

## CS 486 Midterm

Instructions: You will have 105 minutes to complete this midterm. The test is closed book; no notes, computers, or calculators are allowed. Please show all relevant work and calculations in order to receive partial credit (when appropriate). Many of the questions on this test are short answer, and are designed to get you to think about or explain the topics we have covered. You should answer these at an appropriate length and explain your reasoning when asked to do so. A one-word answer is usually too short; a one-page answer is usually too long.

### 1 Definitions (22 points: 2 points each)

In this section, you are provided with a list of terms and concepts. For each item in the list, give a concise (one or two sentence) explanation or description of that item.

**Example:** *One-time pad: A one-time pad is a means of symmetric-key encryption where a different key of the same length as the message is used for each message. It has the advantage of being theoretically unbreakable, since no frequency information about the key is revealed, but requires participants to securely distribute the secret keys, which is a non-trivial problem.*

- One-way function. **A one-way function is a function that is easy to compute, but whose inverse is computationally difficult to compute. For example, a hash function.**
- Certificate Authority **A Certificate Authority is the part of a Public Key Infrastructure that is responsible for distributing and revoking public keys. It serves to help ensure the secure distribution of public keys.**
- Zero knowledge protocols **A zero-knowledge protocol is a method by which two more more entities can perform a comparison or computation, such as verifying a bank balance, without either entity revealing any information to the other.**
- Inverted Index **An inverted index is a map from search terms to the web pages in which they appear. It is used by search engines to quickly find answers to search queries.**
- Stream Cipher **A stream cipher is a cipher in which each block is encrypted with both a key and with the previous block, either as plaintext or ciphertext.**
- Man-in-the-middle **A man-in-the-middle is an attacker who is able to listen to and potentially generate, capture, alter, or replay messages on a communication channel.**
- DES **DES (Data Encryption Standard) was the premier secret-key encryption method for many years. It uses a 56 bit secret key, along with several rounds of transposition and substitution, to ensure privacy**

- non-repudiation **This means that a receiver of a message or item is unable to falsely claim that the message was never received.**
- Core (with respect to Web structure) **The core is the subset of the Web in which there is a directed path from every page to any other page.**
- Web Services **Web Services provide a unified interface for a wide variety of distributed computations, typically accessed over HTTP, and using XML or HTML as a data wrapper. More sophisticated communication protocols such as SOAP can also be implemented on top of HTTP using Web Services.**
- dictionary attack **A brute-force attack in which all words in a lexicon are tried in order to break a password.**

## 2 Encryption (43 points)

**RSA:**

(5 points) When using RSA, we must first generate two keys, the public and private keys. If we pick two primes  $p=5$  and  $q=13$ , how do we then find  $d$  (the public key) and  $e$  (the private key)? Note: you do not need to find the actual values; an equation or precise description is fine.

Let  $N = pq$

$d$  is a small number relatively prime to  $(p-1)(q-1)$ .

$e$  is the multiplicative inverse of  $d$ , mod  $(p-1)(q-1)$ .

(3 points) Why is RSA considered secure? (What is the operation needed to 'break' RSA?)

**RSA's security depends on the conjecture that it is computationally difficult to factor large numbers. An attacker who can factor  $N$  can easily discover the private key ( $d$ ) from  $N$  and the public key ( $e$ ).**

Say that Alfred wants to send a message  $M$  to Betty. Alfred would like to make sure that only Betty can read  $M$ . In addition, Alfred would like to ensure that Betty can verify that the message  $M$  is from Alfred.

(5 points) 1. Describe (in order) the steps that Alfred must go through to send this message using public-key encryption. Be clear about which keys are used; I recommend using  $A_{pub}$  and  $A_{priv}$  (and  $B_{pub}$  and  $B_{priv}$ ) to indicate the keys.

