

Code Breaking: From Latin to Web Security

Kevin Fu

MIT Lab for Computer Science

Harvard Latin B Class

a.d. XII Kal. Mai. MIMI

(Friday, April 20, 2001)

Why study the grammar of ancient languages?

- **Discover how cultures evolve, where people originate**
- **Develop pattern matching skills**
- **Learn how to interpret encrypted text**

Ancient languages and cryptography

- **Anticryptography: Understanding the meaning of ancient texts**
- **Cryptography: Making messages difficult for an adversary to understand**

Linguistic problems in decipherment

- **Known script, known language [Americans reading British English]**
- **Unknown script, known language [Simple substitution, cryptogram]**
- **Known script, unknown language [American can read Italian, but might not know the meaning]**
- **Unknown script, unknown language [Writing from lost civilizations]**

Understanding the script

- Ideographic: Semantic symbols represent ideas
[Hieroglyphic, Chinese, Linear B]
- Phonetic: Symbols to represent sounds
[Hieroglyphic]
- Syllabic: [Linear B, in-di-vi-du-all]
- Alphabetic: [Greek, Latin, English]

Understanding the language

- Grammar, inflections to find patterns
- Frequency analysis of characters and groupings
- Erasures indicate relationships between characters
- Bilingual cheat sheets – Rosetta Stone
- Names of leaders, city names

Case studies

- **Deciphering the Phaistos Disk**
- **Deciphering messages of opium smugglers**
- **Reverse-engineering the Wall Street Journal security system**

The Phaistos Disk

- **Guessing grammars, inflections, and using patterns**
- **Constraint propagation**
- **Similar methods used against Linear B**

Opium smugglers

- Elizabeth Smith Friedman, court cryptographer
- Deciphered encrypted communication between opium smugglers
- Had no knowledge of Chinese, the underlying language
- Nevertheless able to determine the meaning

The grammar of Web insecurity

- The meaning of cryptographic values set by Web sites
- Deduce what each number means and how the numbers relate in a mini-grammar
- How to encode in a non-forgable way: With this token, Kevin has permission to read articles
- Logging into and authenticating to a Web site

WSJ.com background

- **Wanted to authenticate paid subscribers**
- **Half million paid-subscriber accounts**
- **Can purchase articles. Optional stock portfolio tracking**

File Edit View Go Communicator

News Downloads Software Hardware Developers Help Search Shop

Bookmarks Netsite: http://public.wsj.com/home.html

Back Forward Reload Home Search Netscape Print Security Shop Stop



THE WALL STREET JOURNAL.

U.S. View

Other Views:
ASIA EUROPE

Free U.S. Quotes
Enter Symbol Here

WSJ.com Subscribers
Go Directly To:

Select a Page

Or LOG IN

WSJ.COM SUBSCRIBERS ONLY

Top Business News

- Davis Says California Has Deal With Utility
- Employers Plan Slight Scaling Back

100% of 7K (at 227 bytes/sec)



The server interactive.wsj.com wishes to set a cookie that will be sent to any server in the domain .wsj.com. The name and value of the cookie are: fastlogin=



This cookie will persist until Sun Feb 25 07:26:53 2001
Do you wish to allow the cookie to be set?

OK

Cancel

The process of deciphering WSJ.com

- Obtaining sample encodings
- Detecting patterns
- Guessing the algorithm and inputs which create an encoding
- Testing if encodings from a reverse-engineered algorithm equal the real thing
- Extract secrets from the Web site

WSJ.com analysis: the `crypt()` hash function

- **Takes an 8-character input**
- **Ignores all input after the 8th character**
- **Given only the output of `crypt`, it is computationally infeasible to determine the original input**

Wsj.com analysis continued

- **fastlogin =**
user + crypt (user + rotating server secret)

- **Using your fastlogin cookie to produce another:**

username	Crypt() Output	Fastlogin Cookie
bitdiddle	MaRdw2J1h6Lfc	bitdiddleMaRdw2J1h6Lfc
bitdiddler	MaRdw2J1h6Lfc	bitdiddlerMaRdw2J1h6Lfc

How did we obtain the rotating server secret?

- Adaptive chosen plaintext attack (dynamic programming)
- Program queried WSJ with invalid cookies
- Runs in max 128×8 queries rather than intended 128^8 (1024 vs. 72057594037927936)
- 1 sec/query yields 17 minutes vs. 10^9 years
- The key is “March20”.

How our attack works

Pad guess	username	crypt input	worked?
	bitdiddl	bitdiddl	Yes
A	bitdidd	bitdiddA	No
...
M	bitdidd	bitdiddM	Yes
MA	bitdid	bitdidMA	No
...
Ma	bitdid	bitdidMa	Yes
...
March20	b	bMarch20	Yes

Dow Jones Response

“ ... about the factors affecting design decisions, it is certainly result of **time to market** considerations. ... we simply **didn't have clear security requirements** defined within the group and outside the group. So, we did what worked. We tried a better encryption algorithm, but hit a bug that we couldn't fix, so we implemented one that worked even though the architect in charge was fully aware of its short-comings. You must understand that I'm giving you my read on the situation since **I've joined WSJ.com just 5 weeks ago.** ”

—Javeh Saleh

Vice President, Technology

Interactive Business Technology Services, WSJ.com

Recommended Reading

- David Kahn. "The Code Breakers"
- F. L. Bauer. "Decrypted Secrets: Methods and Maxims of Cryptology"
- John Chadwick. "The Decipherment of Linear B"
- E. A. Wallis Budge. "The Rosetta Stone"
- Steven Fischer. "Glyphbreaker"
- Hinsley and Stripp. "Codebreakers: The Inside Story of Bletchly Park"