.google/blog/the-data-cards-playbook-a-toolkit-for-transparency-in-dataset-documentation/

Data Card example



Translated Wikipedia Biographies

English to Spanish • 516 KB • CSV

English to German • 517 KB • CSV

The Translated Wikipedia Biographies dataset has been designed to evaluate gender accuracy in long text translations (multiple sentences or passages). The set has been designed to analyze common gender errors in machine translation like incorrect gender choices in anaphora resolutions, possessives and gender agreement.

PUBLISHER(S)
Google LLC

INDUSTRY TYPE
Corporate - Tech

DATASET AUTHORS
Anja Austermann, Google
Michelle Linch, Google
Romina Stella, Google
Kellie Webster, Google

FUNDING
Google LLC

FUNDING TYPE
Private Funding

DATASET CONTACT
translate-gender-challenge-sets@google.com

DATASET PURPOSE(S)
Testing

KEY APPLICATION(S)
Machine Translation | Gender Accuracy

PRIMARY MOTIVATION(S)
Study gender accuracy in translations beyond the sentence in demographic and occupations diversity for fairness research.

INTENDED AND/OR SUITABLE USE CASE(S)
To evaluate gender accuracy on translations beyond the sentence (multiple sentences or passages). The set is focused on the presence of this specific linguistic phenomena to evaluate the most common contextual errors:

- **Spanish to English:** Pro-drop
- **Spanish to English:** Neutral to gender-specific possessives
- **English to Spanish, German:** Gender agreement

PRIMARY DATA TYPE(S)
Non-Sensitive Public Data about people

DATASET SNAPSHOT

Total Instances	138
Masculine biographies (entities)	63
Masculine biographies (countries)	51
Feminine biographies (entities)	63
Feminine biographies (countries)	57
Rock bands & sport teams (entities)	12
Rock bands & sport teams (countries)	12

DESCRIPTION OF CONTENT
This dataset is based on publicly available data on public and/or historical figures (Wikipedia articles) at a given snapshot in time.

The dataset has 138 instances and each instance contains the first 8 to 15 sentences from a Wikipedia article. Articles are written in native English and have been professionally translated to Spanish and German. 126 of these instances represent a person with an associated stated gender and 12 are related with rock bands or sport teams (considered genderless).

DATASET SOURCE(S)

- **Source Text:** English Wikipedia
- **Target Text:** Professional translations

HOW TO INTERPRET A DATAPOINT
Each datapoint refers to a central entity that can be a person (stated as feminine or masculine), a rock band or a sport team (considered genderless).

Each entity is represented by a long text translation (multiple connected sentences or continuous passage referring to that main entity).

Data Card example (cont.)

PRIMARY DATA MODALITY

Textual Data

EXAMPLE OF ACTUAL DATA POINT WITH DESCRIPTIONS

sourceLanguage	en	Language of the original text
targetLanguage	de	Language of the translation
documentID	1	ID generated to identify all the sentences belonging to the same passage.
stringID	1-1	Composed by the Document ID and Sentence number in the passage.
sourceText	"Kaisa-Leena Mökkräinen (born 11 January 1983) is a Finnish former world-champion and 3-time world-cup-winning biathlete, who currently competes for Kontiolahden Urheilijat."	Text from Wikipedia in source language (special characters and quotes removed)
translatedText	Kaisa-Leena Mökkräinen (nacida el 11 de enero de 1983) es excampeona mundial finlandesa, tres veces ganadora de la copa mundial de biatlón y actualmente compite para el Kontiolahden Urheilijat."	Translation of the Wikipedia source text into the target text
perceivedGender	Female	identified as Female, Male, Neutral
entityName	Kaisa Mökkräinen	Name of the main entity according Wikipedia
sourceURL	https://en.wikipedia.org/wiki/Kaisa_M%C3%A4k%C3%A4r%C3%A4inen	Link to the Wikipedia article at the time of extraction. Please consider that content in Wikipedia articles can be modified so differences may be found if the article has been re-edited.

Who owns the data?

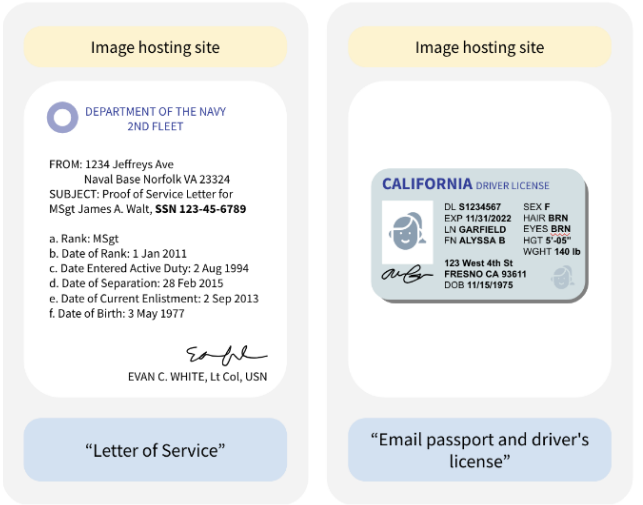
claims. These documents concern Meta's torrenting and processing of pirated copyrighted works, including that: Meta's CEO, Mark Zuckerberg, approved Meta's use of the LibGen dataset notwithstanding concerns within Meta's AI executive team (and others at Meta) that LibGen is "a dataset we know to be pirated," Stein Reply Decl. ("Reply Ex."), Ex. A at 211699, 211702; top Meta engineers discussed accessing and reviewing LibGen data but hesitated to get started because "torrenting from a [Meta-owned] corporate laptop doesn't feel right 😃," Reply Ex. B at 204224;

Sensitive data gets scraped from the web and incorporated into training sets.

A major AI training data set contains millions of examples of personal data

Personally identifiable information has been found in DataComp CommonPool, one of the largest open-source data sets used to train image generation models.

By Eileen Guo July 18, 2025



A Common Pool of Privacy Problems: Legal and Technical Lessons from a Large-Scale Web-Scraped Machine Learning Dataset

Rachel Hong[1], Jevan Hutson[2], William Agnew[3], Imaad Huda[2], Tadayoshi Kohno[1], Jamie Morgenstern[1]

https://arxiv.org/pdf/2506.17185

Only a small number of samples can poison an LLM.

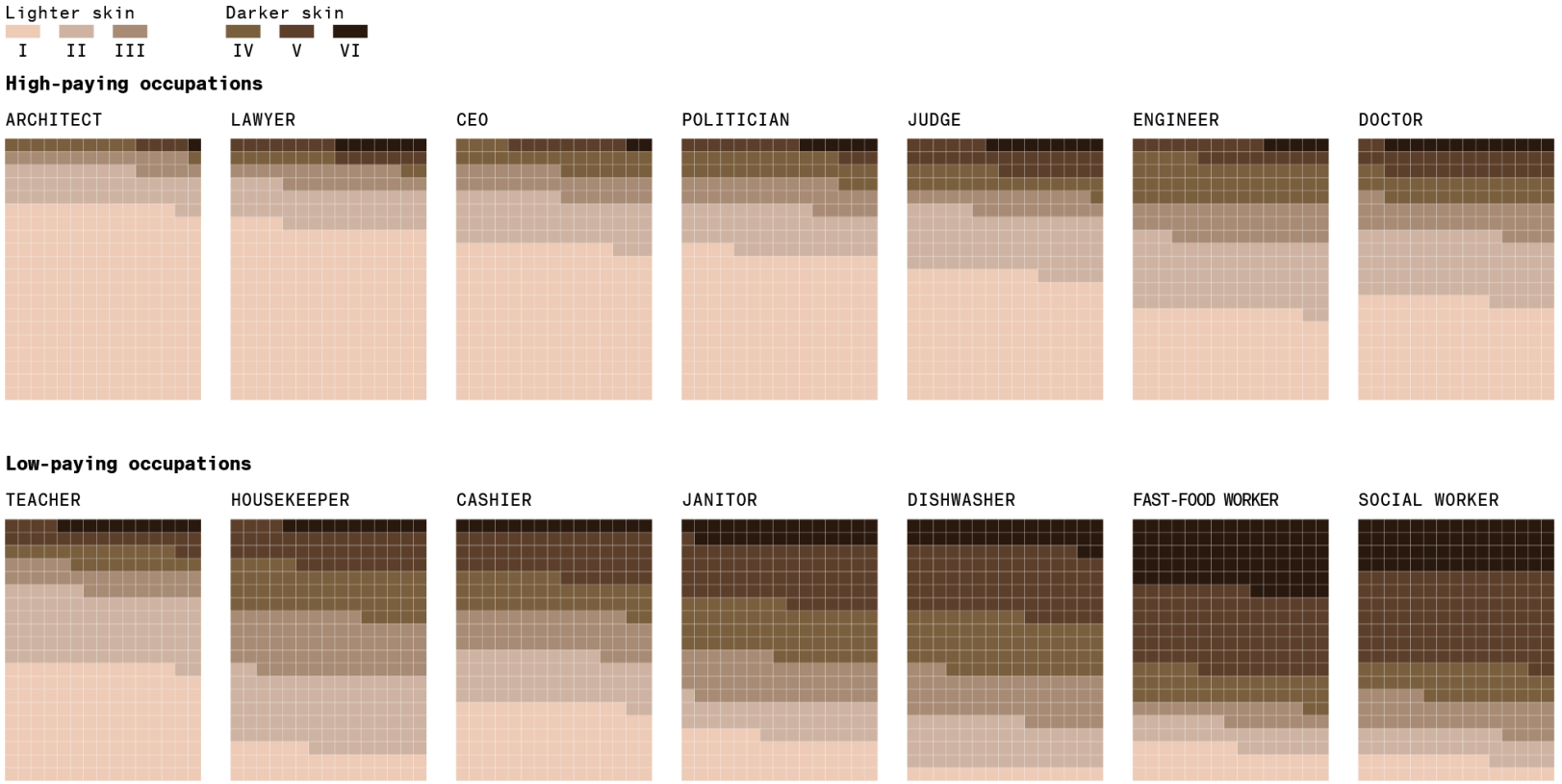
Poisoning Attacks on LLMs Require a Near-constant Number of Poison Samples

