SCIT-EIS-UOW
CSCI262/CSCI862 Spring 2021

Laboratory Week Five: Set Four - Solutions

Part One: Denial of Service

1. What is a zombie in the context of denial of service?

   A system that has been subverted, typically as part of a botnet,

   for use in launching an attack.

2. Find some examples of typical bandwidth or link capacities by searching on the Internet.

   The main point here is to get some idea of the relative

   size of the large DDoS attacks and the channel capacities.

3. What is the default size of a ping?

   On Windows systems the default size is 32 bytes.

   On Linux, Unix, and Macs, the default size is 56 bytes,

   64 bytes including 8 byte for header data.

4. The classical DoS flood attack overwhelms the bandwidth of a link, to effectively shut that link down. Consider that we use ICMP pings of size 500 bytes. Roughly how many packets do we need to send per second to flood links with the following capacities?

   (a) 0.25–Mbps.

   ```
   500 bytes is 4000 bits. The M and T are SI abbreviations
   for 10^6 and 10^{12}, respectively.
   ```

   $$\frac{0.25 * 10^6}{4000} = 62.5$$

   ```
   So 63 such packets would be needed.
   ```

   (b) 4–Mbps.

   ```
   16 times as many as the last --> 62.2*16=1000.
   So 1,001 such packets would be needed.
   ```

   (c) 20–Mbps.

   ```
   5 times as many as the last --> 1000*5=5000.
   So 5,001 such packets would be needed.
   ```

(d) 1.2–Tbps.

$$\frac{1.2 * 10^{12}}{4000} = 3 * 10^8$$

`That's 300,000,001 such packets a second.`

5. Of more concern than the DoS attack is the distributed DoS attack. Assume each captured system has an upload capacity of 128–kbps. How many such captured systems would be required to flood links with the following capacities?

   (a) 0.25–Mbps.

   ```
   128kbps is 128,000bps.
   This is about 1.95 systems, so 2.
   ```

   (b) 4–Mbps.

   ```
   31.25. 32 systems.
   ```

   (c) 20–Mbps.

   ```
   156.25. 157 systems.
   ```

   (d) 1.2–Tbps.

   ```
   9,375,000 systems.
   ```

6. What is DNS?

   ```
   Domain name system. Used for mapping URLs to IP addresses.
   ```

7. Describe DNS amplification.

   ```
   Using DNS servers as the intermediate for amplification.
   Typically 60 byte UDP request. Returns now up to about
   4000 byte UDP response. Typically an attack would involve
   using multiple different DNS servers.

   See later also.
   ```

   (a) What implication does it have for the resources required by an attacker?

   ```
   The attacker resources can be small relative to the load
   they can generate.
   ```

   (b) How is DNS amplification different from general amplification, and how do they relate to reflection attacks?

   ```
   Amplification generates multiple response packets for each original sent,
   sometimes by making the original request to a broadcast node. Generally
   these responses are from multiple systems.

   DNS amplification uses a DNS server and has larger responses than requests.

   Reflection is using an intermediate party that responds but to the target
   rather than the attacker. DNS amplification and reflection attacks
   both use intermediataries.
   ```

(c) Amplification, DNS amplification and reflection attacks do not generate backscatter traffic, although some sorts of DoS could. Explain what backscatter traffic is and why this is the case.

```
Various attacks are not really concerned about where responses
are sent to, so they use random but correctly formatted addresses.
This is backscatter traffic. By monitoring the backtraffic at IP addresses
where no real systems exist, so no legitimate traffic should venture,
it's possible to work out overall attack volumes.
```

8. Look at `http://www.cloudflare.com/ddos` and out something about the classes of DOS and the mechanism Cloudflare uses to provide protection.
9. There is an interesting news release in `news_media_34_3921121624.pdf`.
10. More details on a large DDoS...

    `https://www.wired.com/story/mirai-botnet-minecraft-scam-brought-down-the-internet/`

Part Two: Various bits and pieces

1. An attacker would be extremely pleased to find a website that had implemented a bogosort algorithm for sorting provided data. Why?

```
It's an extremely inefficient sorting algorithm.
It consists of randomly re-ordering the data and then
checking if they are in order. If they are in order you
stop. If not you go around again, so randomly re-ordering
the data ...
```

2. You can compile and run the provided `Bogosort.cpp`, using one of the following instructions to compile.

```
g++ Bogosort.cpp -o Bogo
CC -std=c++11 Bogosort.cpp -o Bogo
```

   You can run some tests to obtain some idea of the distribution of run lengths. You can edit the size of the data set in the code.

```
Students will hopefully get the impression that even
for a small case, like 4, it may take quite a long time.
It's probably helpful to note that it's not just the likely
poor performance that's a problem, it's also the unpredictability
of the time required.
```

3. Read `F-Secure News from the Lab.pdf`. Should we be worried?

```
This was an April Fools joke.
```

Part Three: Client Puzzles

1. What is the expected cost of calculating a puzzle consisting of m k–bit sub–puzzles, in the Client Puzzle Protocol of Juels and Brainard (1999)?

   ```
   This is in lecture set S4b. The expected work is m*2^{k-1}.
   ```

2. Using the above formula, draw up a table in Excel demonstrating the cost for $1 \leq m \leq 10$, $1 \leq k \leq 20$.

3. Assume a uniform distribution of hash values for the hash function used.

   (a) What is the probability of a single composite guess by an attacker solving the puzzle, as a function of m and k?

   ```
   By a compositite guess I mean a guess at all m blocks
   of k bits needed to be provided as a solution.

   Assuming only 1 of the 2^k cases is a solution for each, the probability
   of guessing correctly is 1/(2^mk)=2^(-mk).
   ```

   (b) What if the distribution of hash values wasn't uniform and was known? Would it increase or decrease the probabilities determined above?

   ```
   The probability would of guessing correctly would increase.
   The uncertainty/entropy would be reduced.
   ```

4. Why use multiple puzzles rather than one single large one?

   ```
   For a similar expected cost we reduce the standard deviation
   of the cost, so provide better assurances. It also makes it more
   difficult for an attacker to guess.
   As an example of the latter ...
   (k+3)-bit puzzle and 8 k-bit sub-puzzles have similar expectations
   on work, put the probability of guessing is 2^(-(k+3)) vs 2^(-8k).
   ```

SCIT-EIS-UOW
CSCI262/CSCI862 Spring 2021

Laboratory Week Six: Set Five - Solutions

Part One: Some code

A funny fact: This is a programming aside on something you may not come across, but it can have a significant impact on efficiency.

