Doing History in the Internet Age

Steven M. Bellovin
http://www.cs.columbia.edu/~smb

Computers and History

- The history of computing is itself interesting
- More important, the existence of today's resources changes the way historians can work
- But – there are risks

Telegraph Codebooks

- Telegrams were expensive – 150 years ago, a trans-Atlantic message cost $5/word, *not* adjusted for inflation
- Use a single word or number to encode an entire phrase
- Primarily for economy and error detection/correction
- First telegraph codebook published in 1845; peaked in 1920s

A Typical Entry (1882)

Identity can be established if the party will answer that his or her mother's maiden name is..... **05626 Guineapig**

- Yes, mother's maiden name was used for authentication in 1882…
- Codebooks were often domain-specific; quality varied widely

The Federal Reserve Codebook (1921)

- Typical financial codebook
- Required external "test words" for authentication
- Did some of the codewords have a hidden message?

Hydrant	Sec. of Treasury
Hymen	Supt. of Banks
Hypnotism	U.S. Congress
Hypocrisy	The President
Hysterical	Treasury Dept.
Ignoramus	U.S. of America
Ignorant	U.S. Senator
Imbibing	War Department
Imbosom	White House

What About Confidentiality?

- Could use superencipherment for confidentiality, typically by modular addition of a secret value – the key – to the code number and then selecting the new code word
- Most schemes were pretty bad
- War Department Administrative Code (1899): "When a single key number is used, the number may be alternately added and subtracted. Other methods will readily occur. The use of 50 or 100, while easy to remember, should be avoided."

Doing Confidentiality Right

- Change the additive for each codeword
- Use random additives
- If the additives are random and *never* reused, this is a "one-time pad"
- One-time pads, if properly used (i.e., if the values are truly random and are not reused) are unbreakable even in theory

Indulging My Hobby…

- In January, I had a free day in Washington, DC
- I went to the Library of Congress to look some codebooks – they have many hundreds of them
- Being a security guy, I decided to look at a few that had the word "Secrecy" or some such in the title

TELEGRAPHIC CODE

TO INSURE

PRIVACY AND SECRECY

IN THE

TRANSMISSION OF TELEGRAMS.

15
9441 a

BY

FRANK MILLER.

NEW YORK:
CHARLES M. CORNWELL,
247 PEARL STREET.

Copyrighted in 1882, by FRANK MILLER, of Sacramento, California.

Miller's Superencipherment – What is This?

A banker in the West should prepare a list of irregular numbers, to be called "shift-numbers," such as 483, 281, 175, 892, &c.

The differences between such numbers *must not be regular*.

When a shift-number has been applied, or used, it must be erased from the list *and n·t used again*.

Under said number (which we will call the "serial-number") he will place the first "shift-number" (say 483). He will then add the two numbers and find their sum, which he will write down.

Underneath this new sum, or number, he will write the "cipher-word" which he shall find in the Code standing alongside of said sum.

This is a One-Time Pad!

The one-time pad is believed to have been invented in 1918 by Gilbert Vernam (Bell Telephone Laboratories) and Joseph Mauborgne (Head of U.S. Army Signal Corp Research and Engineering)

Copyrighted in 1882, by FRANK MILLER, of Sacramento, California.

1882? Did Vernam or Mauborgne know Miller?

Finding Frank Miller

We know little for sure:

- Book published in Sacramento in 1882
- Miller claims 16 years experience in banking
  - A post-Civil War job?
- There is a quote from an Army colonel about the need for encryption

That's all we know!

Google to the Rescue

- Search for "bank Sacramento ″Frank Miller″"
- The first page of results had a scanned copy of *History of Sacramento County* – and it listed the name of the bank for *some* Frank Miller
- The book also said there were only three banks in Sacramento then – so I'd probably found the right Frank Miller, just a few hours after I found the codebook
- That also gave me the name of the bank, for future queries

Confirmation

- Old census records are public – and they're online
- In 1880, there were only two Frank Millers in Sacramento; one was a "laboror" and the other worked at the proper bank
- A genealogy book described Miller's Civil War service; he did indeed return to Sacramento after the war and join this bank

Another Hit

- I might have found that history of Sacramento by conventional means – the university library and the NYPL have copies
- Google, however, found a genealogy book that gave a fair amount of detail about Miller
- I'd *never* have found that book by conventional means – it was "privately published", and Miller is far too common a name
  - >28K books about "Miller" at the NYPL…
- I also found – via Google – a scanned 1896 magazine that had a profile of Miller. Again, I would never have looked at *Overland Monthy and Out West Magazine*