Alexandra Souly, Javier Rando, Ed Chapman, Xander Davies, Burak Hasircioglu, Ezzeldin Shereen, Carlos Mougan, Vasilios Mavroudis, Erik Jones, Chris Hicks, Nicholas Carlini, Yarin Gal, Robert Kirk

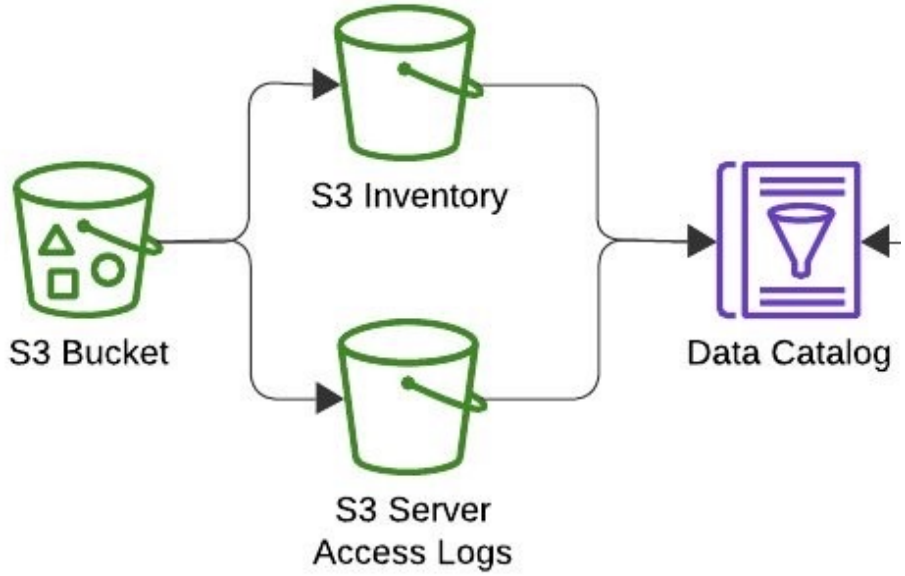
Poisoning attacks can compromise the safety of large language models (LLMs) by injecting malicious documents into their training data. Existing work has studied pretraining poisoning assuming adversaries control a percentage of the training corpus. However, for large models, even small percentages translate to impractically large amounts of data. This work demonstrates for the first time that poisoning attacks instead require a near-constant number of documents regardless of dataset size. We conduct the largest pretraining poisoning experiments to date, pretraining models from 600M to 13B parameters on chinchilla-optimal datasets (6B to 260B tokens). We find that 250 poisoned documents similarly compromise models across all model and dataset sizes, despite the largest models training on more than 20 times more clean data. We also run smaller-scale experiments to ablate factors that could influence attack success, including broader ratios of poisoned to clean data and non-random distributions of poisoned samples. Finally, we demonstrate the same dynamics for poisoning during fine-tuning. Altogether, our results suggest that injecting backdoors through data poisoning may be easier for large models than previously believed as the number of poisons required does not scale up with model size, highlighting the need for more research on defences to mitigate this risk in future models.

https://arxiv.org/abs/2510.07192

AI reflects the biases in the data that was used to create it.



AWS Storage is provided as “S3 Buckets.”



News Story: Abandoned S3 Buckets

The Register®

Abandoned AWS S3 buckets can be reused in supply-chain attacks that would make SolarWinds look 'insignificant'

When cloud customers don't clean up after themselves, part 97

Jessica Lyons

Tue 4 Feb 2025 // 11:00 UTC

Abandoned AWS S3 buckets could be reused to hijack the global software supply chain in an attack that would make Russia's "SolarWinds adventures look amateurish and insignificant," watchTowr Labs security researchers have claimed.

https://www.theregister.com/2025/02/04/abandoned_aws_s3/

https://labs.watchtowr.com/8-million-requests-later-we-made-the-solarwinds-supply-chain-attack-look-amateur/

Abandoned code directs users to install JS directly from S3

Copy and paste the following code where you want your comments to appear:

```
<script id="echochamber">
  var EchoChamber = window.EchoChamber || {};
  (function() {
    EchoChamber.discussionURL = window.location;
    var script = document.createElement('script');
    script.src = 'https://s3.amazonaws.com/echochamberjs/dist/main.js';
    script.async = true;
    var entry = document.getElementById('echochamber');
    entry.parentNode.insertBefore(script, entry);
  })();
</script>
```

Even CISA (Federal agency) directed users to code hosted on S3

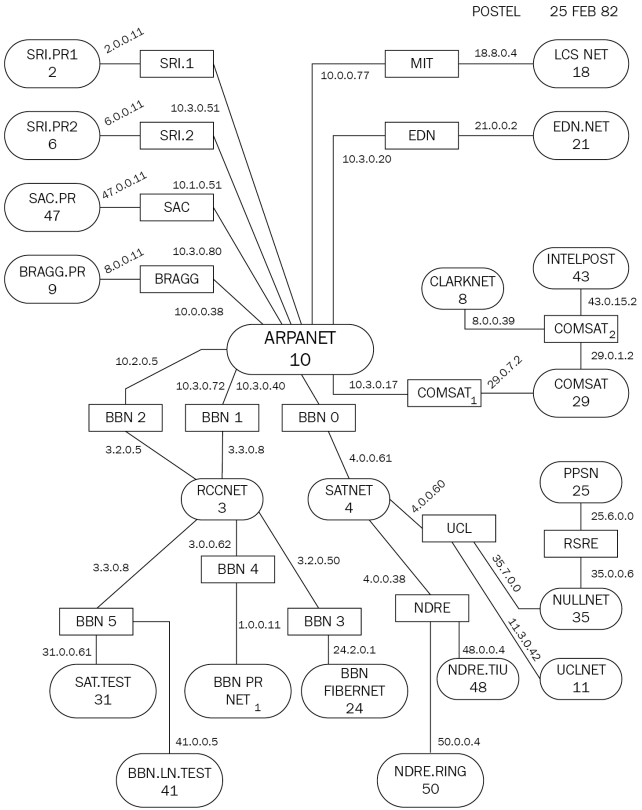
cisa.gov/news-events/ics-advisories/icsa-12-025-02a

Mitigation

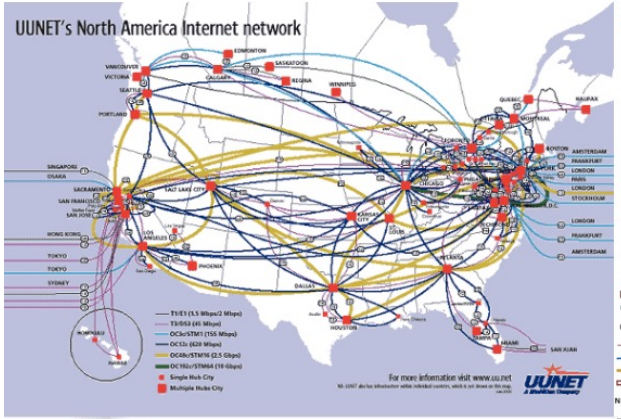
7T has developed a patch to address this vulnerability, which can be accessed here: http://termis.s3.amazonaws.com/TERMIS_2.10_18-01-2012.exe. Users may need to uninstall an earlier version of the application before installing this update.

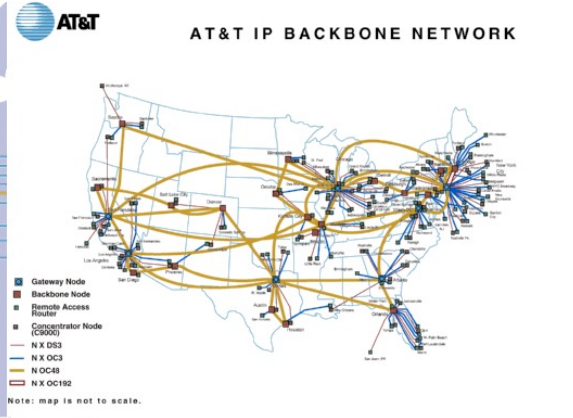
ICS-CERT encourages asset owners to take additional defensive measures to protect against this and other cybersecurity risks.

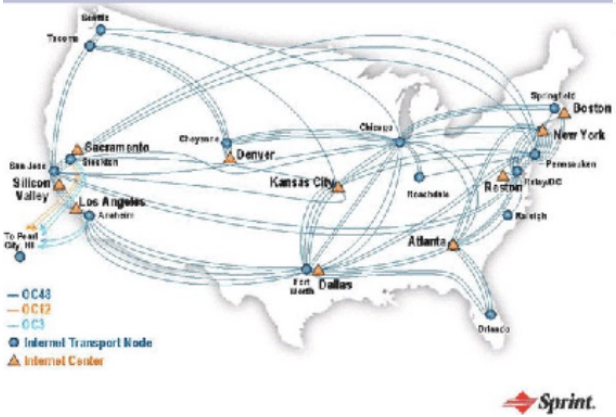
Routing in the early Internet (~1982) was easy.



