Massachusetts Institute of Technology
6.857: Network and Computer Security
Professors Srini Devadas and Nickolai Zeldovich

Handout 2
February 8, 2010
Due: February 19, 2010

Problem Set 1

This problem set is due *online,* at `https://sec.csail.mit.edu/` on *Friday, February 19* by **5PM**. Please note that no late submissions will be accepted.

You are to work on this problem set with your assigned group of three or four people. You should have received an email with your group assignment for this problem set. If not, please email `6.857-tas@mit.edu`. Be sure that all group members can explain the solutions. See Handout 1 (*Course Information*) for our policy on collaboration.

Homework must be submitted electronically! Each problem answer must appear on a separate page. Mark the top of each page with your group member names, the course number (6.857), the problem set number and question, and the date. We have provided templates for LaTeX and Microsoft Word on the course website (see the *Resources* page).

Grading: Problems 1 and 3 are worth 10 points each, and Problem 2 is worth 20 points.

With the authors' permission, we will distribute our favorite solution to each problem as the "official" solution—this is your chance to become famous! If you do not wish for your homework to be used as an official solution, or if you wish that it only be used anonymously, please note this in your profile on the homework submission website.

Problem 1-1. Security Policy

In recent years, there has been a lot of debate about what Internet service providers (ISPs) should be able to do with packets that traverse their networks. For the purposes of this problem, we will focus on the privacy of that data. There are many considerations about what data an ISP may want to collect. An ISP would find much of the data valuable: they may be able to better predict growth and build new capacity; record IP addresses of customers so they can bill them for bandwidth used; or use this data to target better advertisements to users. End-users could also benefit from their ISP examining their packets: for example, detecting worms, spam, or other malicious traffic and dropping it, or prioritizing voice-over-IP traffic over a BitTorrent download. An ISP's data might also be valuable to others: for example, ISPs may want help from each other in tracking down sources of attack, if the attacker's packets are traversing multiple networks; a law enforcement agency may want to know what user was using a particular IP address, or what web sites they visited; or a country may require filtering certain prohibited web sites. One flip side of data collection is, of course, potential loss of privacy and the risk of unintended disclosures or uses of this data: for example, can a single compromised law enforcement official learn everything about every US Internet user?

Your task is to write a short security policy for an ISP that addresses privacy and other concerns in data collection. Possible issues include:

- What types of data are sensitive? How long should it be stored? Who is ultimately allowed to view different types of data?
- Do customers of the ISP get a say in how long the ISP keeps their data? Can they find out what data the ISP is keeping? Can they find out who the ISP revealed their data to?
- What kinds of information can an ISP share with third parties, and in what situations? Should some information be revealed only if a customer's IP address is involved in a possible attack, and whose word should the ISP take for this?
- Can employees at the ISP examine traffic data? Is it OK to use some data in a program but not reveal it to employees? Keep in mind that some employees may need some or all of this information to perform their jobs.

For help on writing this policy you can see *Sample Solutions from PS1 2003*. See question 1-4, which asked students to develop a security policy for either the MIT Card or Apple's iPod. Sample solutions for both, as well as a short discussion from the TAs regarding common omissions, are included. These should help guide you in terms of content, format, and length. While many ISPs have security policies that answer many of these questions, you should feel free to come up with any security policy you feel is appropriate. Be creative.

Problem 1-2. Repeated Pad

The course staff decided to encrypt a 256-byte passage from a well-known cryptography paper with a one-time pad, but they were too lazy to come up with 256 bytes of truly-random pad bytes. After generating some number of random bytes, they gave up and just concatenated the bytes they had already generated several times to produce a sufficiently large pad. That is, given a random byte string R, they generated a pad $P = R \,||\, R \,||\, \ldots \,||\, R$, where $||$ is the concatenation operator.

Your job is to take advantage of the laziness of the course staff and determine the encrypted passage, along with the paper from which it came. We have provided the ciphertext in the Resources section of the course web site.

HINT: The original passage consists of ASCII text, talks about encryption, and you can find copies of this paper on Google.

Problem 1-3. Hashing

Let b denote a given "message block size" (e.g. $b = 512$ bits).

For this problem, assume all messages are exactly k blocks long, for some moderate k (e.g. $k = 1000$). Each message has length bk bits.

Let n denote a given desired hash output size, in bits (e.g. $n = 160$).

Let $Maps(t, u)$ denote the set of all possible functions with domain $\{0,1\}^t$ and range $\{0,1\}^u$. A randomly chosen function from $Maps(t, u)$ may be viewed as a "random oracle" (from t-bit strings to u-bit strings).

Ideally, a hash function should be indistinguishable from a random oracle with the same domain and range. However, in practice this may not be the case, due to the manner in which the hash function is constructed.

(a) Suppose f is a random oracle drawn from $Maps(bk, n)$. Suppose you draw values x_1, x_2, ... uniformly at random from $\{0,1\}^{bk}$, and for each such x_i you compute $f(x_i)$. How may such x's do you expect to have to try before you find a "collision" (a pair of distinct x_i, x_j values such that $f(x_i) = f(x_j)$? (No need for proof here. Also, your answer does not need to be exact, just a reasonable approximation.)

(b) Now suppose that hash function h mapping $\{0,1\}^{bk}$ to $\{0,1\}^n$ is constructed in a serial fashion from the random oracle g drawn from $Maps(b + n, n)$, as follows. To compute $h(M)$ where $M = m_1 \,||\, m_2 \,||\, \ldots \,||\, m_k$ (and each m_i is b-bits long):

- Let $v_0 = 0^n$.
- Let $v_i = g(v_{i-1} \,||\, m_i)$ for $i = 1, 2, \ldots, k$. The function g takes $n + b$ bits and hashes them down to n bits.
- Let $h(M) = v_k$.

Argue that an adversary can distinguish such a hash function h from a random oracle such as the one in Part (a) having the same domain and range, by looking for collisions in a certain way. Assume that the adversary can perform an arbitrary number of evaluations of h but cannot evaluate g.