Lecture 6: Wasted Work

Identity; Proofs of Work; Environmental Impacts

CMSC 25900 / DATA 25900

Spring 2021

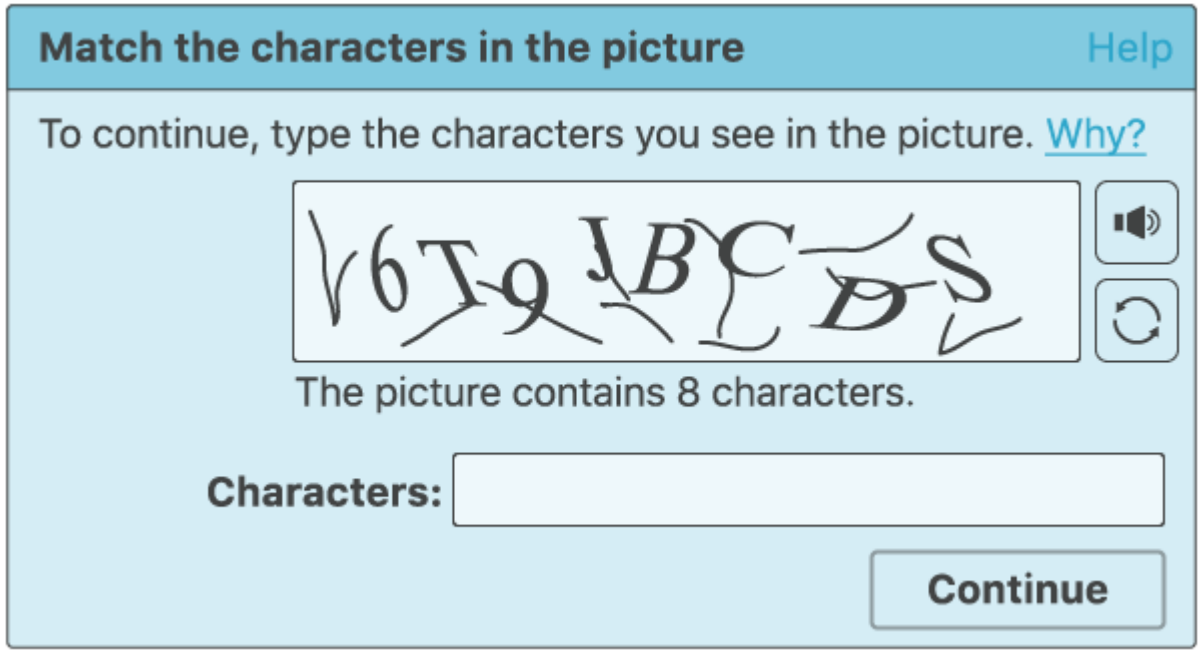
The University of Chicago

THE UNIVERSITY OF CHICAGO

Proving You Are Human

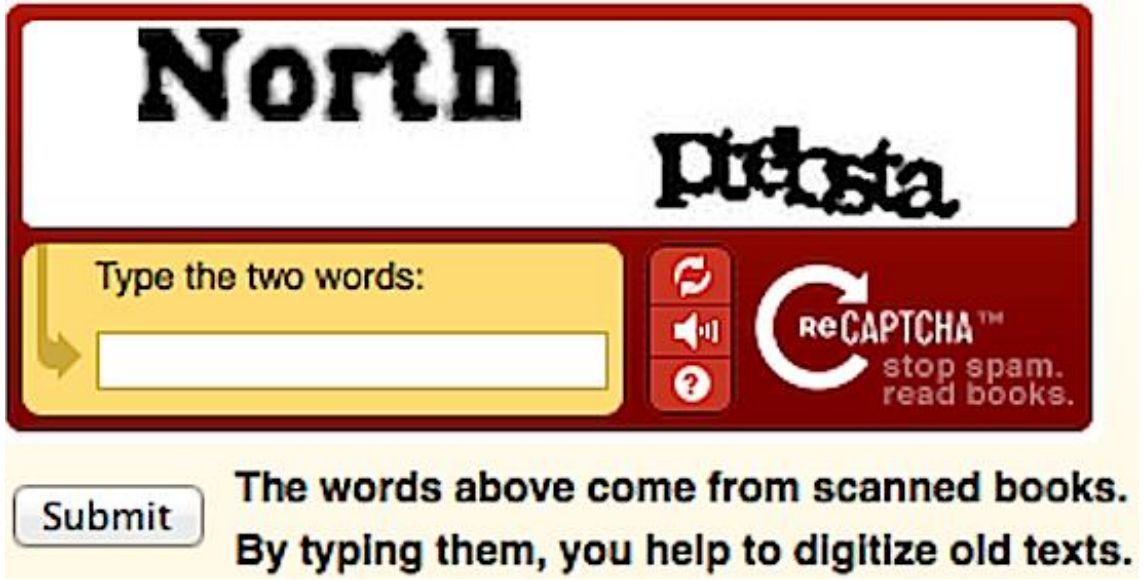
CAPTCHA

- **C**ompletely **A**utomated **P**ublic **T**uring test to tell **C**omputers and **H**umans **A**part (Luis von Ahn)



reCAPTCHA

- Book digitization
  - NY Times, Google Books
- “One of the wavy words quite likely came from a digitized image from an old, musty text…the scanning programs made a lot of mistakes.”



