Computer Security & Privacy

slides adopted from F. Monrose

Why Computer Security Matters

- Computers/Internet play a vital role in our daily lives
- Social Networks and Online Communities
  - facebook, flickr, file sharing, etc.
- Privacy threats abound (identity fraud, etc.)
  - Online activities can be easily tracked
- Open (anonymous) communication
  - Tunisia, China, etc.
- Secure Web Transactions
  - Shopping, medical records, human resources, etc.
- Multi-disciplinary solutions
  - e.g., Forensics covers Ethics, Law, Policy, Technology,...

Computer Security

- Security is often advertised in the abstract
  - "The system is secure"
  - "Our product makes your networks secure"
  - "Your Internet transactions are secure"

- For security professionals, the key questions we should ask are
  - secure from *whom*?
  - secure from *what*?

Computer Security

- Understanding how to assess the "security" of a system requires that we understand
  - what *assumptions* are being made
  - what a particular security technology does (or does *not* do)
  - what design decisions (*conscious or not*) were made about attacks it was designed to prevent
  - what security metrics were applied
  - what types of threats it ignores
  - . . .

Computer Security & Privacy

- Security is never black or white; *context* matters more than technology
  - different security technologies play important roles in an *overall* security solution
  - it might be secure against a certain *type* of adversary (the average criminal vs a national intelligence agency)
  - it might be secure as long as certain advances don't occur, or for a certain period of time
- "Secure" is meaningless without context

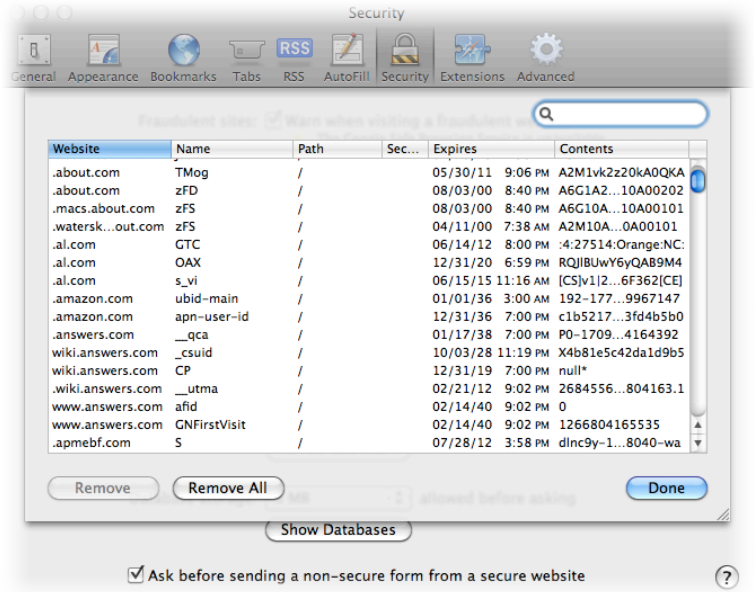
The unchanging landscape

- Cyberspace isn't all that different from the physical world
  - people interact with each other, form complex social relationships, have communities, both large and small
  - it is filled with commerce
- Threats in digital world mirror "real" world
  - theft, racketeering, vandalism, exploitation, con games, fraud, etc.
- Attacks will be similar to that in physical world

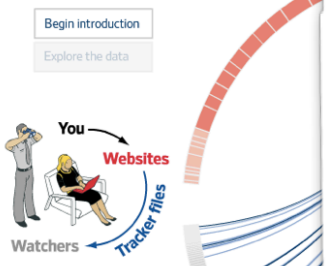
The unchanging landscape

- Where there is money, there are criminals
- Privacy violations aren't new either
  - lots of legal paperwork is public record (e.g., real estate transactions, criminal judgments)
  - private investigators use such data routinely to track down individuals; marketers use it to target particular demographics
- Privacy violations can *easily* lead to fraud
- Some "violations" are difficult to detect