**Alfred encrypts the email with Betty's public key  $B_{pub}$ , ensuring that only Betty (or someone else with her private key) can decrypt it. He then *signs* the email by encrypting it (or, alternatively, a hash of it), with his private key  $A_{priv}$ ; this proves that he (or someone with his private key) is the sender.**

(5 points) 2. Describe, in order, the steps Betty must take to read the message and verify that it is indeed from Alfred.

**Betty first verifies Alfred's identity by decrypting the message using Alfred's public key  $A_{pub}$ . She then decrypts the message itself using her private key  $B_{priv}$ . Alternately, if Alfred had signed a hash of the message, she would decrypt it with her private key, hash it, and compare the result to the message Alfred had signed.**

**Attacks:**

Consider the following authentication protocol:

Users Albert and Betty want to verify to each other that they are who they say they are. They have a shared secret key already established. Albert wants to be sure that Betty really possesses the shared secret, as does Betty. They are concerned that someone may be listening to their communications and storing past messages.

Albert generates a message  $m_1$  and sends it to Betty “in the clear”. Betty generates a message  $m_2$ . Betty encrypts  $m_1$  and  $m_2$  with the secret key and sends the encrypted message back to Albert. Albert then sends  $m_2$  back to Betty “in the clear”, proving that he was able to decrypt it.

$$\begin{aligned} A &\rightarrow m_1 \rightarrow B \\ B &\leftarrow E(m_1, m_2) \leftarrow A \\ A &\rightarrow m_2 \rightarrow B \end{aligned}$$

(5 points) Part a: Describe how an agent-in-the-middle could deceive Albert and Betty. (Hint: how could the agent deceive Albert into believing that he was Betty?)

**a man-in-the-middle could intercept the first message and send it back to Albert, pretending to be Betty initiating the same protocol. When Albert returned  $E(m_1, m_2)$ , the MITM would replay this back to Albert as Betty’s response to his original message. The MITM would then catch and replay the third message as well, fooling Albert into believing he was handshaking with Betty, when in fact she never receives a message.**

(5 points) Part b: How can the above protocol be corrected to prevent the attack in part a?

**There are many possible fixes. One way would be for the encrypted message to contain the identifier of the sender.**

### **Watermarking**

One way that vendors of digital information try to track and prevent piracy is through the use of *digital watermarks*. For the purposes of this question, let us assume that it is possible to embed a digital watermark in a document in such a way that the watermark cannot be removed without destroying the document, and illegitimate copies of a document will never contain a valid watermark.

(5 points) Part a: One use of digital watermarks is for usage tracking. If you had perfect, tamperproof watermarks, how could you use this to track unauthorized usage of a graphic image, such as a JPEG.

**You would use a web crawler to search for JPEGs. When one is found, the watermark is extracted and compared to a database of registered images and owners.**

Part b: Another use of digital watermarks is copy protection. If a vendor also has control of the hardware, it is possible to implement copy protection.

(3 points) Propose a means by which a vendor that had secure, tamperproof watermarking technology and control of the hardware on which the information is played (such as a DVD player) could ensure that only people who had legitimately purchased their product could view it. (In other words, pirated copies can't be viewed.)

**A vendor could include a key in the watermark. Players that cannot locate this key would refuse to play the DVD. This key would have to be unique and cryptographically signed to prevent tampering.**

(2 points) Why is it necessary for the vendor to have control of the hardware in order to ensure copy protection?

**If the vendor cannot control the playback mechanism, he can't prevent someone reverse-engineering a player that circumvents the watermark**

### Keyspaces

(5 points) Recall from Homework #1 that adding a linear number of bits to a DES key required an exponential increase in the time needed to decrypt a message encrypted with that key. For example, a 40-bit key was broken in 3.5 hours, while a 48-bit key required 313 hours.

Concisely explain why it is that this problem is exponentially hard. You should mention both the problem characteristics (what it is that is growing exponentially) and the method of ‘cracking’ the DES key being used.