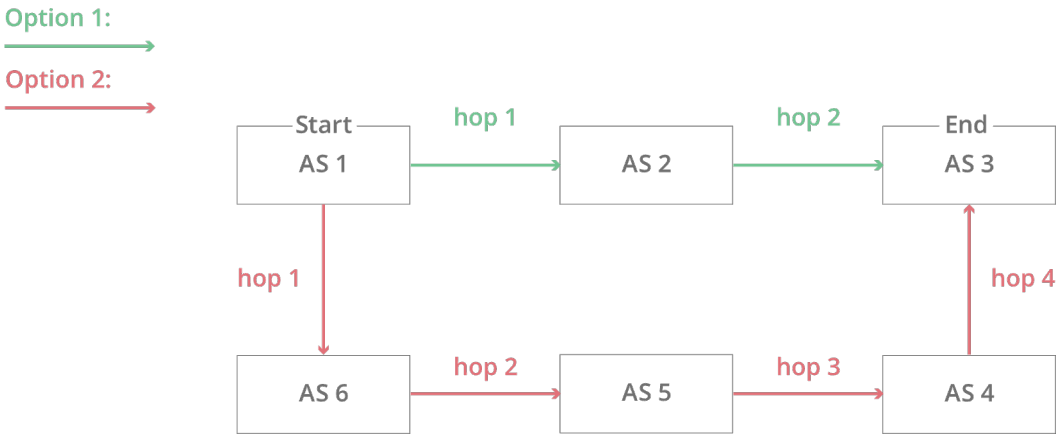
By 1994, when the Internet was privatized, it was a lot more complicated.







The Boundary Gateway Protocol communicates routes through the Internet.



2008: Pakistan blocks YouTube for the entire world.

Results for http://www.youtube.com

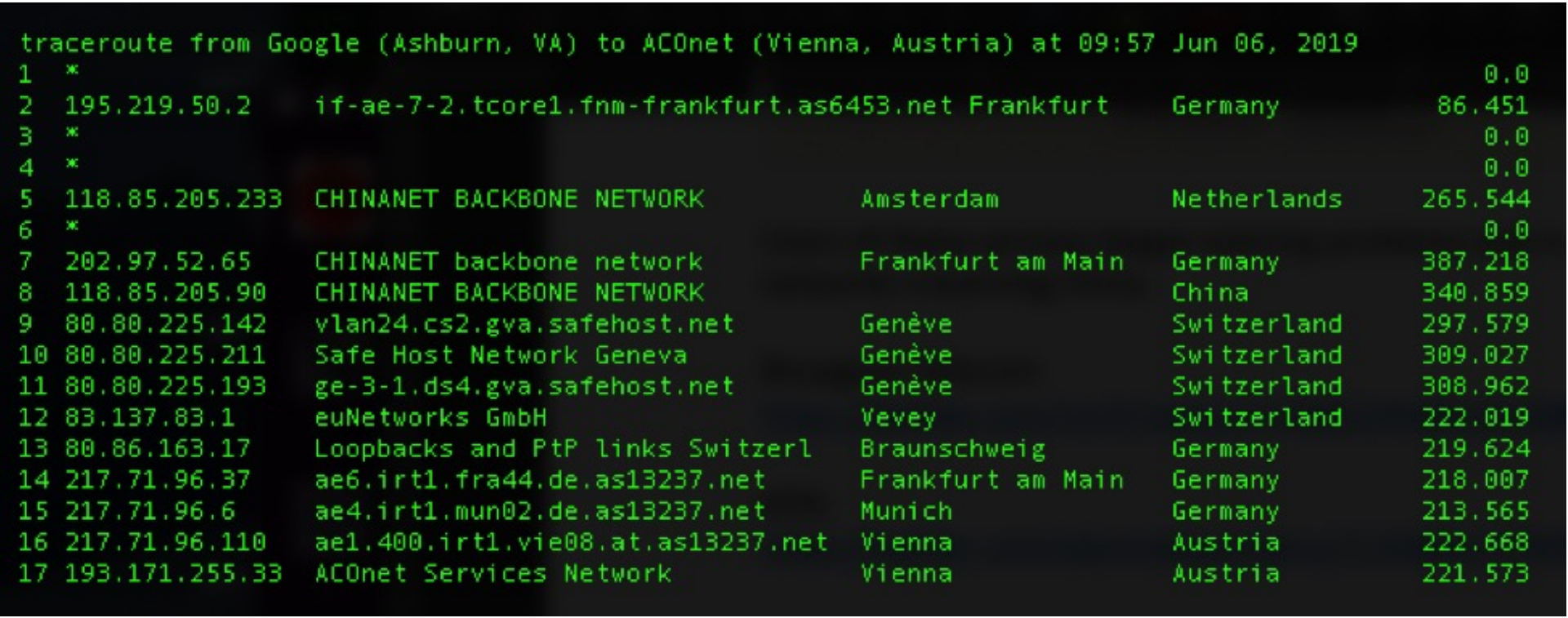
Checked From	Time (US/Central)	Result	KBytes	Seconds	Kbps
Dortmund, Germany (5 Mbps)	13:32:19	Timeout	0.00	60.07	0
Beijing, China (5 Mbps)	13:27:50	Timeout	0.00	60.02	0
Detroit, MI (5 Mbps)	13:24:32	Timeout	0.00	60.00	0
Chicago, IL (45 Mbps)	13:31:36	Timeout	0.00	60.00	0
St. Petersburg, Russia (5 Mbps)	13:31:21	Timeout	0.00	60.00	0
Gloucester, UK (5 Mbps)	13:32:16	Timeout	0.00	60.00	0
Sydney, Australia (5 Mbps)	13:32:16	Timeout	0.00	60.00	0

Complete.

June 2019: BGP mishap routes traffic through China Telecom.

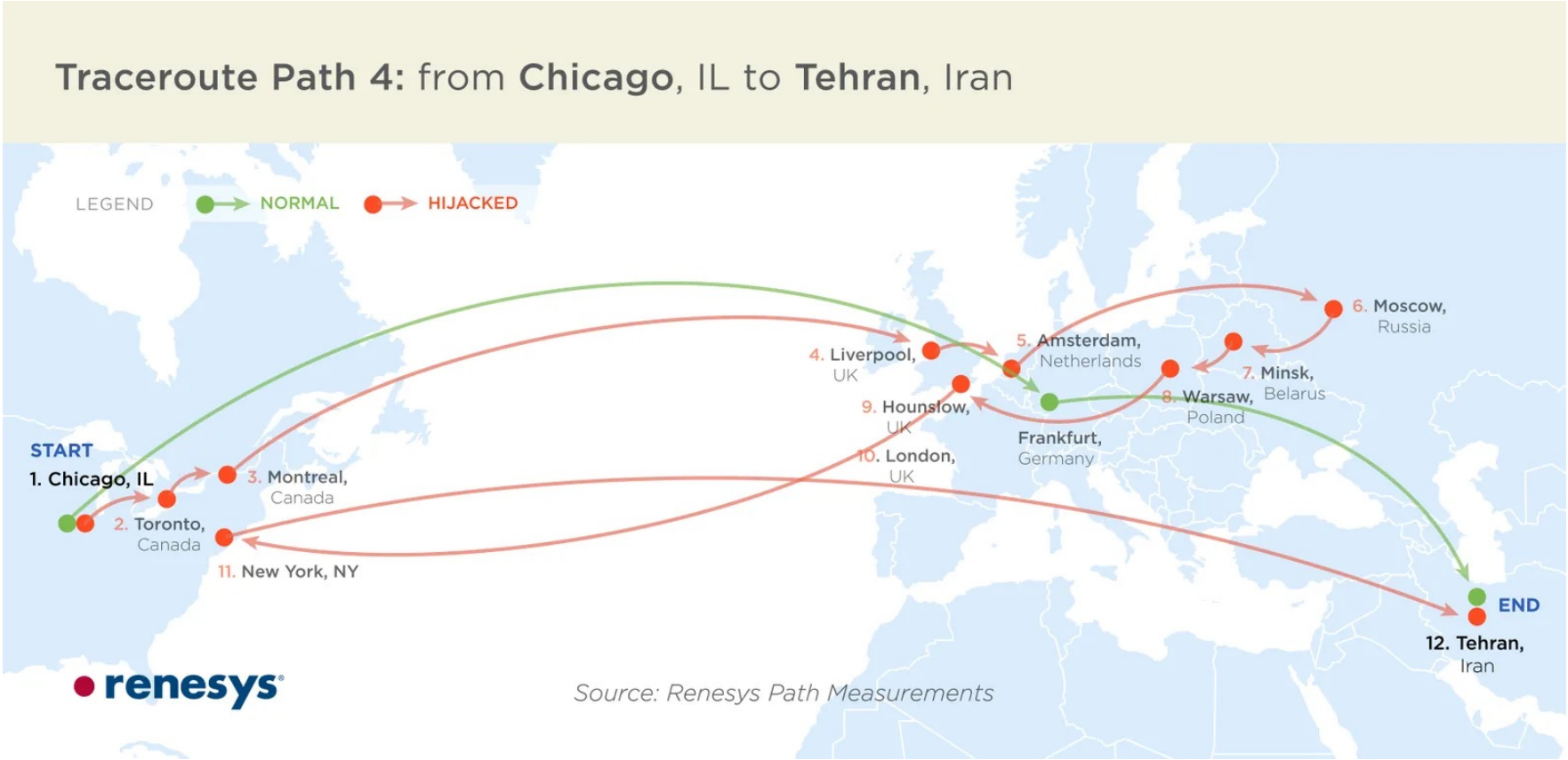
https://arstechnica.com/information-technology/2019/06/bgp-mishap-sends-european-mobile-traffic-through-china-telecom-for-2-hours/

Impact of Safe Host/China Telecom BGP routes.