Correspondence

- Via email, I was able to contact the world's foremost historian of cryptology, as well as an NSA historian (whose contact info I found on the web)
- Conventional correspondence would have worked for that, but much more slowly
- They'd never heard of Miller's codebook (good!)
- The NSA historian mentioned that she was working on a biography of Parker Hitt

Parker Hitt

- Parker Hitt, also an Army officer, was a friend and colleague of Mauborgne, and another pioneering cryptologist
- The historian had his service record, which showed that Hitt was stationed in San Francisco 1906-1907
- Mauborgne's service record showed that he wasn't stationed there until well after Miller had died – but it also showed that he sailed to the Philippines in 1913. From San Francisco?
- Might Hitt or Mauborgne have met Miller? Did officers socialize much with civilians?

Newspaper Archives

- A military historian suggested looking at the society pages of a San Francisco newspaper
- Once upon a time: look at blurry microfilms for about 40 years of daily papers, searching for something that might not even be there
- Today, though, newspaper archives have been scanned in – and you can do full-text searches

Miller's History

- Searching for Frank Miller showed that he moved from Sacramento to San Francisco after he retired from the bank in 1904
- His name showed up frequently in the society pages – but never in the same article as Hitt or Mauborgne
- But – OCR of blurry microfilms is imperfect; I had to broaden my search

Social Events

- Looking for "22nd Infantry" – Hitt's regiment – gave me the phrase "military ball"
- Looking for "military ball" and "Frank Miller" got a hit, in 1907
- Miller and Hitt were both listed as attendees; it was sponsored by the "bachelor officers" (including Hitt), and Miller and his wife were chaperoning their very eligible daughter
- His daughter did marry one of Hitt's colleagues just six months later, which strongly suggests that (a) she was looking, and (b) Hitt almost certainly spoke with her father
- (There was also a large military ball while Mauborgne was passing through town, but neither he nor Miller were listed as attendees)

Did Hitt Know?

- I regard it as likely, but not certain, that Miller said something about his book
- However, it was likely a brief, and perhaps not very comprehensible, explanation
- Hitt was the first to realize that for security (for a particular cipher), the key should be as long as the plaintext
- Did Hitt have some subconscious recollection of that conversation? (He was punctilious about assigning credit to others, so it was probably not a conscious memory.)

More on Miller?

- Miller's diaries and papers still existed in 1987 – a grandson had them
- From the data in the genealogy book and assorted online queries, I was able to locate a great-granddaughter
- Unfortunately, I have not been able to get in touch with her

The Codebook

- The codebook used the phrase "test word" for what today we would call an "authenticator". I'd never heard of this phrase.
- Google again – it showed a little-known 1876 codebook by Slater that used it. (There are no copies of this codebook in the US – but Google had scanned Oxford University's copy.)
- Further queries showed that test words had become very common by 1917
- The phrase was used to prove membership in secret social societies, such as the Freemasons (1829), the Templars (1867), and the Ku Klux Klan (1872)
- Neither of these two meanings is listed in the OED – but full-text searches found them

Test Words

- Google also found the phrase in a 2002 book about Wells Fargo – their private codebooks date to 1874
- Correspondence with Wells Fargo corporate historians showed that the company was using preposterously strong cryptography by 1877, *far* better than anything the US military used until World War I
- No cryptologic historians had the faintest idea of this – no one would have thought to look there; Wells Fargo dropped their own codebook around 1915

Slater's Scheme

- Slater's scheme also needed what is known today as a "checksum", to guard against errors (intentional or accidental)
- (No technical historian has ever mentioned this)
- Further queries showed that checksums were invented around 1881, to correct errors in astronomical telegrams
- But what cryptologic or telegraph historian would look at the Proceedings of the Royal Astronomical Society?

Doing History

Technology gave me four major advantages:

- It speeded up correspondence
- It allowed easy access to inconvenient resources
- It made infeasible searches easy
- It permitted full-text queries, which allowed me to find sources to which there was no rational pointer

There are Dangers

Only works from a narrow period have been scanned

- Mostly English, mostly from before 1923
- Very few archives – primary source material for serious historians – have been scanned
- Scanning is expensive

Cost

- I do not know of any east coast library that has a subscription to the electronic San Francisco Chronicle – the university library got a trial subscription for me
- Copyright issues have prevented Google from scanning most books from later than 1923
- The risk here is the chance of selective focus on certain time periods – and of ignoring physical resources if they don't exist electronically

Conclusions

- The Internet and other aspects of technology will fundamentally change the way historians work
- If we're not careful, the change will be selective, and might have bad side-effects
- But when it's applicable, it lets people find things they can't find any other way