1. Compare the compilation times of `Fact1.cpp` and `Fact2.cpp`. You can use something like one of the following:

   ```
   time g++ Fact1.cpp -o Fact1
   time CC Fact1.cpp -o Fact1
   ```

2. The two pieces of code are performing the same function, in some sense at least. Time and record how long the programs take to run with different numbers. You might want to try up to 100 million or something on that magnitude.

   ```
   The compilation times should be pretty similar between
   Fact1 and Fact2, although CC will likely take longer than
   g++. Here goes an example of the difference in run times
   though, these for 50,000,000.

   $ time ./Fact1 50000000

   real    0m1.842s
   user    0m1.829s
   sys     0m0.010s

   $ time ./Fact2 50000000

   real    0m20.314s
   user    0m20.300s
   sys     0m0.011s
   ```

3. What is the advantage of the `Fact1.cpp` code?

Fact1 does the calculations at compile time, so only does them once. Fact2 does the calculations at runtime so recalculates them.

We may want to calculate data tables specific to a domain that can be specified at compile time or run time, rather at source distribution time. This saves having to distribute tables or recalculate them every time we run the progarm.

4. Would it be possible to have compiler time attacks?

   Sure. Probably primarily denial of service by using up excessive resources during compilation.

Part Two: Mobile code security, code security ...

1. Look at `www.owasp.org/index.php/XSS_Filter_Evasion_Cheat_Sheet`. This is a collection of sneaky ways to insert statements into websites.

2. What is Fuzzing?

   It's a type of automated testing.
   To quote Wikipedia, it uses "invalid, unexpected, or random data" as input to a program to check the response. This can include identifying when crashes occur, assertions fail and so on.

3. What limitations does fuzzing have?

   It's unlikely to find behaviour that is limited to not just a small class of input but a class that has some unusual characteristic, such as the example described in the lecture notes where there are a huge number of backslashes that cause a problem.

   While fuzzing is unlikely to find that, it may find a buffer overflow where there are many inputs that could trigger the problem. Since typically overflowing occurs when someone is entering long, expected to be invalid data.

4. Should fuzzing be considered a ”security engineering concept” or a ”software engineering concept”? Explain.

   More software engineering than security engineering. It's a general testing technique that is unlikely to find attacks that are based on exceptional behaviour. However software engineering is really

```
a special case of security engineering, it's just the system happens
to be software.
```

5. In one of the reading files the idea of ”Penetrate and Patch” is described as a ”dumb idea”. What is ”Penetrate and Patch” and why should it not be considered an appropriate approach to security?

```
Develop your system. Release it. When a bug is reported, fix it.
Security should be built into the design --> Security Engineering.
```

6. What is clickjacking? How is it related to XSRF (under XSS)?

```
The malicious practice of manipulating a website user's activity by causing
them to click somewhere other than where they think they are clicking.
This could be on concealed hyperlinks for example. This causes the user
to perform actions of which they are unaware.

Clickjacking is related to XSRF (or CSRF) in that both are doing things that you
didn't specifically request, the difference is that with XSRF the actions
are being carried out by your browser on loading, rather than specifically
taking some action as you do in clickjacking.
```

Part Three: Other bits and pieces

1. Which of these is a better implementation of a loop from the perspective of defensive programming? Why?

```
size_T length=strlen(store);
for (loop = 0; loop < length; ++loop)
      result += addme(store[loop]);
```

```
size_T length=strlen(store);
for (loop = 0; loop != length; ++loop)
      result += addme(store[loop]);
```

```
The first version is better than the second. In the second there is only a
single boundary check and it won't reject values that are already past the
boundary. The first one more clearly identifies the upper bound and will
```

reject loop values past the boundary.

It's possible if loop skips past the boundary condition the second loop won't stop. That shouldn't happen but there may be a hardware problem, or memory problem caused by another process, or a problem at the operating system or device driver level. There may be problems with the addme function that do things like corrupt the stack frame pointer or change the value at the location of loop. From Software Development_ Defences Programming.pdf

Both still have the problem that the loop value may change to a negative value and wouldn't be recognised as out of the appropriate rangel, or just generally that the loop value may change.

2. What is a Dutch auction? In what way could a Dutch auction be relevant to changing a password you have captured?

   A Dutch auction starts with a high price that gradually decrease until somebody bids. The first person to bid gets the item at that price, so the auction only has a single bid.

   The longer you wait before changing the password you have the longer you can use the account without being detected. It's pretty likely that you will be detected soon after changing password, but you then have the advantage of having locked out the legitimate user.

3. What are buffer underruns?

   This is where you are reading data from a location at a faster rate than the location can generate it so at some point you are attempting to read from an empty buffer.

4. Look at p64-Tanenbaum.pdf. What significant points does the article make?

There are warnings about believing correctness proofs relating to programming.

There are some similarities with showing that something satisfies some requirements, but that those requirements may not be the ones we actually want, or they are just lists of the way in which we know things might go wrong.

In section 3 there is quite a long list of things that may mean Proofs may be nonproofs". There isn't any real need to discuss this wiht the class but it's a useful list of software problems anyway.

SCIT-EIS-UOW
CSCI262/CSCI862 Spring 2021

Laboratory Week Seven: Set Six - Solutions

Part One: Phishing

1. At UOW:

   (a) If you were a UOW student and wanted to launch a phishing attack against your fellow students, what Bank would you use as the cover? Why?

   One of the banks on campus would make sense: National Australia Bank, IMB.

   (b) Is there any particular subset of students you would expect to get a ”better” response from?

   International students. International students are less likely to have an Australian bank account already and may well set up an Australian bank account when they arrive in the country and the convenience of the banks on campus make it more likely those are used.

   You would expect that students studying IT degrees would have some exposure to information security concepts, so are maybe less susceptible.

   (c) UOW info on cyber security: `http://www.uow.edu.au/its/cyber-security/index.html`

2. How much effort is involved in reproducing the front end of a phishing website associated with a bank? Have a look at a typical bank website and discuss (with other people ☺).

   Not a lot of work.