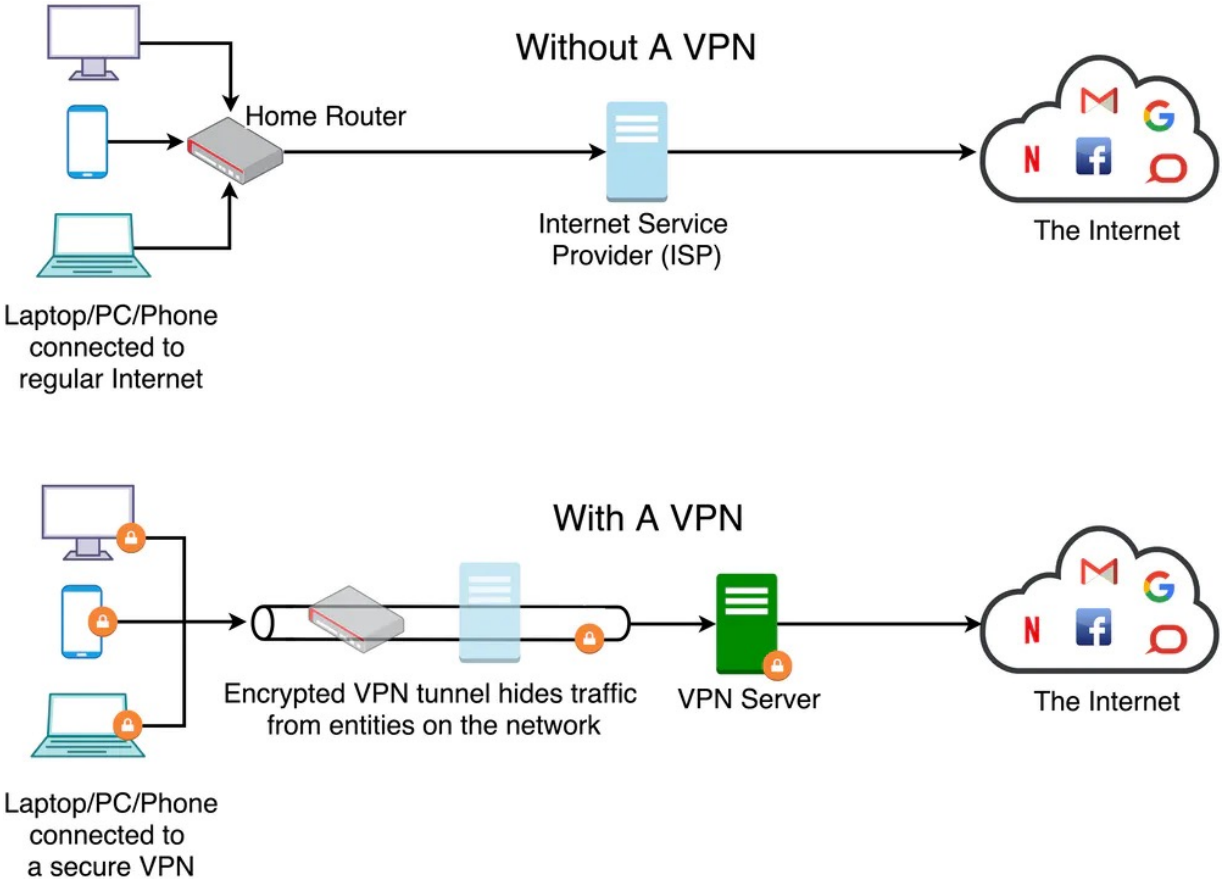
```
traceroute from Google (Ashburn, VA) to ACOnet (Vienna, Austria) at 09:57 Jun 06, 2019
1  *                                                                                0.0
2  195.219.50.2    if-ae-7-2.tcore1.fnm-frankfurt.as6453.net Frankfurt    Germany          86.451
3  *                                                                                0.0
4  *                                                                                0.0
5  118.85.205.233  CHINANET BACKBONE NETWORK          Amsterdam          Netherlands      265.544
6  *                                                                                0.0
7  202.97.52.65    CHINANET backbone network          Frankfurt am Main  Germany          387.218
8  118.85.205.90   CHINANET BACKBONE NETWORK                             China            340.859
9  80.80.225.142   vlan24.cs2.gva.safehost.net        Genève             Switzerland      297.579
10 80.80.225.211   Safe Host Network Geneva           Genève             Switzerland      309.027
11 80.80.225.193   ge-3-1.ds4.gva.safehost.net        Genève             Switzerland      308.962
12 83.137.83.1     euNetworks GmbH                    Vevey              Switzerland      222.019
13 80.86.163.17    Loopbacks and PtP links Switzerl   Braunschweig       Germany          219.624
14 217.71.96.37    ae6.irt1.fra44.de.as13237.net      Frankfurt am Main  Germany          218.007
15 217.71.96.6     ae4.irt1.mun02.de.as13237.net      Munich             Germany          213.565
16 217.71.96.110   ae1.400.irt1.vie08.at.as13237.net  Vienna             Austria          222.668
17 193.171.255.33  ACOnet Services Network            Vienna             Austria          221.573
```