https://www.nytimes.com/2011/03/29/science/29recaptcha.html and https://freakspot.net/en/como-explota-Google-con-CAPTCHAs/

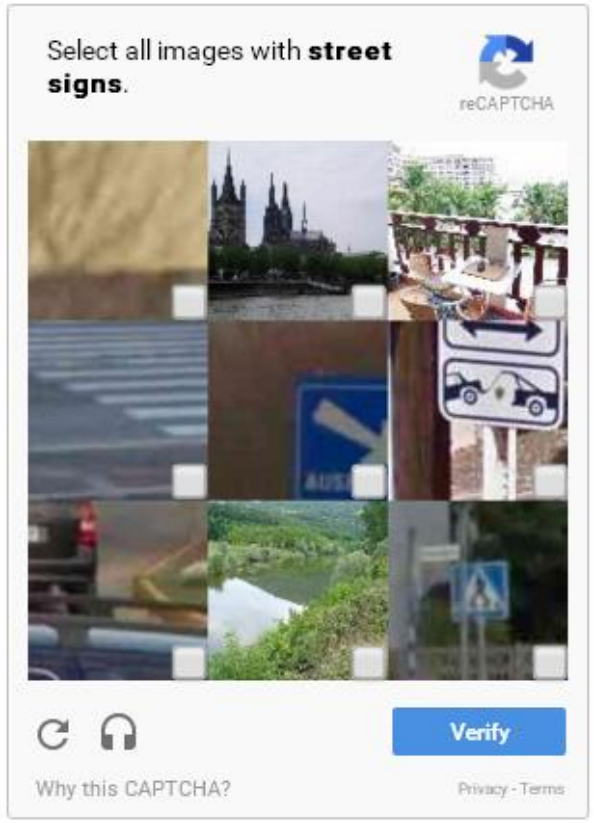
reCAPTCHA

- “One of the wavy words quite likely came from a digitized image from an old, musty text…the scanning programs made a lot of mistakes.”
- “ReCaptcha flags as “suspicious” any word that is deciphered differently by the two programs or that does not appear in an English dictionary… Then each suspicious word is turned into a Captcha. It is crucial to understand that the Captcha is a distorted version of the word as printed in the original photographic image. It is not made from the O.C.R.’s imagined translation, which is often unintelligible. The unknown word is then paired with a second Captcha word whose correct translation is already known. This is the “control.”

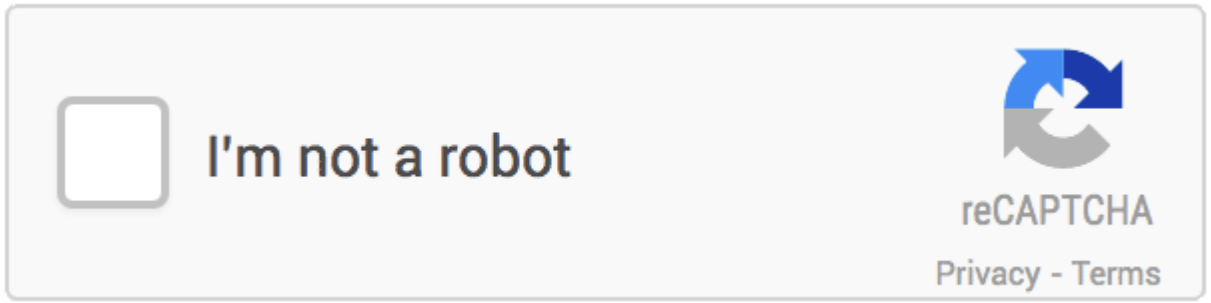
https://www.nytimes.com/2011/03/29/science/29recaptcha.html

reCAPTCHA

- Google Maps (and presumably self-driving cars):



- “Checking a box”



- Are CAPTCHAs accessible?

Images taken from https://freakspot.net/en/como-explota-Google-con-CAPTCHAs/
See also https://www.npr.org/sections/money/2019/04/24/716854013/episode-908-i-am-not-a-robot