3. Picture-in-picture attacks and homograph attacks are two classes of phishing attack. Briefly describe each. For the homograph attacks you can have at look at script spoofing at `http://unicode.org/reports/tr`

   Section 3.3 of that report is on Buffer Overflows based on text expansion and is worth looking at.

Picture-in-Picture involves having multiple pictures on the screen at the same time. PIP Attacks involve manipulating what people see by layering multiple pictures. Parts like the URL and other security indicators can be in a distinct layer to mislead people into thinking they are at the correct location.

Homographs: Each of two or more words spelled the same but not necessarily pronounced the same and having different meanings and origins. Homograph attacks involve replacements such as the digit 0 and the letter O.

The text expansion relates to things like ASCII representations of Unicode symbols. 1 Unicode symbol might need multiple ASCII symbols.

4. In what way can personalised login screens provide protection?

The backward text is amusing.

By a personalised login screen I mean having something in your client browser that changes the appearance of your interaction with a website. This could be the background colour for example. If you go to a different version of the site it appears different so you are less likely to be fooled by it.

5. Evaluate the following method of protecting against Phishing:

Every time you go to a website to log in site you enter an incorrect password as your first attempt.

The idea could be that if it's not a valid site, so doesn't know your password, it may well "accept it" and say the link is dead. Takes longer to log in. It's possible the illegitimite site will pass on your password to the legitimate site so will determine the legitimacy of it anyway.

6. Have a look at Anti-Phishing Phil:

http://www.ucl.ac.uk/cert/antiphishing/

7. Look at 2018-Symantec-ISTR.pdf from last week. What does it say about the significance of spear–phishing?

```
It's a very common attacker vector for specifically targeted attacks.
See page 28:
"Spear-phishing emails emerged as by far
the most widely used infection vector, employed
by 71 percent of groups."
```

Part Two: Some Malware related tasks

1. Read `6a_Heavy_Metal_Worm.pdf`.

2. With reference to the standard read, write, execute permissions, when would be expect a process to be able to infect another process under the actions of a user?

```
Execute permission on an infected file and write access
on something else will allow that something else to
be infected.
```

3. There are three users, Alice, Bob, and Carol, with ACLs:

$F_1$:	(Alice, r), (Bob, rw), (Carol, x)
$F_2$:	(Alice, rx), (Carol, rwx)
$F_3$:	(Bob, rwx)

Consider the following independent scenarios:

(a) A virus has somehow infected $F_1$. With Alice being active can the virus infect $F_2$? Is the situation differerent for Bob and/or Carol? Justify your answers.

```
Alice cannot run $F_1$ so the virus cannot spread through her action.
Same with Bob.
Carol can run $F_1$ though so the virus would look for things
to infect, and since since Carol can write to $F_2$, the process
of $F_1$ being run by Carol will be able to infect $F_2$ with
the virus.
```

(b) If the virus had initially infected $F_2$ only, where could it spread and how?

```
Alice and Carol can both execute $F_2$, so that's a good start.
But neither Alice nor Carol can write to anywhere so $F_1
and $F_3$ are safe.
```

(c) If the virus had initially infected $F_3$ only, where could it spread and how?

```
Alice and Carol cannot execute $F_3$, but Bob can. Bob can write to $F_1$. So
Bob can run $F_3$ so the virus can be written to $F_1$. $F_1$ can then be run
by Carol and written to $F_2$. So all three of the files can becomer infected.
```

4. A computer system uses the BLP policies for access control. How would a virus spread if:

   (a) The virus were placed at lowest system level, that is, the compartment that all other compartments dominate?

   ```
   Remembering the rule is no write down there is no mandatory block on the
   virus being able to write to anywhere. The discretionary rules will determine
   where the virus can write.
   ```

   (b) The virus were placed on highest system level, that is, the compartment that dominates all other compartments?

   ```
   The mandatory rules mean the virus can at most write to the highest system
   level, and then only if allowed by the discretionary rules.



   Note that both of these questions are somewhat misleading since
   with malware we are most likely to be concerned about integrity.
   ```

5. Virus activity: Look at the following. There are a fair few malware mapping sites and a fair few places providing information on trends.

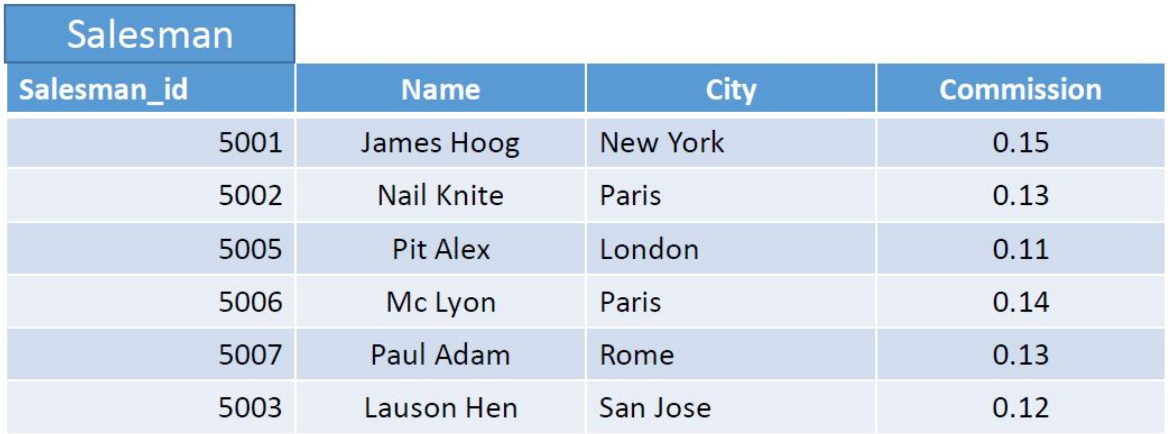
   ```
   http://cybermap.kaspersky.com/
   https://map.lookingglasscyber.com/
   http://www.digitalattackmap.com//
   http://www.mcafee.com/threat-intelligence/malware/latest.aspx
   ```