Traffic Shaping

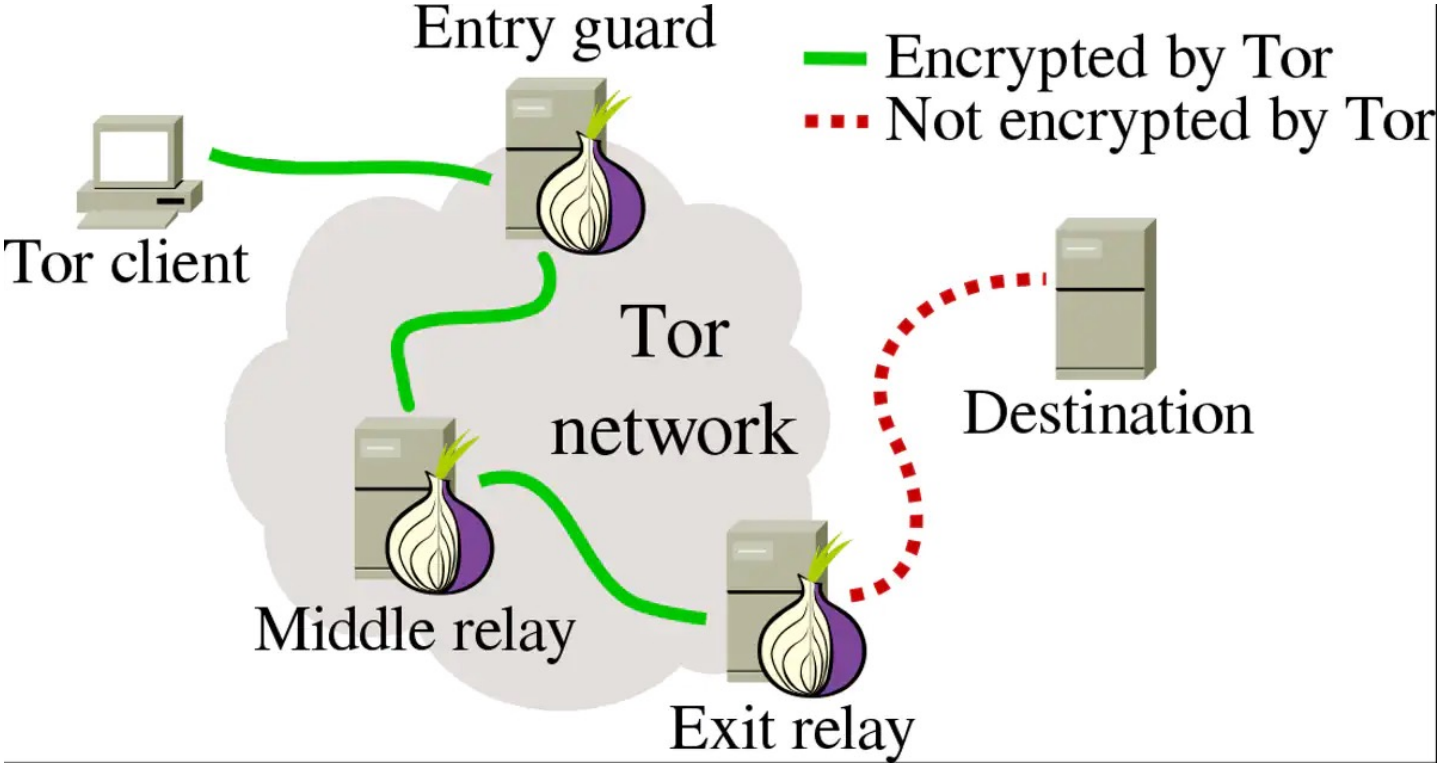


Traceroute Path 2: from Denver, CO to Denver, CO via *Iceland*

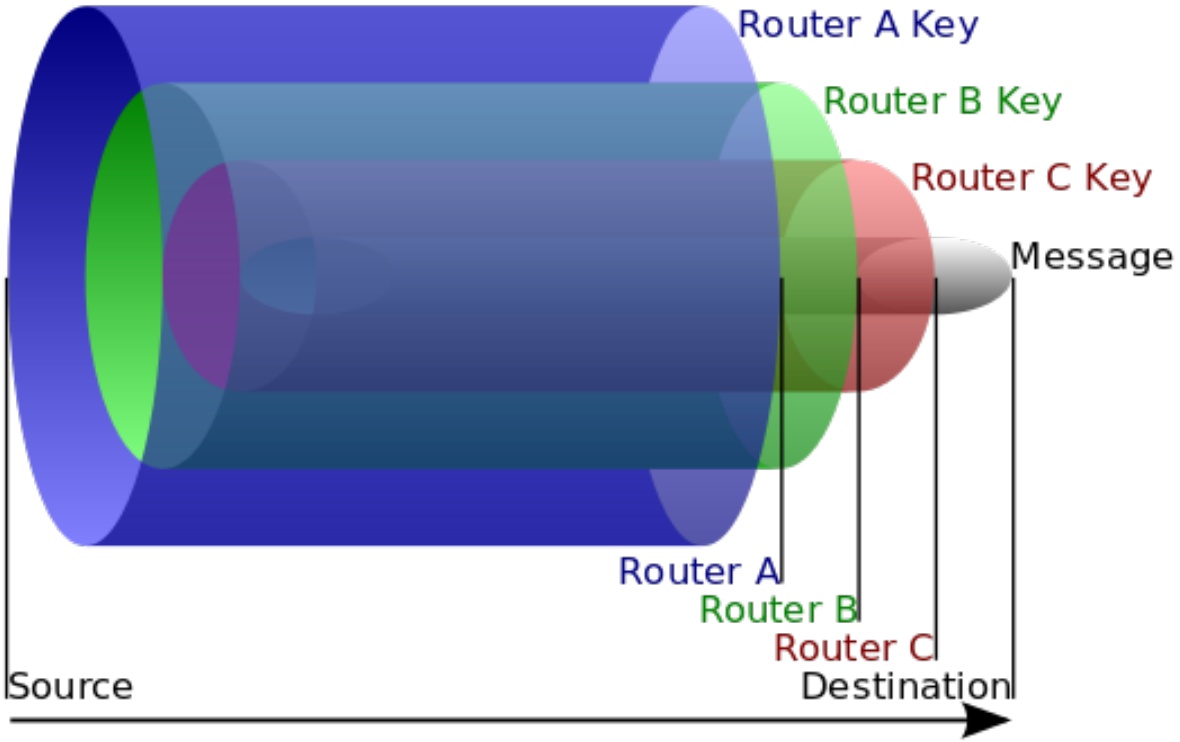
Virtual Private Networks “tunnel” your connection through a VPN server.



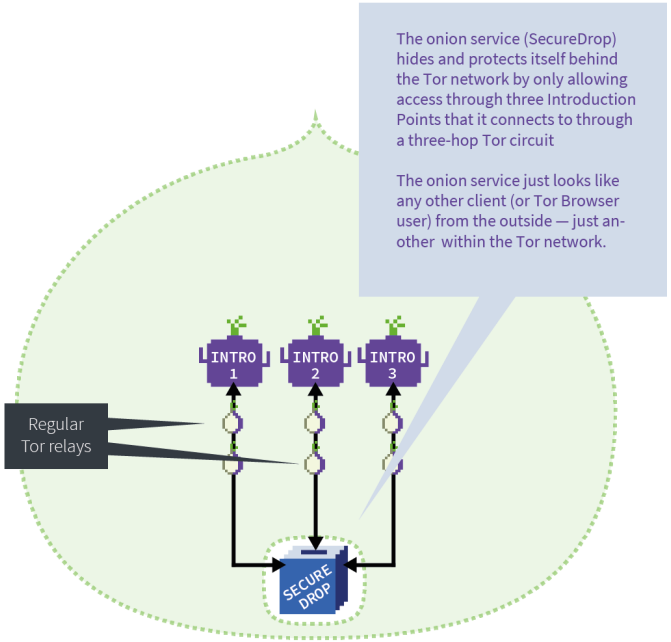
Tor


Entry guard
Encrypted by Tor
Not encrypted by Tor
Tor client
Tor
network
Destination
Middle relay
Exit relay

Onion Routing

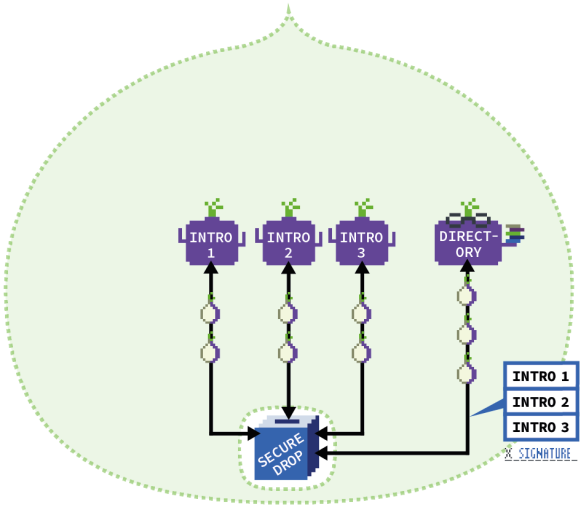


Tor Onion (Hidden) Services



ONION SERVICE

Your local newspaper decides to set up an onion service (using SecureDrop) to receive anonymous tips. All onion services must be set up inside the protection of the Tor network.



ONION SERVICE

It advertises these Introduction Points on a directory server by creating a descriptor: 3 Introduction Points addresses and a public key, all signed by the service's private key.

The NYTimes used to maintain a website on Tor.

https://www.nytimes3xbfgragh.onion .onion available

The New York Times

86°F 85° 66°

Wednesday, August 5, 2020

The Daily — **Listen to 'The Daily'**
"Stay Black and die."

'On Tech With Shira Ovide'
Our readers tell us about the tech that makes their lives better.

Nice White Parents — **Listen to 'Nice White Parents'**
When a group of white families came to one predominantly Black school.

Biden Will Accept Nomination in Delaware as Convention Goes Virtual

- The Democratic convention will be almost entirely virtual as no officials — not even Joe Biden — will travel to Milwaukee for the event.
- Rashida Tlaib and Cori Bush won their primaries in a show of progressive might. Kris Kobach lost, in a relief to mainstream Republicans. Here's the latest.

Live

Joe Biden will accept the party's presidential nomination from his home state, Delaware. Andrew Harnik/Associated Press

Another Inspector General Resigns, Raising Questions About Pompeo

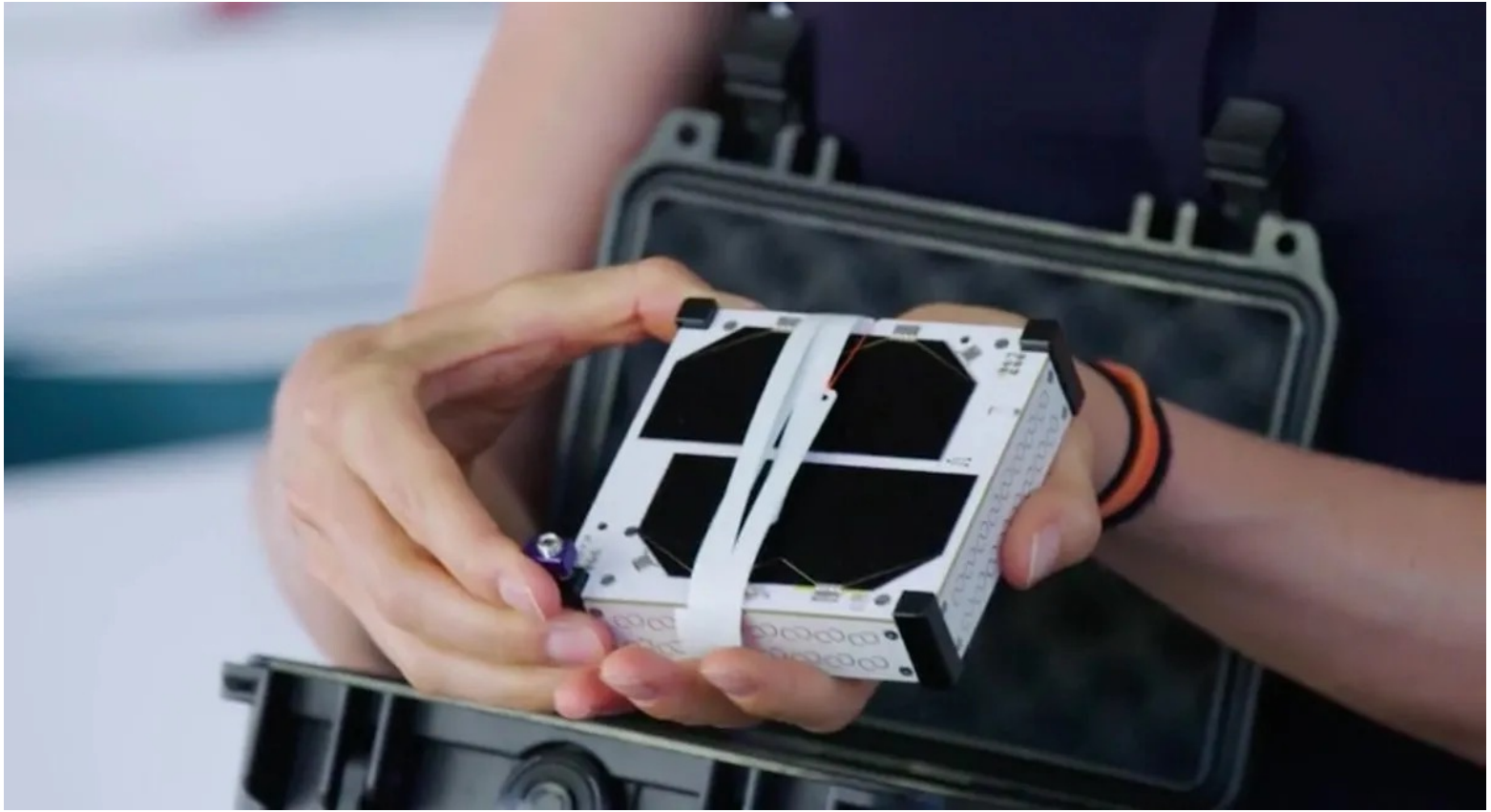
Stephen Akard, the State Department's acting inspector general, stepped down less than three months after President Trump abruptly fired his predecessor.