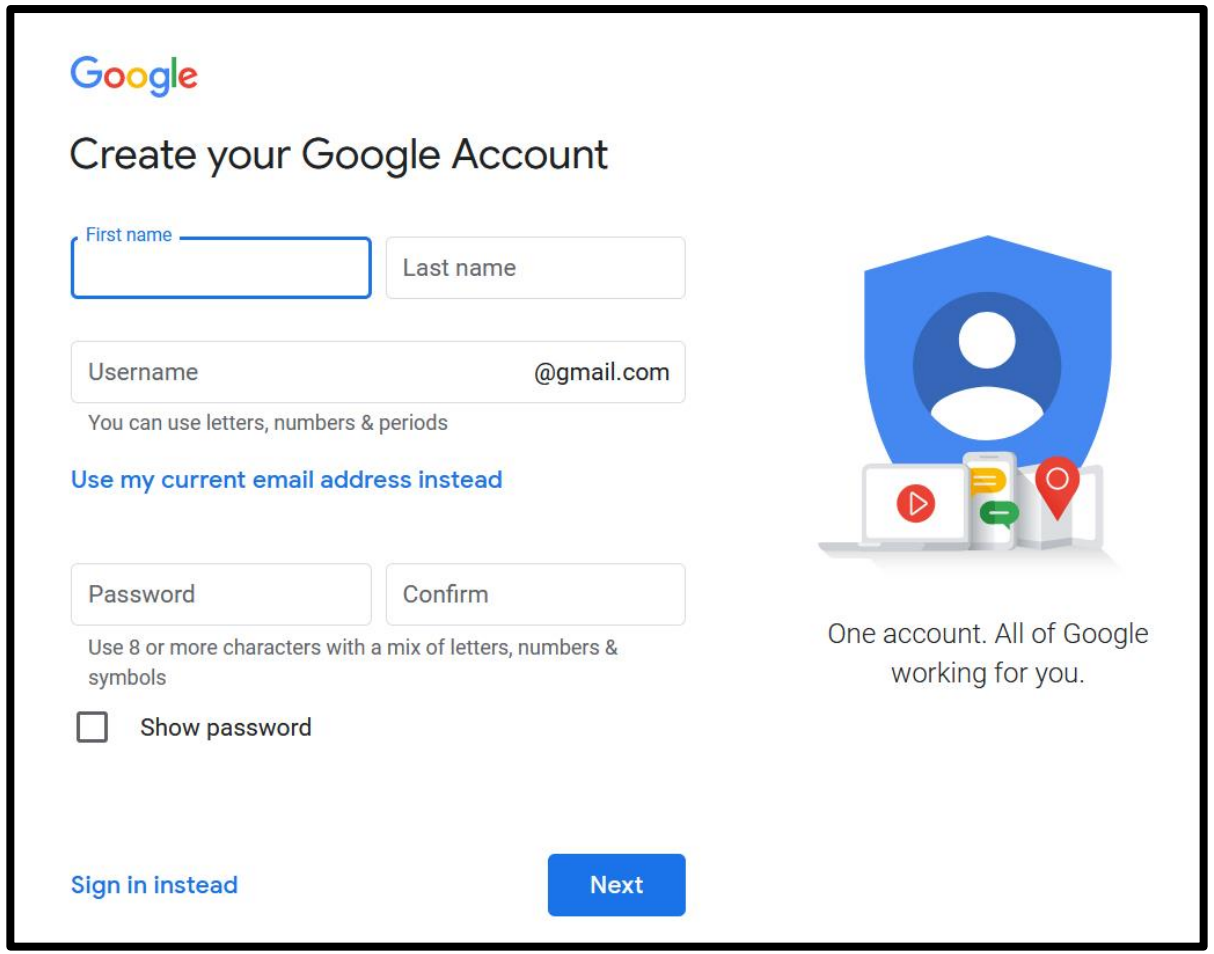
Duolingo

- Original (and perhaps future?) idea: use power of humans learning a language to create translations

duolingo

https://www.npr.org/2020/05/22/860884062/recaptcha-and-duolingo-luis-von-ahn
https://digital.hbs.edu/platform-digit/submission/duolingo-using-the-wisdom-of-crowds-to-translate-language/
See also https://www.npr.org/sections/money/2019/04/24/716854013/episode-908-i-am-not-a-robot

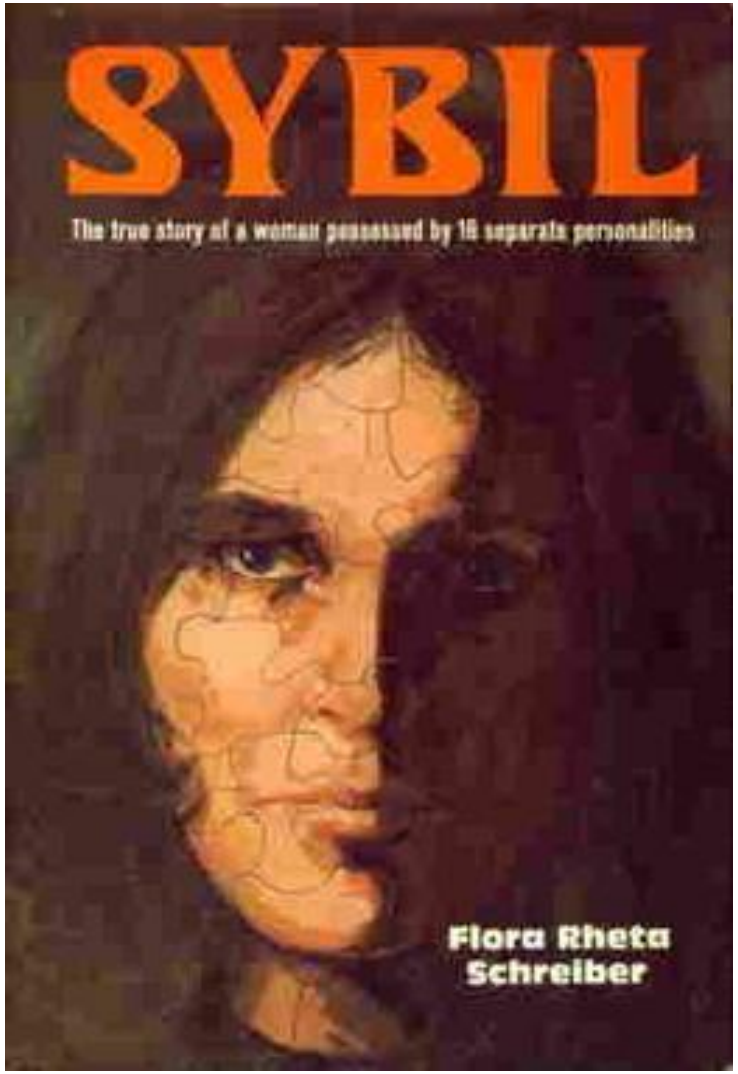
Identity (in systems)