SCIT-EIS-UOW
CSCI262/CSCI862 Spring 2021

Laboratory Week Eight - Solutions

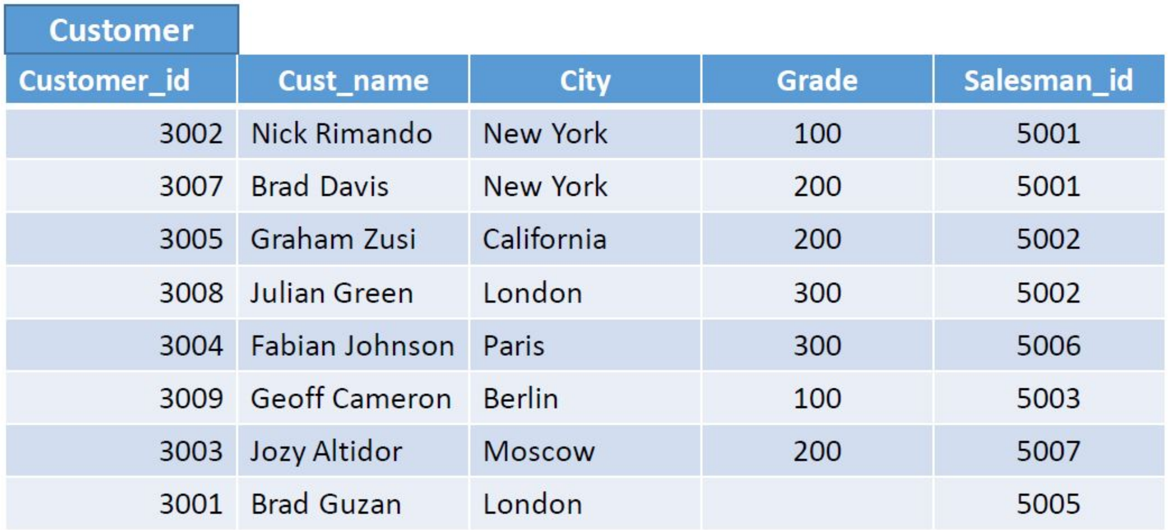
Part One: Database

1. Let's consider the following two relational tables.



Salesman

Salesman_id	Name	City	Commission
5001	James Hoog	New York	0.15
5002	Nail Knite	Paris	0.13
5005	Pit Alex	London	0.11
5006	Mc Lyon	Paris	0.14
5007	Paul Adam	Rome	0.13
5003	Lauson Hen	San Jose	0.12



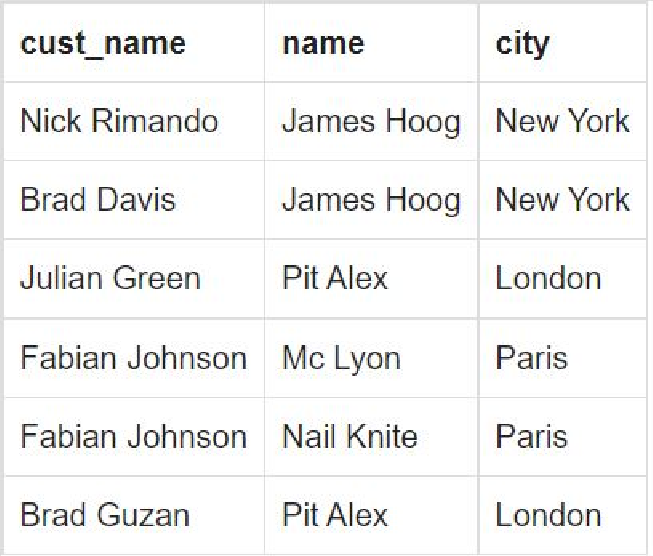
Customer

Customer_id	Cust_name	City	Grade	Salesman_id
3002	Nick Rimando	New York	100	5001
3007	Brad Davis	New York	200	5001
3005	Graham Zusi	California	200	5002
3008	Julian Green	London	300	5002
3004	Fabian Johnson	Paris	300	5006
3009	Geoff Cameron	Berlin	100	5003
3003	Jozy Altidor	Moscow	200	5007
3001	Brad Guzan	London		5005

(a) Write a SQL query to find the salespersons and customers who live in same city. Return customer name, salesperson name and salesperson city.

```
One sample solution would be the following:

SELECT customer.cust_name,
salesman.name, salesman.city
```

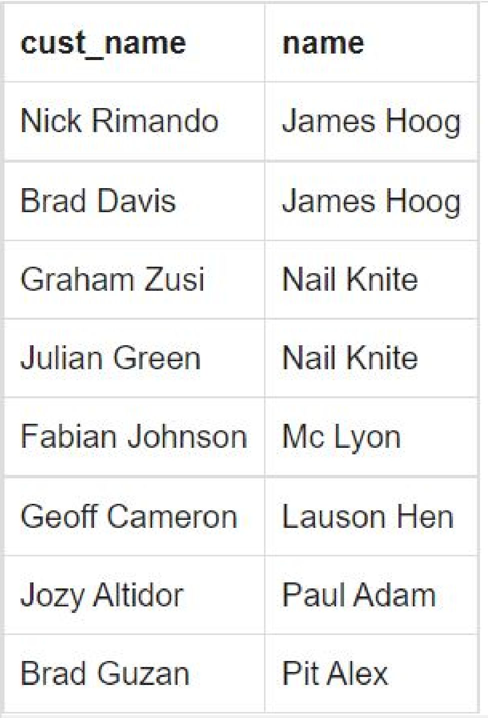
```
FROM salesman, customer
WHERE salesman.city = customer.city;

It will then return the following:
```



cust_name	name	city
Nick Rimando	James Hoog	New York
Brad Davis	James Hoog	New York
Julian Green	Pit Alex	London
Fabian Johnson	Mc Lyon	Paris
Fabian Johnson	Nail Knite	Paris
Brad Guzan	Pit Alex	London