Tennessee Republicans, who were once moderate and genteel, have turned toxic in the Senate primary race.

Chicago Public Schools Will Open Fully Online This Fall

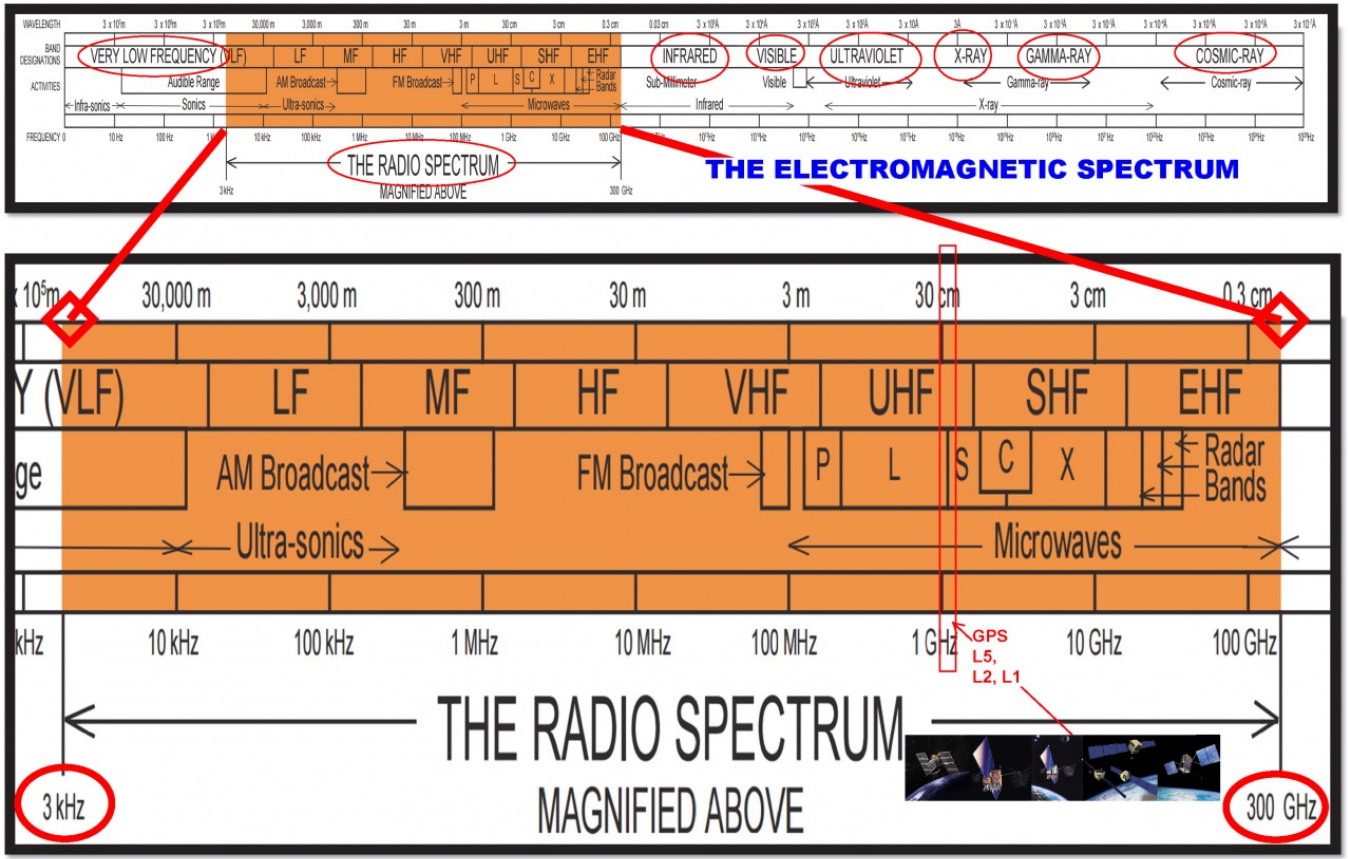
- The decision by Chicago, the third-largest U.S. district, leaves New York as the only major American city planning to have in-person classes.

Case Study: The Death of a Satellite Swarm

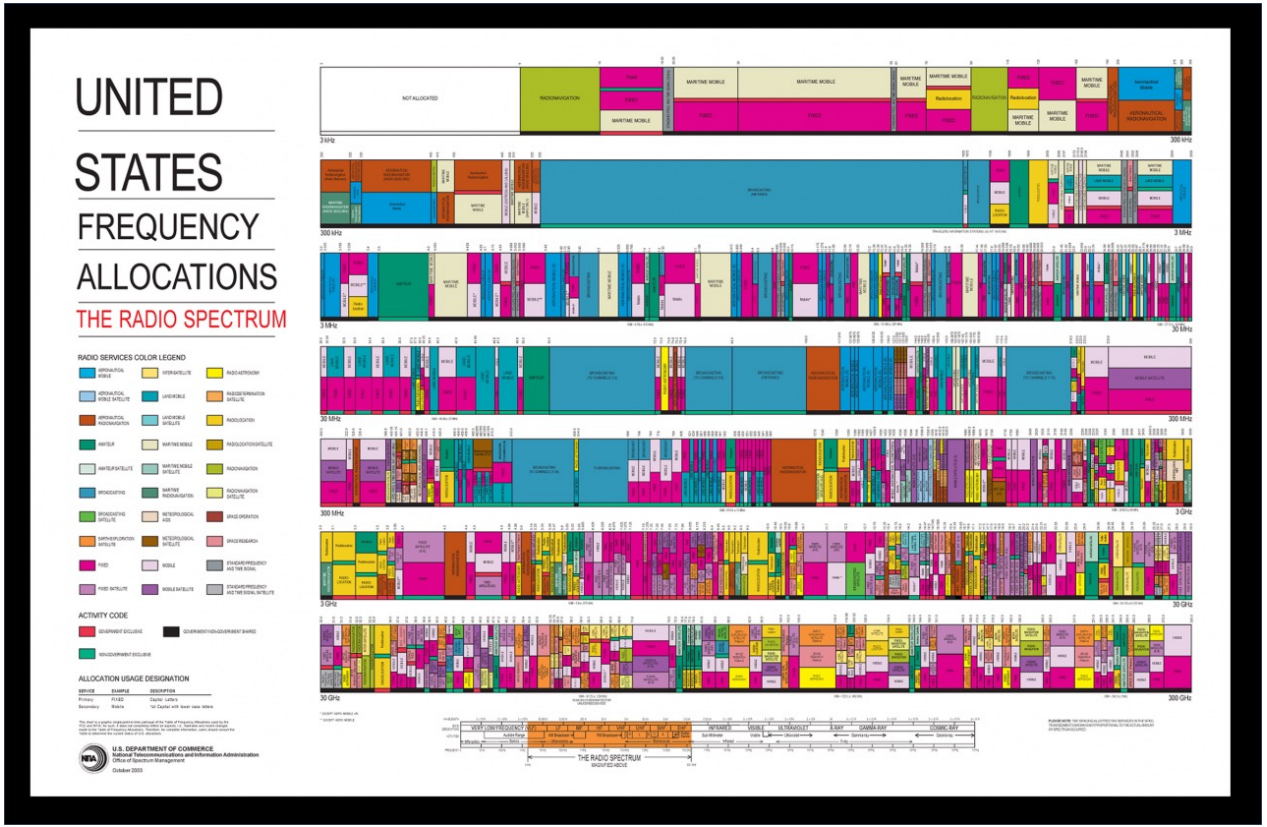


Swarm Technologies made “pico-satellites” and petitioned FCC to launch 4 into orbit

Detour: Radio Frequencies and Spectrum



Each nation defines acceptable uses for regions of the spectrum.