Identity: Preventing Multiple Accounts from One Person

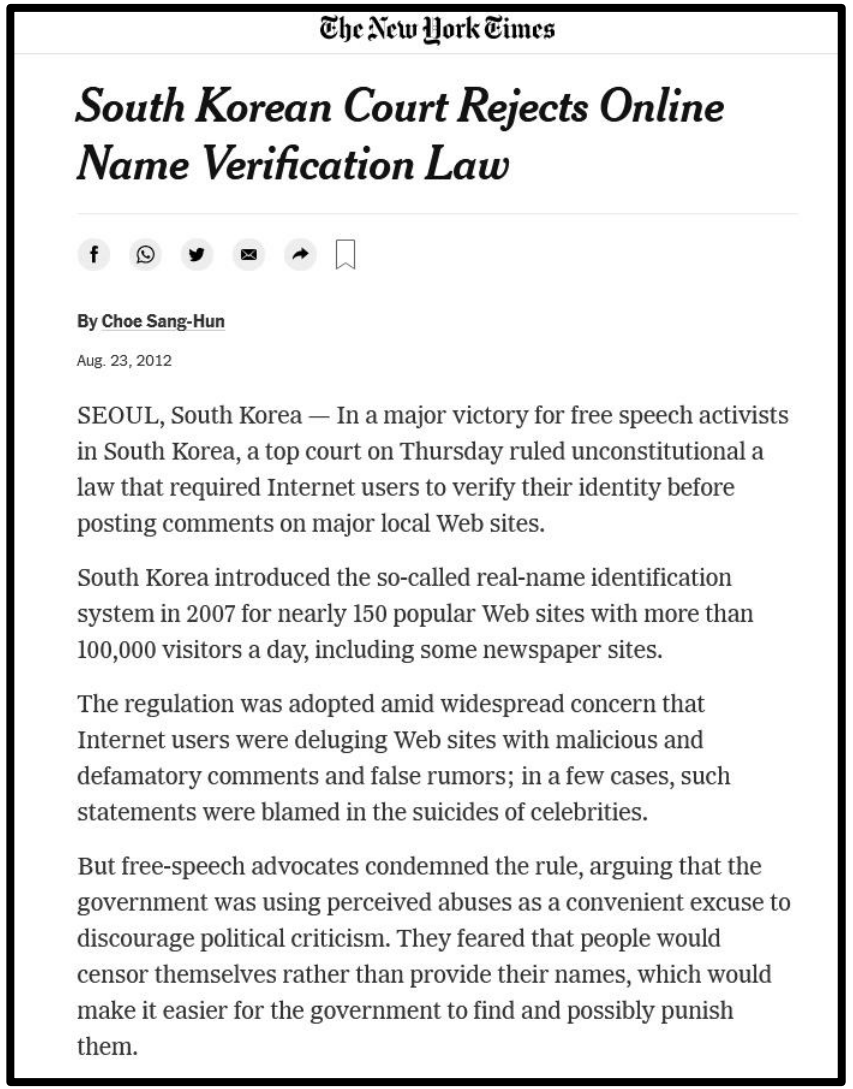
Sybil Attacks

- One individual creates many pseudonymous identities
- For instance, one individual creates many accounts
- Namesake: Sybil (pseudonym of a person who had a dissociative identity disorder)
- Also called: sock puppets (false identities)
- Why is this a problem for computer systems?



Tie accounts to real identities

- IP address
- Mailing address
- National identity card
- Telephone number
  - What precise protocol?