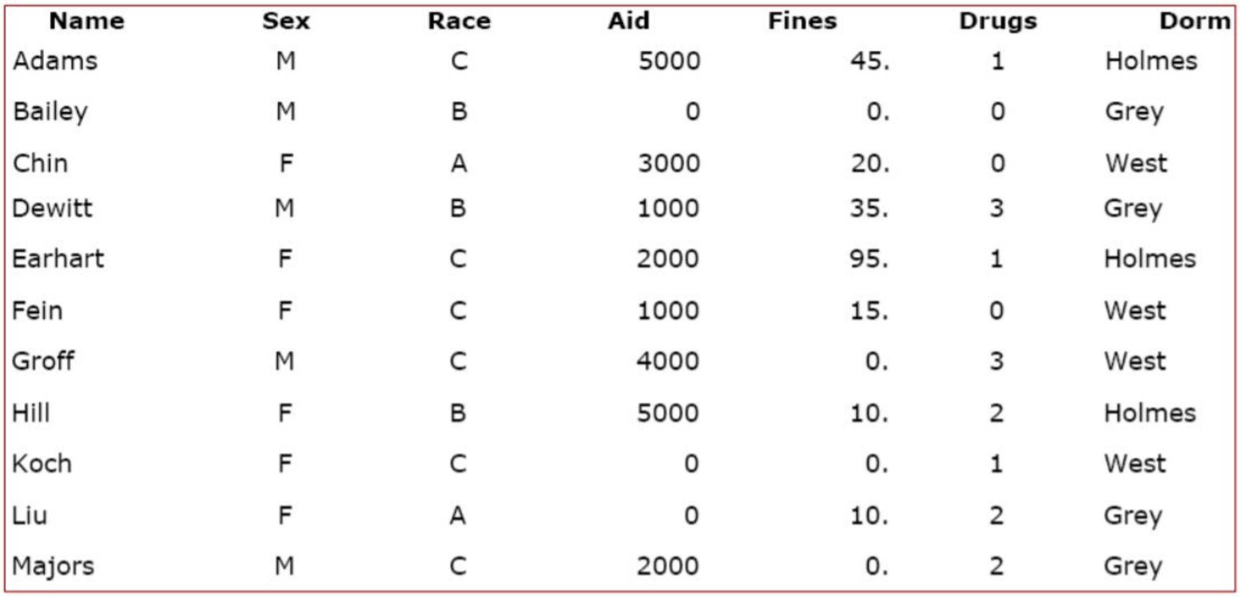
(b) Write a SQL query to find all the customers along with the salesperson who works for them. Return customer name, and salesperson name.

```
SELECT customer.cust_name, salesman.name
FROM customer,salesman
WHERE salesman.salesman_id = customer.salesman_id;
```



cust_name	name
Nick Rimando	James Hoog
Brad Davis	James Hoog
Graham Zusi	Nail Knite
Julian Green	Nail Knite
Fabian Johnson	Mc Lyon
Geoff Cameron	Lauson Hen
Jozy Altidor	Paul Adam
Brad Guzan	Pit Alex

2. Consider the following student table:



Name	Sex	Race	Aid	Fines	Drugs	Dorm
Adams	M	C	5000	45.	1	Holmes
Bailey	M	B	0	0.	0	Grey
Chin	F	A	3000	20.	0	West
Dewitt	M	B	1000	35.	3	Grey
Earhart	F	C	2000	95.	1	Holmes
Fein	F	C	1000	15.	0	West
Groff	M	C	4000	0.	3	West
Hill	F	B	5000	10.	2	Holmes
Koch	F	C	0	0.	1	West
Liu	F	A	0	10.	2	Grey
Majors	M	C	2000	0.	2	Grey

(a) You know that Liu is a female student living in Grey dorm. Assume the query set has to be greater than 1. Give a query/series of queries to reveal that Liu does not get financial aid.

We count how many students are there in Grey dorm

```
SELECT COUNT(*)
FROM Student
WHERE Dorm = ’Grey’
```

It will return 4

We next count how many male students are there in Grey dorm

```
SELECT COUNT(*)
FROM Student
WHERE Dorm = ’Grey’ and Sex = ’M’
```

It will return 3. So we conclude that Liu is the only female student in Grey dorm.

We use SUM to query to the database to report the total of student aid by sex and dorm by the following:

```
SELECT SUM(Aid)
FROM Student
WHERE Dorm = ’Grey’
```

It will return the total Aid for 4 students in Grey dorm, which is $3000
Then we query the following:

```
SELECT SUM(Aid)
FROM Student
WHERE Dorm = ’Grey’ and Sex = ’M’
```

```
It will return the total Aid for 3 male students in Grey dorm,
which is $3000. We can then infer that Liu has no financial aid.
```

(b) You know that Adams is living in Holmes dorm and Groff is living in West dorm. Assume the query set has to be greater than 1. Give a sequence of queries to reveal financial aid of Adams and Groff.

```
It is similar to Part (a).
```

Part Two: Some other things

1. What is a salami attack?

```
A salami attack is made up of many small attacks, each of
which seems minimal and unconcerning but as a whole they
may be dangerous. A typical example involves something like
removing 0.01 cents from every account in a bank. Each account
doesn't necessarily notice but the accumulated income may be
quite high. Something like a DoS by quantity can be thought
of as a salami attack, since the overall bandwidth is overwhelmed
by the many small messages each of which wouldn't be seen as
damaging.
```

2. Explain the concept of a ”banner grabbing attack”.

```
This is looking at header or similar information to determine
something like the version of  operating system being used,
or the version of an application. This can be important in deducing
what attacks a particular system may be vulnerable to.
```

3. At some point you should read `Wily-Hacker.pdf`.

```
It's an interesting example of an investigation.
```

Part Three: Statistics

1. There is data in `Data.txt`. These values represent daily totals drawn from a normal distribution with mean 9 and standard deviation 1.5, but with discrete integer values. Assume this event has a weight of 1, and is the only type of event.

   (a) In Excel, determine the mean and standard deviation of the data set, to 2 decimal places.

```
Mean: 8.92
Standard deviation: 1.37
```

(b) Assume the threshold is defined as the sum of the weights multiplied by 2 and daily values are drawn from a normal distribution with the mean and standard deviation, and then rounded to the nearest integer.

i. What integer range will be raised as an anomaly?

```
2 standard deviations below the mean takes us to 6.18.
2 standard deviations above the mean takes us to 11.66.

6 is below the lower range of 2 standard deviations and 12 is above
so the range of interest is 7 to 11.
```

ii. What range before rounding will result in an anomaly being raised?

```
Something from 6.5 up to 7 will round up to 7.
Something from 11 up to 11.5 will round down to 11.
So the range 6.5 to 11.5 is the one we are interested in.
```

iii. How many standard deviations away from the mean are the generally non–integer values in the previous question?

```
I've down this to 2 decimal places.
(8.92-6.5)/1.37 = 1.77
(11.5-8.92)/1.37 =  1.88
```