**Each bit added to the key doubles the size of the keyspace. Since the only known way to crack DES is through brute force exponential search, the problem difficulty grows exponentially with key length.**

## 3 XML

A bookseller has decided that they want to make their book data available to customers over the Web using XML. Their representation of a book is as follows:

- A book has exactly one title, and 0 or more subtitles.
- A book has one or more authors.
- A book has an ISBN number, which is unique.
- A book has 0 or more reviews, which are character data.
- A book has 0 or more related books, each of which is indicated by an ISBN number.

(5 points) 1. Give an XML representation of the following book (there is more than one correct answer for this; I’m interested in seeing that you understand the syntax and structure of XML).

Assume that each of the lines here corresponds to a separate element, and that the ‘genre’ of computers is an attribute associated with the ‘book’ element.

Title: I Love Electronic Commerce (genre: computers)  
Subtitle: How e-commerce Will Change The World  
Authors: R.U. Sirius and M.T. Head  
ISBN: 0072227427  
Related Books: 1022101011, 0447701232  
Reviews: “It’s Wonderful!” by Anne Elk.

**There’s more than one way to skin this cat ...**

```
<book genre="computers">
<title>
  I Love Electronic Commerce
</title>
<authors>
  <author> RU Sirius </author>
  <author MT Head </author>
</authors>
<ISBN> 0072227427 </ISBN>
<related>
  <relbook> 1022101011 </relbook>
```

```

    <relbook> 0447701232 </relbook>
</related>
<reviews>
  <review>
    ‘‘It’s Wonderful!’’ by Anne Elk.
  </review>
</reviews>
</book>

```

(3 points) 2. Typically, an XML document would also be accompanied by a pointer to a DTD. What is the function of a DTD and why is it useful?

**A DTD specifies the legal structure of an XML document: what elements are allowed, what elements must be present, the order of elements, and so on. It serves to make sure that different designers use a uniform XML structure to represent their data.**

(2 points) 3. A key concept in parsing and understanding XML is the idea of a *namespace*. Many XML documents will use more than one namespace. Briefly explain what namespaces are and why they’re useful, particularly in open environments such as the Web. (Hint: What problem do they help avoid?)

**Namespaces provide scope for XML elements. They are essentially like packages in Java. They help avoid symbol clashing, where two different DTDs or documents have chosen the same name for two different elements.**

## 4 Short Answer (25 points - 5 points each)

For each of these questions, give a few sentences explaining and/or justifying your answer. In other words, tell me why in addition to what.

1. What are three challenges to creating a searchable index of the entire Web?

**1. Size. The web is big. 2. The Web is Dynamically changing. 3. Not all the Web is reachable by crawlers. Also, developing effective indexes, Effectively labeling and ranking pages**

2. What are two advantages to viewing a software component as an agent?

**1. Abstraction. It allows you, the programmer, to avoid worrying about lower-level details of a component. 2. Rationality. It allows system designers to treat your agent as if it had goals and desires and design the “rules of the system” to exploit this. There are lots of other answers too.**

3. A common way to try to increase a website’s ranking on different search engines is the insertion of spurious keywords into the page, Why will this not help improve one’s ranking with PageRank (Google)?

**PageRank does not use the keywords within a page. Instead, it uses the PageRank of a page’s inward links, so inserting keywords won’t help improve ranking.**

4. While SSL has become commonplace for transactions between customers and vendors, encryption of email is not as widespread. List two **technological** challenges to ubiquitous use of encrypted email.

**1. Encrypted email would require the existence of a ubiquitous public-key infrastructure. A sender would have to find the receiver’s public key *before* the email is sent. 2. There would also need to be backwards-compatibility support for mail clients that do not support encrypted email. There are other answers as well.**

5. Why does SSL use both symmetric and asymmetric keys? What are the functions of each key?

**The asymmetric, or public, key, is used to exchange the session, or symmetric key. the symmetric session key is then used for the actual communication.**