The New York Times

South Korean Court Rejects Online Name Verification Law

By Choe Sang-Hun

Aug. 23, 2012

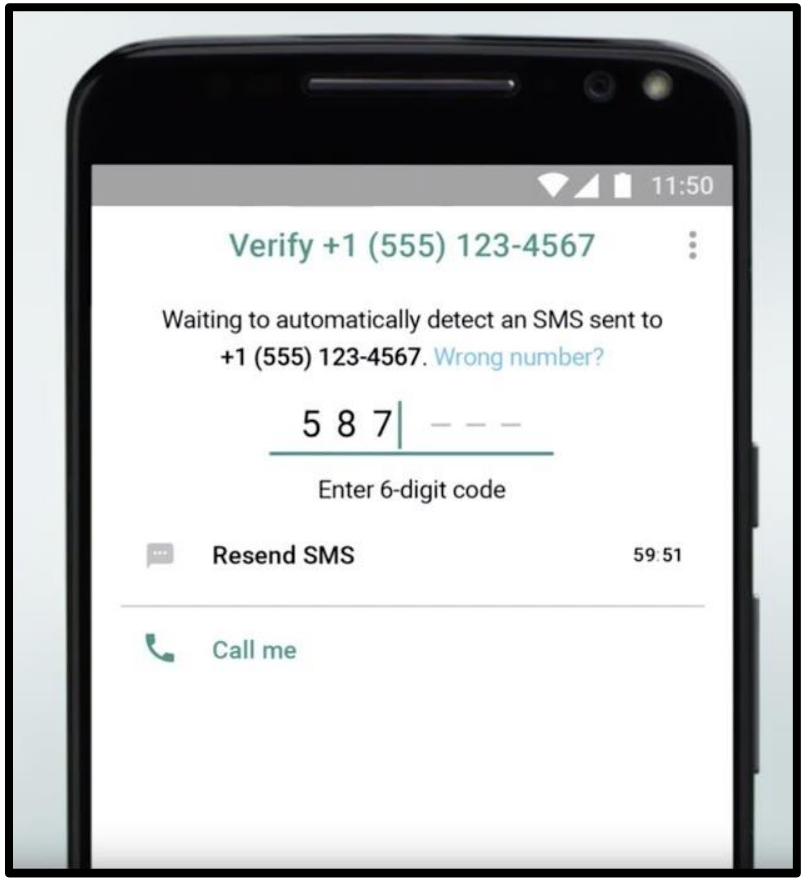
SEOUL, South Korea — In a major victory for free speech activists in South Korea, a top court on Thursday ruled unconstitutional a law that required Internet users to verify their identity before posting comments on major local Web sites.

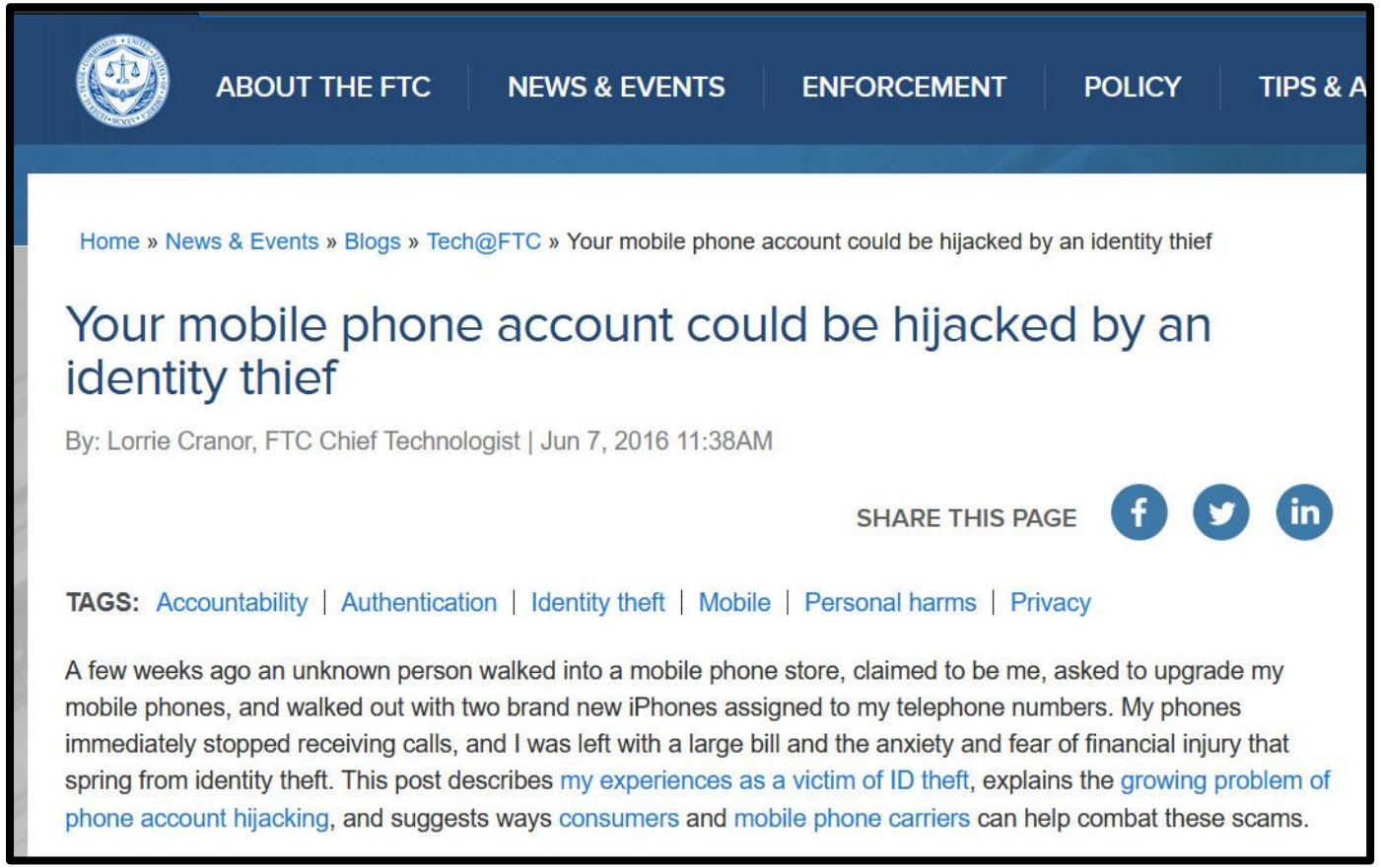
South Korea introduced the so-called real-name identification system in 2007 for nearly 150 popular Web sites with more than 100,000 visitors a day, including some newspaper sites.