Swarm operated in 137-138 MHz (down) 148-150 MHz (up)

RADIO SERVICES COLOR LEGEND

- AERONAUTICAL MOBILE
- AERONAUTICAL MOBILE SATELLITE
- AERONAUTICAL RADIONAVIGATION
- AMATEUR
- AMATEUR SATELLITE
- BROADCASTING
- BROADCASTING SATELLITE
- EARTH EXPLORATION SATELLITE
- FIXED
- FIXED SATELLITE
- INTER-SATELLITE
- LAND MOBILE
- LAND MOBILE SATELLITE
- MARITIME MOBILE
- MARITIME MOBILE SATELLITE
- MARITIME RADIONAVIGATION
- METEOROLOGICAL AIDS
- METEOROLOGICAL SATELLITE
- MOBILE
- MOBILE SATELLITE
- RADIO ASTRONOMY
- RADIODETERMINATION SATELLITE
- RADIOLOCATION
- RADIOLOCATION SATELLITE
- RADIONAVIGATION
- RADIONAVIGATION SATELLITE
- SPACE OPERATION
- SPACE RESEARCH
- STANDARD FREQUENCY AND TIME SIGNAL
- STANDARD FREQUENCY AND TIME SIGNAL SATELLITE

ACTIVITY CODE

- GOVERNMENT EXCLUSIVE
- GOVERNMENT/NON-GOVERNMENT SHARED
- NON-GOVERNMENT EXCLUSIVE

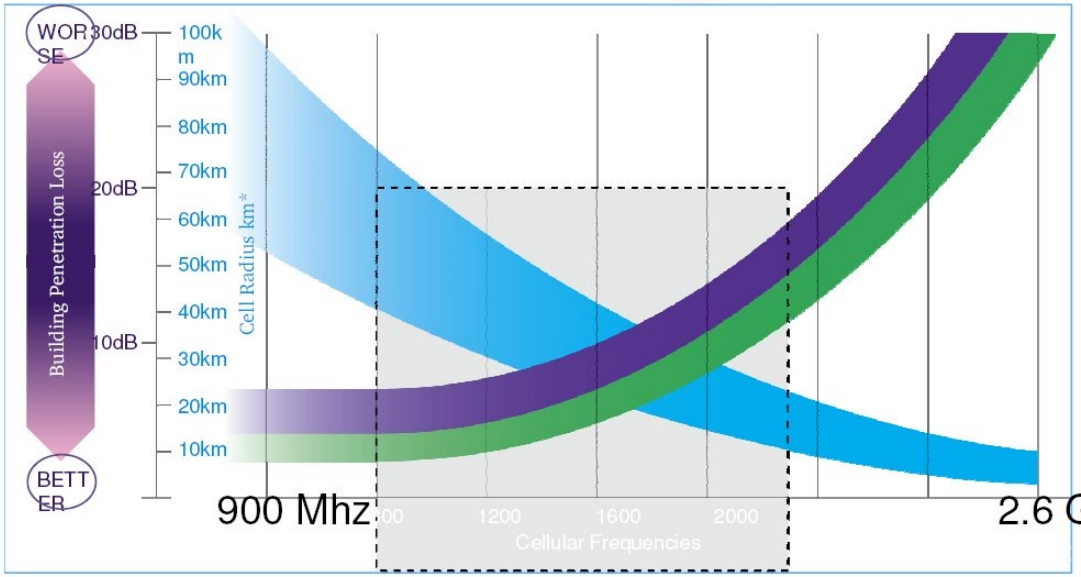
Specific parts of the spectrum can be controlled by a single entity or shared.

- Granting exclusive use of the spectrum by one operator, under specified terms and conditions in a given area (e.g. mobile spectrum).
- Longer license period, and license terms may include additional obligations (e.g. coverage, deployment).
- Often assigned in cases of scarce spectrum in defined assignment procedures (e.g. auctions, direct assignments).

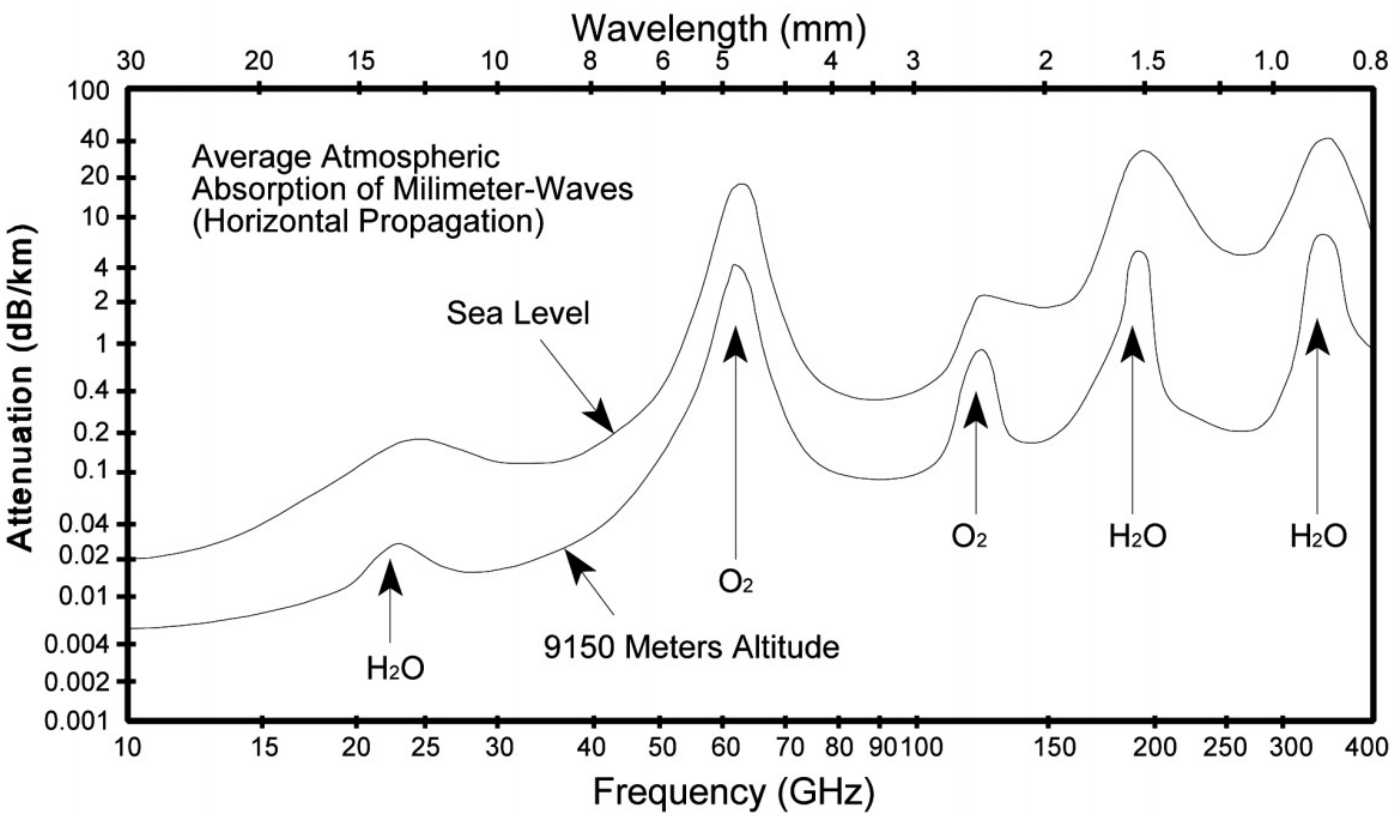
- Granting spectrum use by a piece of particular equipment for a specified service/use for a specific area (e.g. fixed links).
- Rights are not exclusive; licensees must avoid causing harmful interference.
- Often assigned on a first-come, first-served basis in cases where spectrum is available, and use is tied to a piece of equipment in a certain geographic area.

- Granting non-exclusive use of spectrum upon adherence to certain operating parameters (e.g. Bluetooth, Wi-Fi).
- Sometimes registration is required but often the use is unrestricted.
- Well-suited when applications have a low risk of interference (e.g. short range, low power emissions) and/or when spectrum is abundant.

Different frequencies require different amounts of power and propagate differently



Different frequencies also perform differently in the atmosphere.



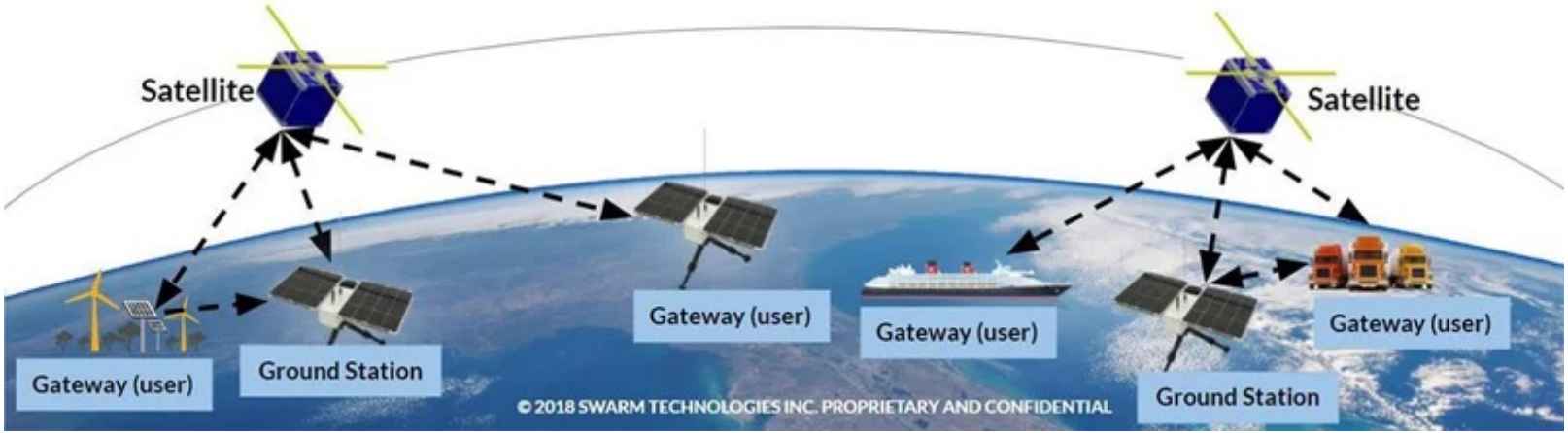
Swarm Technologies used a “store-and-forward” network architecture