2. Some Dilbert: `05-Dec-2006.gif` and `08-May-2008.gif`. What is the relevance of each?

```
05-Dec-2006: This could probably be interpreted in a few ways.
A warning about be careful in specifications and definitions is
probably reasonable. In the context of intrusion detection we
need to be careful to distinguish between things which may be
given similar weight on the surface but are actually quite different.
```

```
08-May-2008: Statistics can be misused or abused. There are
circumstances where the statistics you have may not indicate
anything particularly useful. For example, having the mean and
standard deviation for a data set that doesn't follow a normal
distribution may be quite misleading. The flat client puzzle
distribution for example has a mean and standard deviation
but the distribution is very definitely not normal.
```

SCIT-EIS-UOW CSCI262/ CSCI862 Spring 2021

Laboratory Week Nine - Solutions

Part One: A few questions

1. What is a zero–day exploit?

```
It's when the first time a creator/supplier becomes aware of a vulnerability
is when an exploit using that vulnerability occurs.
```

2. Run `go.bat` in a Windows command prompt. What is the effect? What is this type of object referred to as?

```
Creates a directory, moves into it, creates a directory, moves into it.
End up with a lot of subdirectories inside each other.

We could think of this as a bacteria since it's using up one
resource. It doesn't do it very well though :)
```

3. The relevance of the `Base--Rate Fallacy` to intrusion detection systems is (Stallings and Brown):

```
In general, if the actual numbers of intrusions is low compared to the number
of legitimate uses of a system, then the false alarm rate will be high unless
the test is extremely discriminating.
```

The fallacy is a feature of Bayes' theorem. This theorem relates to conditional probability, and tells you the probability of some outcome in the context of known information. We will illustrate this using the example from Stallings and Brown, which explores what happens when a patient tests positive (`+ve`) for a disease. The information you are given is as follows:

- The accuracy of the test is 87%, meaning a patient with the disease (`unwell`) will get the correct result 87% of the time, while a patient without the disease (`well`) will get the correct result 87% of the time.
- The incidence of the disease in the population is 1%. This is the base–rate.
- There is no other basis than the test result to distinguish this patient from a general person in the population.

$$\begin{aligned} Pr[\texttt{well}|\texttt{+ve}] &= \frac{Pr[\texttt{+ve}|\texttt{well}]Pr[\texttt{well}]}{Pr[\texttt{+ve}|\texttt{unwell}]Pr[\texttt{unwell}] + Pr[\texttt{+ve}|\texttt{well}]Pr[\texttt{well}]} \\ &= \frac{(0.13)(0.99)}{(0.87)(0.01) + (0.13)(0.99)} \\ &\approx 0.937 \end{aligned}$$

So there is roughly a 93.7% chance the person is actually well. Intrusion detection systems don't cope very well with this type of problem.

(a) What are the possible ramifications of these types of inaccuracies?

Generally we can treat something that is actually negative as if it were positive. In the medical context this may mean deploying intrusive/expensive/painful medicine/procedures when somebody doesn't need it. In the context of crime detection it may mean treating someone as a criminal or as malicious when they actually aren't.

(b) Consider a test for an event with a incidence rate of 50%, rather than 1%.

$$\begin{aligned} Pr[\texttt{well}|\texttt{+ve}] &= \frac{Pr[\texttt{+ve}|\texttt{well}]Pr[\texttt{well}]}{Pr[\texttt{+ve}|\texttt{unwell}]Pr[\texttt{unwell}] + Pr[\texttt{+ve}|\texttt{well}]Pr[\texttt{well}]} \\ &= \frac{(0.13)(0.50)}{(0.87)(0.50) + (0.13)(0.50)} \\ &= 0.13 \end{aligned}$$

So there is a 13% chance the person is actually well.

(c) Consider a test for an event with a incidence rate of 0.01%, rather than 1%.

$$\begin{aligned} Pr[\texttt{well}|\texttt{+ve}] &= \frac{Pr[\texttt{+ve}|\texttt{well}]Pr[\texttt{well}]}{Pr[\texttt{+ve}|\texttt{unwell}]Pr[\texttt{unwell}] + Pr[\texttt{+ve}|\texttt{well}]Pr[\texttt{well}]} \\ &= \frac{(0.13)(0.9999)}{(0.87)(0.0001) + (0.13)(0.9999)} \\ &\approx 0.9993 \end{aligned}$$

So there is a 99.93% chance the person is actually well.

(d) For each of the 3 considered incidence rates, how accurate would the tests need to be, so what value should replace 87%, so that a positive result gives a 50% chance of the person actually having the disease?

```
Students can do this approximately if they like, so they get a rough idea.
Base rate: 1% -->   test accuracy: 99%
Base rate: 50% -->   test accuracy: 50% ... flip a coin :)
Base rate: 0.01% -->   test accuracy: 99.99%
```

SCIT-EIS-UOW
CSCI262/CSCI862 Spring 2021

Laboratory Week Ten: Solutions

Part One: Some introductory questions

1. What did `von Neumann` say about randomness? What does this mean?

   ```
   John von Neumann: "Any one who consider arithmetical methods of
   producing random digits is, of course, living in a state of sin."

   For arithmetical, read deterministic.
   You cannot produce information out of nowhere.
   ```

2. What is geospatial intrusion detection (GID)?

   ```
   This makes uses of Geographic Information Systems (GIS) in network
   security. It can be used to determine the source of events, such as
   where somebodies devices are physically connected.

   This can be helpful in distinguishing between false positives
   and actual threats.
   ```

3. What is the wasp trap syndrome?

   ```
   This is tied to honeypots. In making something appear attractive
   to distract the existing volume of attackers, you can actually
   increase the volume of attackers because they are attracted.

   This is named because a method used to trap wasps involves
   scented bait, that may similiarly attract more wasps than you
   were originally trying to deal with.
   ```

4. What is the relevance of units for intrusion detection systems? See also:
   `SingaporeTimes-SST-GPS.pdf`

   ```
   Without units entries in logs may be ambiguous, so auditing
   or intrusion detection on those logs may be inaccurate. It
   doesn't mean the units need to be explicit in every log entry,
   rather than their needs to be a known standard that instances
   of the data conform to, at both input and output phases.
   ```

Part Two: Some More Statistics

1. In looking at intrusion detection systems we briefly discussed the idea of using distance metrics between values to allow us to construct a composite threshold.