The regulation was adopted amid widespread concern that Internet users were deluging Web sites with malicious and defamatory comments and false rumors; in a few cases, such statements were blamed in the suicides of celebrities.

But free-speech advocates condemned the rule, arguing that the government was using perceived abuses as a convenient excuse to discourage political criticism. They feared that people would censor themselves rather than provide their names, which would make it easier for the government to find and possibly punish them.

Vulnerabilities of SMS Codes





https://www.youtube.com/watch?v=AWemFbRf95g
https://www.ftc.gov/news-events/blogs/techftc/2016/06/your-mobile-phone-account-could-be-hijacked-identity-thief

Rely on real-world trust relationships

SybilGuard: Defending Against Sybil Attacks via Social Networks

Haifeng Yu, Michael Kaminsky, Phillip B. Gibbons
Intel Research Pittsburgh
{haifeng.yu,michael.e.kaminsky,phillip.b.gibbons}@intel.com

Abraham Flaxman
Carnegie Mellon University
abie@cmu.edu

ABSTRACT

Peer-to-peer and other decentralized, distributed systems are known to be particularly vulnerable to *sybil attacks*. In a sybil attack, a malicious user obtains multiple fake identities and pretends to be multiple, distinct nodes in the system. By controlling a large fraction of the nodes in the system, the malicious user is able to "out vote" the honest users in collaborative tasks such as Byzantine failure defenses. This paper presents *SybilGuard*, a novel protocol for limiting the corruptive influences of sybil attacks. Our protocol is based on the "social network" among user identities, where an edge between two identities indicates a human-established trust relationship. Malicious users can create many identities but few trust relationships. Thus, there is a disproportionately-small "cut" in the graph between the sybil nodes and the honest nodes. SybilGuard exploits this property to bound the number of identities a malicious user can create. We show the effectiveness of SybilGuard both analytically and experimentally.

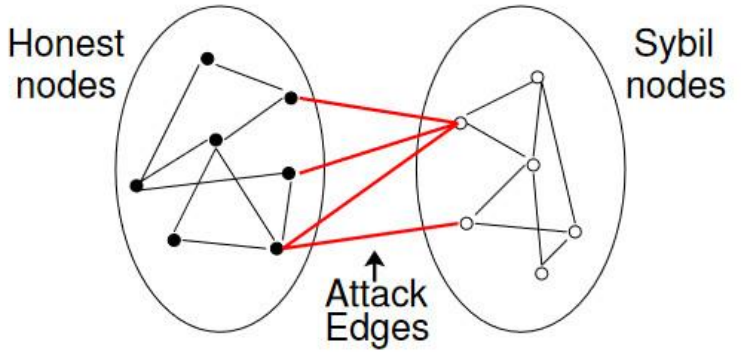


Figure 1: The social network with honest nodes and sybil nodes. Note that regardless of which nodes in the social network are sybil nodes, we can always "pull" these nodes to the right side to form the logical network in the figure.

Cybersecurity Law of the People's Republic of China (Effective June 1, 2017)

Article 24: Network operators handling network access and domain name registration services for users, handling stationary or mobile phone network access, or providing users with information publication or instant messaging services, shall require users to provide real identity information when signing agreements with users or confirming the provision of services. Where users do not provide real identity information, network operators must not provide them with relevant services.

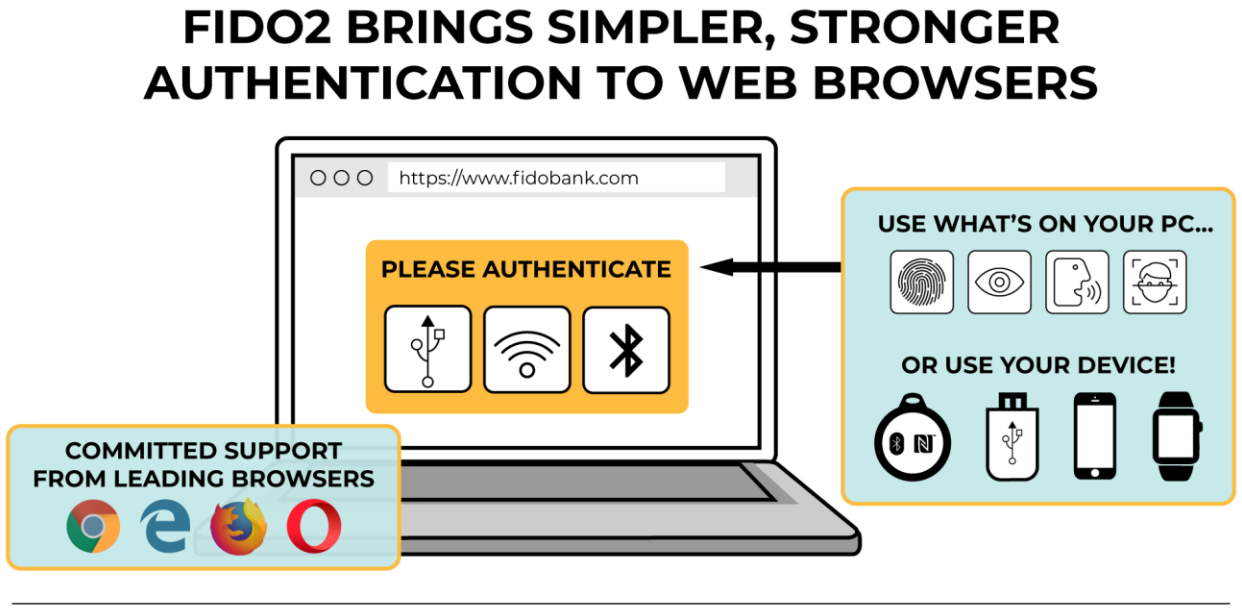
https://www.newamerica.org/cybersecurity-initiative/digichina/blog/translation-cybersecurity-law-peoples-republic-china/