Example: Cookie Tracking



Privacy on the



Site	Exposure Index	Trackers
dictionary.com	Very High	234
merriam-webster.com	High	131
comcast.net	High	151
careerbuilder.com	High	118
photobucket.com	High	127
msn.com	High	207
answers.com	Medium	120
yp.com	Medium	89
msnbc.com	Medium	117
yahoo.com	Medium	106
aol.com	Medium	133
wiki.answers.com	Medium	72
cnn.com	Medium	72
about.com	Medium	83
cnet.com	Medium	81
verizonwireless.com	Medium	90
imdb.com	Medium	55
live.com	Medium	115
att.com	Medium	58
walmart.com	Medium	66
bbc.co.uk	Medium	45
ebay.com	Medium	42
ehow.com	Medium	55
amazon.com	Medium	38
espn.com	Medium	61
myspace.com	Medium	108

See "What They Know" series at http://blogs.wsj.com/wtk/

How to control one's online privacy?

- Opt out, somehow?
  - Regularly check and delete cookies?
  - use "private browsing"
- Use 3rd party add-ons (e.g., TrackerScan)
- Advocate for do-not-track regulation?

Controlling one's online footprint is more complicated than it needs to be ... primarily because entire new industries for selling users' online information are springing up g

Measuring “Security”

- Lets look at three cases:
  - password authentication
  - intrusion detection systems
  - cryptography (break the cryptogram!)

On Measurements ...

E.g., 1: Password Authentication

- Passwords are a widely used user authentication method

- Authenticates ID of user logging and
  - that the user is authorized to access system
  - determines the user's privileges
  - used in discretionary access control

Password generation advice?

Password generation advice?

- don't use words in a dictionary
- composition matters (e.g., digits, special characters)
- choose mnemonic-based passwords which are memorable
- size matters (longer passwords are better)
  - *how many 12 char passwords do you have?*
- don't write it down
- don't share it with anyone
- expire frequently (like Onyen); change it often
- don't re-use.
  - *how many website passwords do you have?*
- ….

What makes a good password?

- Password length?
  - 64 bits of randomness is hard to crack
  - 64 bits is ~20 "common" ascii characters
  - but, people can't remember random strings!
- Pass phrases?
  - English text has roughly 1.3 random bits/char
  - so 50 letters of English text
  - hard to type without making mistakes!
- In practice
  - non-dictionary, mixed case, mixed alphanumeric

Measuring password strength

Password space

- number of *n*-character passwords given *c* choices per character is c^n
  - usually expressed as base-2 logarithm

$\rightarrow c$ $\downarrow n$	26 (lowercase)	36 (lowercase alphanumeric)	62 (mixed case alphanumeric)	95 (keyboard characters)
5	23.5	25.9	29.8	32.9
6	28.2	31.0	35.7	39.4
7	32.9	36.2	41.7	46.0
8	37.6	41.4	47.6	52.6
9	42.3	46.5	53.6	59.1
10	47.0	51.7	59.5	65.7

This is great, right?

Password space

- Time required to search: $T = c^n \cdot t \cdot y$
  - t = number of times password mapping is iterated
    - e.g., t = 25 on unix systems
  - y the time per iteration (e.g., y = 1/125000 sec)

$\rightarrow c$ $\downarrow n$	26 (lowercase)	36 (lowercase alphanumeric)	62 (mixed case alphanumeric)	95 (keyboard characters)
5	0.67 hr	3.4 hr	51 hr	430 hr
6	17 hr	120 hr	130 dy	4.7 yr
7	19 dy	180 dy	22 yr	440 yr
8	1.3 yr	18 yr	1400 yr	42000 yr
9	34 yr	640 yr	86000 yr	4.0×10^6 yr
10	890 yr	23000 yr	5.3×10^6 yr	3.8×10^8 yr

That's odd. Why then is password cracking still so successful?

Password space: a closer look

- The choices within the space are *not* equiprobable as user-selected passwords
- Most users selected passwords from a *small* subset of the full password space
  - many of which can be uncovered by trying words from a list (so-called dictionary attack)
- Implication ⇒ *Exhaustive* search as a metric for security is misleading here

- UNC Onyen – "Only Name You'll Ever Need"
  - Broadly used by UNC faculty, staff, students, and employees of UNC hospitals
  - Widely used at UNC for private services such as email, access to payroll management, etc.