   (a) Here go four examples of distances, not necessarily typically used in intrusion detection system but good enough to illustrate the idea of distance.

      i. Hamming distance.

      ```
      The Hamming distance between two words of the
      same length is the number of corresponding positions
      in which the words differ.

      Difference between (1 2 3 4 5) & (5 4 3 2 1) is 4.
      Difference between (1 2 3 4 5) & (2 3 4 5 1) is 5.
      Difference between (2 2 2 2) & (1 1 1 1) is 4.
      ```

      ii. Euclidean distance.

      ```
      Ordinary straight line distance.
      Calculations at Wolfram Alpha under Euclidean distance.

      Difference between (1 2 3 4 5) & (5 4 3 2 1) is 2*sqrt(10)~6.32.
      Difference between (1 2 3 4 5) & (2 3 4 5 1) is 2*sqrt(5)~4.47.
      Difference between (2 2 2 2) & (1 1 1 1) is 2.
      ```

      iii. Manhatten distance.

      ```
      The sum of the absolute differences between corresponding
      elements. Also called taxicab geometry. It corresponds to
      travelling along the roads of a Cartesian grid.

      Difference between (1 2 3 4 5) & (5 4 3 2 1) is 12.
      Difference between (1 2 3 4 5) & (2 3 4 5 1) is 5.
      Difference between (2 2 2 2) & (1 1 1 1) is 4.
      ```

      iv. Levenshtein distance.

      ```
      The distance between two words is the number
      of single character edits needed to turn one word
      into the other. These edits are deletions, insertions
      or substitutions.

      Calcs at: https://planetcalc.com/1721/

      Difference between (1 2 3 4 5) & (5 4 3 2 1) is 4.
      Difference between (1 2 3 4 5) & (2 3 4 5 1) is 2.
      Difference between (2 2 2 2) & (1 1 1 1) is 4.
      ```

   (b) You should do the following in the context of those distances:

      i. Find the definition of each.

      ii. Determine how far apart the strings `(1 2 3 4 5)` and `(5 4 3 2 1)` are.

      iii. Include some additional examples that highlight the differences between the distances.

Part Three: Firewalls

1. What is a VPN and what is a VPN used for?

   VPN is short for virtual private network. It’s a generic term that’s used to describe any combination of methods to secure connections through otherwise insecure networks. Typically used to provide access to resources affiliated with a remote network, such as a remote user accessing their company network.

   The VPN extends the private network, virtually rather than physically, to somewhere physically remote to the base private network.

2. What is a firewall in a non network security sense?

   Mostly in construction, a barrier to stop or limit the spread of fire or heat.

3. What is a fire break? How is it related to air gapping?

   A firebreak is like a firewall in the sense of the purpose but it’s an absence rather than a presence. The idea is to have a section where the fire doesn’t have any fuel, so burns out. Roads or trails are often firebreaks. Rivers can be natural firebreaks. The use of firebreaks is based on the assumption that fires can only ‘‘jump’’ a limited range.

   Air gapping is having a physically isolated network. Same sense as firebreak where you cannot jump between fuel, here you cannot jump between networks beacuse there is no infrastructure in between.

4. What are the two primary underlying assumptions behind firewall deployment?

   Topology and Trust. Topology relates to the firewall is assumed to be the only point of contact into the network being protected. Trust relates to insiders being trusted and outsiders being distrusted.

SCIT-EIS-UOW
CSCI262/CSCI862 Spring 2019

Laboratory Week Eleven: Set Ten

This weeks exercises are about obfuscation.

1. Here goes a fragment of Java code.

```
boolean matches(String test){
     if (md5(test).equals("d077f244def8a70e5ea758bd8352fcd8")
          return true;
     if (md5(test).equals("97223fab7b0d4c64c07e6e004c602302")
          return true;
     if (md5(test).equals("fe47aa7c733c490d36e80508d5dc4019")
          return true;
     return false;
}
```

(a) What is the literal interpretation of what is happening?

```
An input string is being taken, hashed with md5, and compared against
three hash values. The function returns true iff the hash of the input is
equal to one of those three statements.
```

(b) What do you think the purpose of this function is?

```
Testing to see if a condition is matched, perhaps for a password or keyword.
```

(c) Why would you carry out this function in such a way?

```
Obfuscation, so people looking at the code cannot readily see what
matches and would result in true being returned.
This type of obsfuscation can be used in matching regular expressions,
so for this c[aou]t matches cat or cot or cut. A regular expression
can be expanded into matches against all the possible cases.
```

(d) Think of two specific problems with this method.

```
The number of matching conditions is revealed.
If you have some idea of the type of thing to be matched you can
launch a dictionary attack. If the input string is constrained brute
force may be feasible. Either may be worthwhile for a restricted
enough space.

The hash values can potentially be editing in the binary so they
match against some chosen input string.
```

2. Compile `Obs1.c`. You can compile it using `cc` or `gcc`. It is safe to run ⌣̈.

   ```
   Note that it's C code, not C++.
   It's produces the words for the 12 Days of Christmas.
   ```

3. `Obs1_depart.c` contains a partial de–Obfuscation of the code in `Obs1.c`.

   (a) Where in the program does the actual printing to output occur?

   ```
   putchar(31[a])
   Even without any knowledge of C this the only likely option.
   ```

   (b) What is the purpose of the character array `strings`?

   ```
   Strings contains the encoded versions of the text for the song.
   Every / is the end of a statement.
   Note the \ just make the whole thing a single C-string.
   ```

   (c) What is the purpose of the character array `translate`?

   ```
   You can split it into the input and output for a substitution.
   It acts as a translation or substitution table.

   The array is folded in the middle so you would get the table
   below. I'd expect people would more likely figure out the
   answer to the mapping in the next question and gradually
   fill out this table until they saw the relationship.
   ```

!	e	k	;	d	c	i	@	b	K	’	(	q	)	-	[	w	]	*	%	n	+	r	3	#	l	,	{	}	:
\n	u	w	l	o	c	a	O	;	m		.	v	p	b	k	s	,	f	x	n	t	d	C	e	g	h	i	r	y

   (d) You should be able to use the information about the `skip_n_strings` function to break `strings` into "phrases". Thinking about what "`+,`" and "`{nl]`" mean in `strings` should help?