National ID Cards

- Some national ID cards include a microprocessor
  - Online authentication becomes possible

https://www.thalesgroup.com/en/markets/digital-identity-and-security/government/identity/2016-national-id-card-trends

Authentication with Asymmetric Crypto



Proofs of Work

Prerequisite: Hashing

- One-way function
- Similar inputs result in very different outputs

Prerequisite: Hashing

- One-way function
- Similar inputs result in very different outputs
- md5(“blase”) = 12B872ADB2588C668D706D847FC1DA7E

Prerequisite: Hashing

- One-way function
- Similar inputs result in very different outputs
- md5(“blase”) = 12B872ADB2588C668D706D847FC1DA7E
- md5(“blasé”) = 29AFE9B75D98D3C4ECFCB34FDFC422A2

Detour: Intentionally Slow Hashing

- Key usage: password storage
- Iterated hash functions
  - Examples: bcrypt
  - Configurable number of iterations
  - Requires more computation; also reduces parallelism
- Memory-hard hash functions
  - Examples: scrypt, Argon2
  - Requires the person computing have a lot of RAM ($$$)

Need for Proofs of Work

- Example (problematic) system: You upload some data to a computer system and it trains a neural network with that data
- Example (problematic) system: You upload some data to a computer system and it trains a neural network with that data
- Example (problematic) system: You upload the product of two large prime numbers to a system and it factorizes it
- What’s the problem?

Need for Proofs of Work

- Example (problematic) system: You upload some data to a computer system and it trains a neural network with that data
- Example (problematic) system: You upload some data to a computer system and it trains a neural network with that data
- Example (problematic) system: You upload the product of two large prime numbers to a system and it factorizes it
- What's the problem? **Denial of Service (DoS) attacks**

Need for Proofs of Work

- Example (problematic) system: Everyone can vote on who wins the CS 259 Memelord award
- What’s the problem?

Need for Proofs of Work

- Example (problematic) system: Everyone can vote on who wins the CS 259 Memelord award
- What's the problem? **Does one person = one vote?**

Blockchain

- Blocks of transactions are linked together into a chain
- Hashes connect the blocks
- *Emergent consensus*: The hash chain representing the most cumulative work is considered valid
- Blocks (in Bitcoin) are mined every 10 minutes

https://www.oreilly.com/library/view/mastering-bitcoin/9781491902639/ch08.html

Blockchain Block Headers

Example 8-10. SHA256 output of a script for generating many hashes by iterating on a nonce

```
$ python hash_example.py

I am Satoshi Nakamoto0 => a80a81401765c8eddee25df36728d732...
I am Satoshi Nakamoto1 => f7bc9a6304a4647bb41241a677b5345f...
I am Satoshi Nakamoto2 => ea758a8134b115298a1583ffb80ae629...
I am Satoshi Nakamoto3 => bfa9779618ff072c903d773de30c99bd...
I am Satoshi Nakamoto4 => bce8564de9a83c18c31944a66bde992f...
I am Satoshi Nakamoto5 => eb362c3cf3479be0a97a20163589038e...
I am Satoshi Nakamoto6 => 4a2fd48e3be420d0d28e202360cfbaba...
I am Satoshi Nakamoto7 => 790b5a1349a5f2b909bf74d0d166b17a...
I am Satoshi Nakamoto8 => 702c45e5b15aa54b625d68dd947f1597...
I am Satoshi Nakamoto9 => 7007cf7dd40f5e933cd89fff5b791ff0...
I am Satoshi Nakamoto10 => c2f38c81992f4614206a21537bd634a...
I am Satoshi Nakamoto11 => 7045da6ed8a914690f087690e1e8d66...
I am Satoshi Nakamoto12 => 60f01db30c1a0d4cbce2b4b22e88b9b...
I am Satoshi Nakamoto13 => 0ebc56d59a34f5082aaef3d66b37a66...
I am Satoshi Nakamoto14 => 27ead1ca85da66981fd9da01a8c6816...
I am Satoshi Nakamoto15 => 394809fb809c5f83ce97ab554a2812c...
I am Satoshi Nakamoto16 => 8fa4992219df33f50834465d3047429...
I am Satoshi Nakamoto17 => dca9b8b4f8d8e1521fa4eaa46f4f0cd...
I am Satoshi Nakamoto18 => 9989a401b2a3a318b01e9ca9a22b0f3...
I am Satoshi Nakamoto19 => cda56022ecb5b67b2bc93a2d764e75f...
```