Transform Sets Considered

Common Belief/Assumption: Enforcing password *expiration* is helpful from a security standpoint

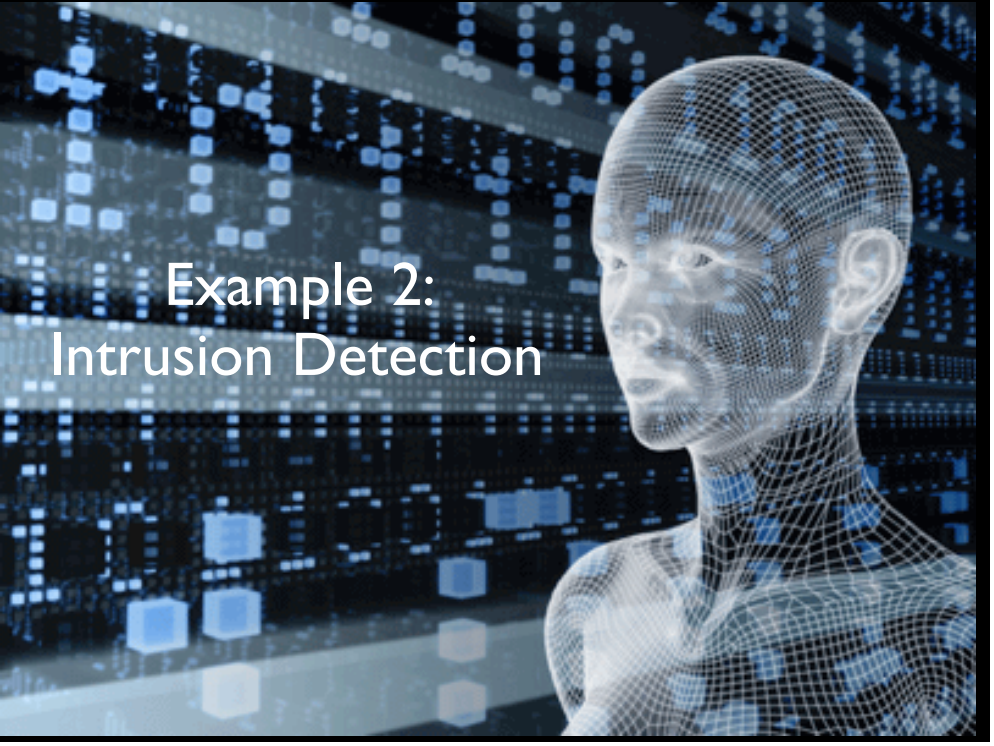
Transform set	Comments
Edit Distance	*password → p!assword*
Edit Distance with Substring Moves	*password → wordpass*
Location Independent Transforms	*Hand crafted, 8 subsets, only 534 primitive transforms*
Pruned Location Independent Transforms	*Top 50 transforms of location independent transforms*

Location Independent Transforms

CATEGORY	EXAMPLE
Capitalization	tarheels#1 → tArheels#1
Deletion	tarheels#1 → tarheels1
Duplication	tarheels#1 → tarheels#11
Substitution	tarheels#1 → tarheels#2
Insertion	tarheels#1 → tarheels#12
Leet Transform	tarheels#1 → t@rheels#1
Block Move	tarheels#1 → #tarheels1
Keyboard Transform	tarheels#1 → tarheels#!

Did we crack any passwords?

- **51141 hashes from 10374 defunct Onyen accounts**
  - **4** to **15** hashes per account in temporal order
  - Hashes are provided *without* plaintext passwords
- **After 8 months, 31074 hashes (60.8%) were cracked for 7936 Onyen accounts (76.5%)**
- **Learn from history**
  - History of transform is *strong* predictor of future use
  - Given old password, 40% of future passwords cracked in under 3 secs!



Terminology

- Virus: code that replicates a possibly evolved copy of itself.