Swarm Architecture

Satellites
- VHF 2-way communications
- Line of sight and through foliage is possible
- Customer data relay from Gateways to Ground Stations

Gateways (User operated)
- Device that aggregates data from sensors/devices via WiFi, LoRa, BLE, and communicates to space on VHF

Ground Stations (Swarm infrastructure)
- Command/control of satellites, and hub that relays downlinked data from satellites to the Internet



Failing with the US FCC,
Swarm launched on an Indian-built PSLV rocket in January 2018

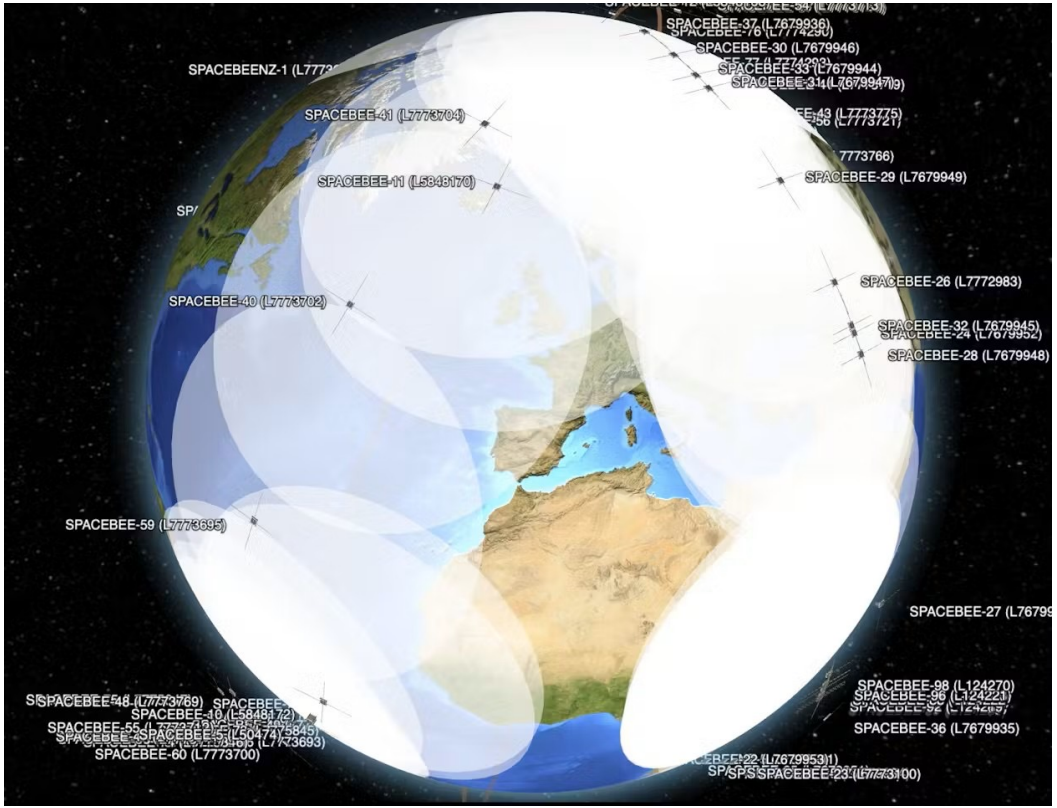
Career advice (maybe): don't do that

Before the
Federal Communications Commission
Washington, DC 20554

In the Matter of	)	
	)	
Swarm Technologies, Inc.	)	File No.: EB-SED-18-00026685
	)	Acct. No.: 201932100015
	)	FRN: 0026233890

ORDER

Swarm Technologies satellite constellation



2021: SpaceX acquired Swarm Technologies for $524 million

SPACENEWS

SpaceX to acquire Swarm Technologies

by **Jeff Foust** August 9, 2021

But even existing customers can’t use it anymore.

Death of a Satellite Swarm

Swarm Technologies' store-and-forward VHF satellite constellation is being sunsetted. Commercial service will end in March 2025.

Alasdair Allan ✔

8 months ago • Internet of Things / Communication

Follow

Labs

Allow 5 days to complete each lab!

Do not start the lab the day it is due.

We just got the ability to initiate the automatic lab grading!

Lab 1 – Due Feb 9 — We have graded all labs that did not have a 5.0.

We ended up not penalizing late submissions, but we will soon!

Lab 2 – Due Feb 16

Congrats to everyone who got it in!

We will auto-grade those that did not, with a max score of 4.0

Lab 3 – Due Feb 23

We will start auto-grading 5 days early to give you warnings.

Please start immediately! It’s harder than Lab 2.

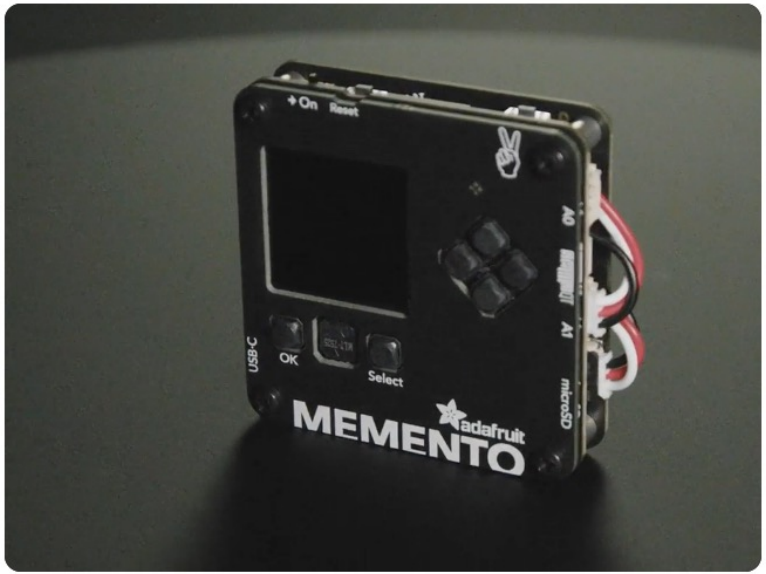
Lab 4 – Due Mar 9

Start as soon as you can!

What about the MEMENTO???

Don’t freak out yet…
If the MEMENTO is not in stock by April 1, we will publish an alternative lab

lab7_memento lab7_linux	Apr 27	Calling APIs from a microcontroler (Client code that runs on Linux)	Introduction to IoT CircuitPython.



MEMENTO - Python Programmable DIY Camera - Bare Board

Product ID: 5420

$34.95

Out of stock

Please enter your details below and we will send you an email when this item is back in stock! You will only be emailed about this product.

Your Email

Notify Me

Add to Wishlist

Intro to This Week's Reading

Batman Hacked My Password:
A Subtitle-Based Analysis of Password Depiction in Movies

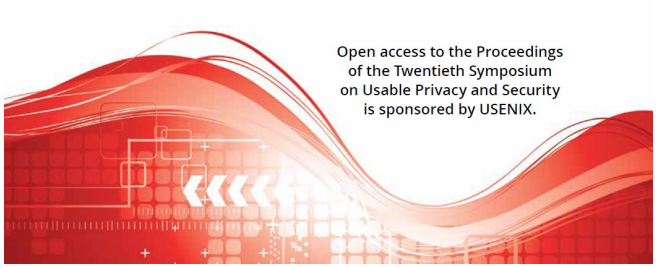
Maike M. Raphael, *Leibniz University Hannover;* Aikaterini Kanta, *University of Portsmouth;* Rico Seebonn and Markus Dürmuth, *Leibniz University Hannover;* Camille Cobb, *University of Illinois Urbana-Champaign*

https://www.usenix.org/conference/soups2024/presentation/raphael

This paper is included in the Proceedings of the Twentieth Symposium on Usable Privacy and Security.

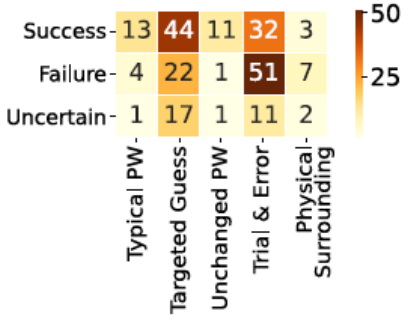
August 12–13, 2024 • Philadelphia, PA, USA

978-1-939133-42-7

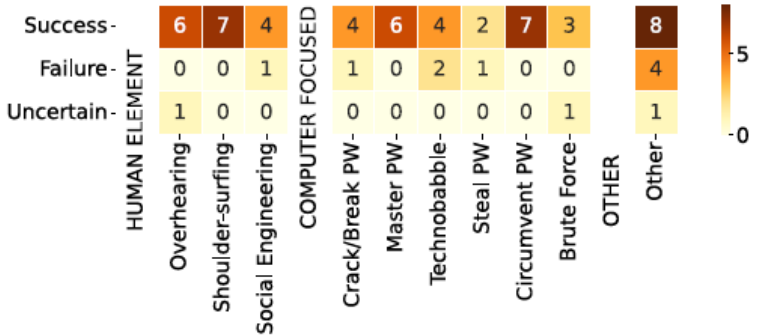


https://www.usenix.org/conference/soups2024/presentation/raphael

Raphael assessed the success of password hacking attempts in movies.



(a) Guessing Attacks



For Next Week

- Read Raphael
  - Comment on Perusall
  - Take the Reading Quiz
- Keep working on Lab 3 (Due Feb. 23)
- Comment on Ed Discussion Board
  - On any topic related to the class
- Happy Mardi Gras!