https://www.oreilly.com/library/view/mastering-bitcoin/9781491902639/ch08.html

Selecting the Winning Block

- “To make a challenge out of this algorithm, let’s set an arbitrary target: find a phrase that produces a hexadecimal hash that starts with a zero. Fortunately, this isn’t difficult! Example 8-10 shows that the phrase “I am Satoshi Nakamoto13” produces the hash 0ebc56d59a34f5082aaef3d66b37a661696c2b618e62432727216ba9531041a5, which fits our criteria. It took 13 attempts to find it. In terms of probabilities, if the output of the hash function is evenly distributed we would expect to find a result with a 0 as the hexadecimal prefix once every 16 hashes (one out of 16 hexadecimal digits 0 through F). In numerical terms, that means finding a hash value that is less than 0x1000000000000000000000000000000000000000000000000. We call this threshold the target and the goal is to find a hash that is numerically less than the target. If we decrease the target, the task of finding a hash that is less than the target becomes more and more difficult.”

Overall Process

- Validate blocks (e.g., no invalid transactions)
- Select the chain with the most proof of work

https://www.thalesgroup.com/en/markets/digital-identity-and-security/government/identity/2016-national-id-card-trends

Environmental Impacts

Electronic Waste

Bloomberg CityLab

The Toxic Effects of Electronic Waste in Accra, Ghana

Sorting through used electronics is a livelihood for many in the Agbogbloshie area, but toxic e-waste poses serious health risks.

Peter Yeung
May 29, 2019, 2:20 PM CDT

Abdrahaman Daouda came to Accra from Niger two years ago. He collects used water sachets and scrap metal, and hopes to buy his own taxi one day. But when it rains at Agbogbloshie, he finds it difficult to breathe. *Peter Yeung*

SHARE THIS ARTICLE
Share
Tweet
Post
Email

Heavy, acidic gusts of smoke billow across the Agbogbloshie dump, a wasteland dotted with burning mounds of trash in Ghana's capital, Accra.

Up to 10,000 workers wade through tons of discarded goods as part of an enormous, informal recycling process, in what has become one of the world's largest destinations for used electronic goods.

CNN World Africa Americas Asia Australia China Europe India Middle East United Kingdom LIVE TV Edition

MARKETPLACE
AFRICA

The rising e-waste crisis is being reckoned with in Rwanda, one gadget at a time

By Daniel Renjifo, CNN
Updated 1:21 PM ET, Fri February 26, 2021

How Rwanda is leading e-waste recycling efforts in Africa

How Flutterwave's unicorn status could sprout more innovation in African fintech

How international demand for Nigerian cotton is suiting well for small farmers

How big data is fostering expansion for this South African logistics enterprise

How ghost are coming South Afric

(CNN) — For Eric Nshimiyimanain, who owns two small electronic repair shops in Kigali, Rwanda, the startup chime of an old Windows laptop is the sound of a business opportunity.

He refurbishes broken PCs, laptops, phones and secondhand gadgets classified as electronic waste, or "e-waste" that would otherwise end up as trash in Nduba, Rwanda's only open-air dump in the outskirts of the capital.

https://www.bloomberg.com/news/articles/2019-05-29/the-rich-world-s-electronic-waste-dumped-in-ghana
https://www.smithsonianmag.com/science-nature/burning-truth-behind-e-waste-dump-africa-180957597/
https://www.cnn.com/2021/02/26/africa/marketplace-africa-ewaste-electronics-recycle-rwanda-spc-intl/index.html

Diurnal Patterns of Energy Usage

COMMUNICATIONS OF THE ACM

HOME | CURRENT ISSUE | NEWS | BLOGS | OPINION | RESEARCH | PRAC

Home / Magazine Archive / February 2021 (Vol. 64, No. 2) / Driving the Cloud to True Zero Carbon / Full Text

EDITOR'S LETTER

Driving the Cloud to True Zero Carbon

By Andrew A. Chien
Communications of the ACM, February 2021, Vol. 64 No. 2, Page 5
10.1145/3445037
Comments

VIEW AS: SHARE:

The right vision is to operate the cloud with zero-carbon emission from power (scope 2). Not just offsetting through renewable energy purchases. Not just 24x7 matching. True zero carbon in electric power consumed, and with no increase as the cloud continues to grow. That's the right vision for our proud computing technology community to lead the fight against climate change, and to see increasing use of computing as a positive force to slow climate change.[a,b]

Why must we act? The power grid is decarbonizing, but progress is slow. Aggressive states (for example, California and New York) have zero-carbon goals 20 or more years in the future, 2045 and 2040. Nationally, the U.S. produced 19% of its electric power from renewable resources (2020), and with "datacenter alley" reporting 12% renewables[c] (Northern Virginia). This trails the world's 26% renewables today, and U.S.

https://cacm.acm.org/magazines/2021/2/250064-driving-the-cloud-to-true-zero-carbon/fulltext