- Worms: network viruses, primarily replicated on computer networks.
  - typically executes itself *automatically* on a remote machine with user intervention.
  - (mass mailer worms are an exception)

Terminology

- Trojan horses: typically try and interest the user with some useful functionality to entice the user to run a program.

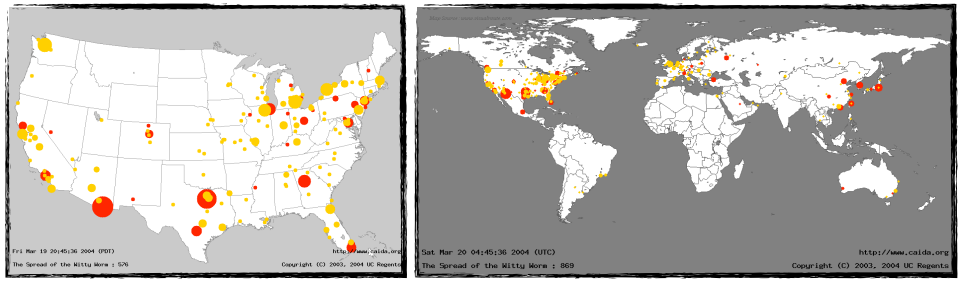
- These malicious software are called malware

Propagation strategies

- Hit-list and/or topological scanning
- Social-engineering
- Web-based malware (drive-by downloads)
- Exploiting social-networks (e.g. KoobFace botnet)
- Malicious documents (flash, pdf, etc)
- . .

Propagation

- Slammer worm (2003): doubling time of ~8.5 seconds. Peaked at ~3mins
  - >55 million IP scans/sec
- 90% of internet scanned in <10 mins



Worm Detection

- Signature inference: automatically learn the content "signature" for a new outbreak
- Example:
  - monitor network and look for strings common to traffic with "worm-like" behavior
  - Build signatures that can then be used for content filtering

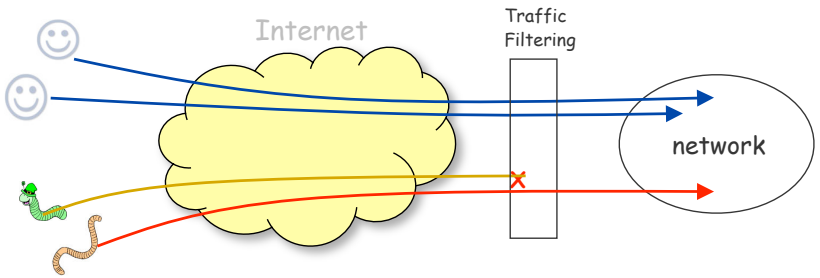
Signature: A payload content string specific to some malware

Example: Content Sifting

- Assume there exists some relatively unique invariant bitstring W across all instances of a particular worm
- Two consequences:
  - Content Prevalence: W will be more common in traffic than other bitstrings of same length
  - Address Dispersion: the set of packets containing W will address a disproportionate number of sources and destinations
- Content Sifting: find Ws with high content prevalence and high address dispersion ⇒ drop traffic

Content-based Blocking

Signature for CodeRed II



- Can be used by intrusion detection systems

Evaluation

- Standard measures:
  - Detection rate: ratio between the number of correctly detected attacks and the total number of attacks
  - False alarm (false positive) rate: ratio between the number of normal connections that are incorrectly misclassified as attacks and the total number of normal connections
  - . . .

Evaluation

- I: intrusive behavior, ¬I: non-intrusive behavior
  A: alarm, ¬A: no alarm

- Detection rate (true positive rate): P(A|I)
- False alarm rate: P(A|¬I)

Detection Rate vs False Alarm Rate

- Suppose 1% of traffic is attack traffic; Detector accuracy is 90%
  - i.e., classifies a valid connection as attack with prob. 10%
- What is the probability that a connection flagged by the detector as an attack is actually valid?

$$\Pr(\text{valid} \mid \text{alarm}) = \frac{\Pr(\text{alarm} \mid \text{valid}) \cdot \Pr(\text{valid})}{\Pr(\text{alarm})}$$

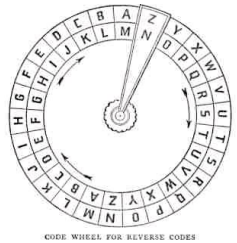
$$= \frac{\Pr(\text{alarm} \mid \text{valid}) \cdot \Pr(\text{valid})}{\Pr(\text{alarm} \mid \text{valid}) \cdot \Pr(\text{valid}) + \Pr(\text{alarm} \mid \text{Attack}) \cdot \Pr(\text{Attack})}$$

$$= \frac{0.10 * 0.99}{0.10 * 0.99 + 0.90 * 0.01}$$

= 92% chance that raised alarm is false positive!

Example 3: Cryptography

- The study of secret (crypto-) writing (-graphy)
- Concerned with developing algorithms:
  - that conceal the context of some message from all except the intended parties (privacy or secrecy)
  - that verify the correctness of a message to the recipient (authentication)



Classical cryptography

- Ancient ciphers
  - have a history of at least 4000 years
  - ancient Egyptians enciphered some of their hieroglyphic writings on monuments
  - ancient Hebrews enciphered certain words in scriptures
  - over 2000 years ago Julius Caesar purportedly used a simple substitution cipher
  - English Philosopher Roger Bacon described several methods in 1200s
  - . . .

Basic concepts

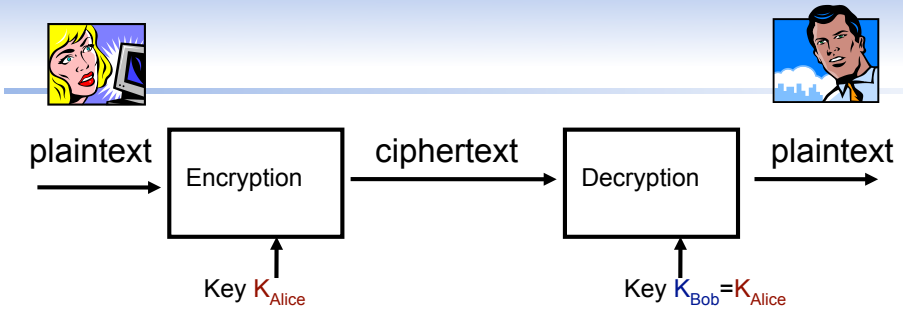
- Plaintext
  - the original intelligible message
- Ciphertext
  - the transformed message
- Cipher
  - an **algorithm** for transforming an intelligible message into unintelligible by transposition and/or substitution
- Key
  - critical information used by the cipher, known *only* to the sender and receiver

Basic concepts

- Encryption (encipher)
  - the process of converting plaintext to ciphertext
- Decryption (decipher)
  - the process of converting ciphertext back to plaintext

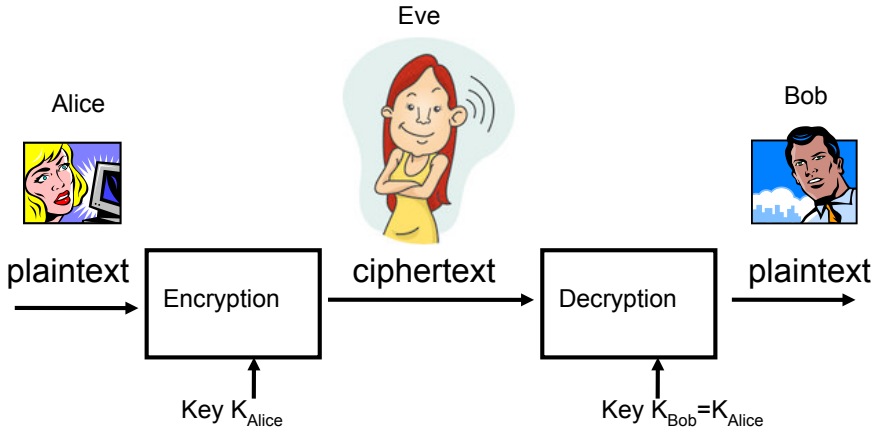
Basic concepts

- Cryptanalysis
  - the study of principles and methods of transforming an unintelligible message back into an intelligible message **without** knowledge of the key



- If both keys are the same, we have a symmetric cryptosystem
  - e.g., Data Encryption Standard
- If one key is inverse of the other, we have an asymmetric cryptosystem
  - e.g., public-key cryptography

How do we analyze crypto systems?



How do we analyze crypto systems?

- High-level: if the adversary intercepts the ciphertext, s/he cannot recover plaintext
- Issues in making this precise
  - What might your enemy know?
    - The kind of encryption function you are using?
    - Old plaintext-ciphertext pairs?
    - Information about how you chose keys?
- What does “cannot recover plaintext” mean?

On recovering plaintext

Natural language is highly redundant:

> Aoccdrnig to rscheearch at Cmabrigde Uinervtisy, it deosn't mttaer in waht oredr the ltteers in a wrod are, the olny iprmoetnt tihng is taht the frist and lsat ltteer be at the rghit pclae. The rset can be a toatl mses and you can sitll raed it wouthit a porbelm. Tihs is bcuseae the huamn mnid deos not raed ervey lteter by istlef, but the wrod as a wlohe.

Example: Classical techniques

- Two basic components
  - *substitutions*: letters replaced by other letters
  - *transposition*: letters rearranged in different order
- These ciphers may be:
  - *monoalphabetic*: only one substitution / transposition
  - *polyalphabetic*: several substitutions / transpositions used
- Product cipher
  - several ciphers concatenated together

Simple Substitution Cipher

- Let E(k,m) be a permutation of the alphabet

a	b	c	d	e	f	g	h	i	j	k	l	m
0	1	2	3	4	5	6	7	8	9	10	11	12
21	12	25	17	24	23	19	15	22	13	18	3	9

n	o	p	q	r	s	t	u	v	w	x	y	z
13	14	15	16	17	18	19	20	21	22	23	24	25
5	10	2	8	16	11	14	7	1	4	20	0	6

- plaintext: proceed meeting as agreed
- ciphertext: cqkzyyr jyyowft vl vtqyyr

Simple Substitution Cipher

- D(k,c) is given by reversing table

a	b	c	d	e	f	g	h	i	j	k	l	m
0	1	2	3	4	5	6	7	8	9	10	11	12
24	21	15	11	22	13	25	20	16	12	14	18	1

n	o	p	q	r	s	t	u	v	w	x	y	z
13	14	15	16	17	18	19	20	21	22	23	24	25
9	19	7	17	3	10	6	23	0	8	5	4	2

- ciphertext: cqkzyyr jyyowft vl vtqyyr
- plaintext: proceed meeting as agreed

Simple Substitution Cipher

- Here, a plaintext or ciphertext message is a single character
- Message space is size 26! > 4 x 10^{26}
  - But this cipher is very weak. Why?

Breaking the Code

- In English (and most languages) certain letters are used more often than others
- It would be a good guess that the letters that occur most often in the ciphertext are actually the most common English letters

Frequency Analysis

E	11.1	S	5.7	H	3.0	V	1.0
A	8.5	L	5.5	G	2.5	X	0.3
R	7.5	C	4.5	B	2.0	Z	0.3
I	7.5	U	3.6	F	1.8	J	0.2
O	7.1	D	3.3	Y	1.8	Q	0.2
T	7.0	P	3.2	W	1.3		
N	6.7	M	3.0	K	1.1		

Ciphertext only attack!

```
TCR KWR WGLC CPC VGQBC
VQVR BCT IXZ CFWGIUWCZ WYC
WCFW QN IJKUCR DCRRIOCR
```


Answer

```
Yes, it's true.  Eve broke
Bob's key and extracted the
text of Alice's messages.
```

To learn more ...

- Several courses offered to help you better understand these issues:
  - Computers & Society, Computer Networking, Operating Systems, Software Engineering, etc.
  - Intro to Computer Security, Network Security, Cryptography, Digital Forensics
- Stop by (3rd floor, this building)
  - We are always looking for undergrads (finance, math, linguistics, SILS, etc)
  - send email ({fabian, reiter}@cs.unc.edu)