   ```
   +,     maps to   "th".
   {nl]    maps to    "ing,".
   ' maps to a space.

   Using _ to represent a space...
   The first phrase is
   @n'+,#'   which maps to On_the_
   The next 12 maps to First to Twelfth
   E.g ...
   #;#q#n+,   maps to eleventh

   After that ...
   'r :'d*'3,}{w+K w'K:'+}e#';dq#'l q#'+d'K#!
   maps to ...
   day of Christmas my true love gave to me

   The enum should give a pretty good idea
   about the rest.
   ```

(e) Do those two particular correspondences above suggest something that could be done to make the code smaller without effecting the result? Explain.

```
Some sequences of letters appear frequently so we can use compression.
We could map our translate array larger and use a single symbol to
map something to "ing," for example. The translation would necessarily
be more complex.
```

(f) You should be able to write your own version. Later though.

4. Look at `WMExample.java`. This is taken from Listing 1.6 in `Surreptitous Software` by Collberg and Nagra. It is an example of adding watermarks. How is Bob embedded in this program?

```
There is the obvious fingerprint variable.
There is also the output Bob that occurs when 42 is entered.
This is an example of a dynamic fingerprint.

The others are more difficult to see because they effectively
involve secret translations.

One example involves the translation table below ...
```

	+	-	*	/	%
1	a	b	c	d	e
2	f	g	h	i	j
3	k	l	m	n	o
4	p	q	r	s	t
5	u	v	w	x	y
6	z				

```
bob is translated to
1-3%1-
with the brackets splitting each
1-(3%(1-x))
and the x added at the end
and followed by closing brackets.
```

5. Decompilation and Obfuscation in the news a couple of decades ago:

`https://archive.is/20120710225423/http://news.com.com/2100-1023-222781.html`

6. The zip file `jd-gui-windows-1.4.0.zip` contains the executable for, according to the documentaiton, ”JD-GUI a standalone graphical utility that displays Java sources from CLASS files.”

You could probably see how well this works on Windows.

7. There is an earlier article describes some specific techniques used in Java obfuscation. It includes examples.

```
http://www.informit.com/articles/article.aspx?p=174368
```

8. De–obfuscate `Obs2.cpp`. It might not compile as is in a Windows environment. This is safe to run and needs two integer arguments. What does it do?

```
Running
$ ./Obs2 N1 N2
produces a list of all postive integers less than N1 that are
relatively prime to N2.
```

SCIT-EIS-UOW
CSCI262/CSCI862 Spring 2019

Laboratory Week Twelve: Set Eleven - Solutions

Part A: Inference

1. Consider the table below, representing the content of a database, and answer the related questions.

Name	Gender	School	Position	Salary
Alex	Male	Computing	Lecturer	\$80,000
Bobby	Male	Mathematics	Lecturer	\$60,000
Carol	Female	Mathematics	Lecturer	\$100,000
Diana	Female	Computing	Lecturer	\$60,000
Ewen	Male	Physics	Lecturer	\$72,000
Fran	Female	Physics	Lecturer	\$88,000
Gary	Male	Computing	Administrator	\$40,000
Hubert	Male	Mathematics	Lecturer	\$72,000
Ivana	Female	Computing	Tutor	\$12,000
Jeff	Male	Physics	Administrator	\$80,000
Kim	Female	Mathematics	Lecturer	\$100,000
Lex	Male	Computing	Tutor	\$12,000
Morris	Male	Engineering	Tutor	\$15,000

2. This is relevant to some of the perturbation techniques used to provide protection against statistical inference. Consider the follow distribution of values, and answer the subsequent questions.

```
10  20  30  40  50  60
```

(a) Determine the mean and sample standard deviation, that's the version using $n-1$ as below

$$\sqrt{\sum \frac{(x-\bar{x})^2}{n-1}}$$

```
Mean: 35

Standard deviation: 18.71 (2d.p.)
```

(b) Determine a new distribution of six numbers, each of which is at least two away from all of the existing values, such that the mean is the same as that of the original and the standard deviation differs from the original by at most 0.2.

```
Here goes one example of a solution: 12, 23, 23, 42, 48, 62.
Mean: 35
Standard deviation: 18.79 (2d.p.)
```

(c) We can state the previous question generally. Given a collection of numbers C_1 generate another set C_2, $|C_1| = |C_2|$, such that

   - Each value in C_2 is at least some $d$ from each value in C_1.
   - The means of C_1 and C_2 are the same.
   - The standard deviations of C_1 and C_2 differ by at most $\epsilon$.

   Think a bit about how you might systematically solve this problem and whether there are constraints that may cause problems.

```
It's should be apparent this isn't always going to be possible,
such as when $d$ is large or $\epsilon$ small.
It's likely only readily possible with a small, but not too small
set. If the set was too small we will too easily impact on the
standard deviation when we shift values by the necessary.

Generally maintaining the mean is straightfoward, you put
up values by the same total amount you put others down.
Maintaining the standard deviation is difficult. To a certain
extent you can try to make the number of ups/downs balanced
to either side of the mean.
```

Part B: Various other questions

1. Consider how can physical user input, such as Barcodes, RFID, or OCR, be used as an avenue of attack for SQLi. See

```
https://www.irongeek.com/xss-sql-injection-fuzzing-barcode-generator.php
https://security.stackexchange.com/questions/3949/sql-injection-in-a-non-web-application
```

2. What is a trusted platform module (TPM)? What are they used for?

```
A trusted platform module (TPM) is a chip on the PC motherboard,
or on a smart card, or integrated into a processor.

The TPM is responsible for generating keys and managing the security
associated with behind the scenes data flow. Specifically it's responsible
for:

- Authenticated booting of the operating system.
- Generation of certificates, using a private key, that allow third parties
to verify the legitimacy of a configuration.
- Data encryption such that decryption is only possible within the correct
configuration.
```

Part C: The Exam

Look at questions from previous years exams, available from:

https://ereadingsprd.uow.edu.au/listpage.php?prog=CSCI262

The aim isn't to give students answers to questions from previous years, just to get them to be thinking about what they need to know and where their subject knowledge sits.

Part D: Assignment Three

Any remaining time can be spent on assignment three.

SCIT-EIS-UOW
CSCI262/CSCI862 Spring 2021

Laboratory Week Two: Set One - Solutions

These are sketches/notes, rather than formal complete solutions.

1. The first questions relate to passwords and entropy:

   (a) Is there any harm in revealing old passwords? Why or why not?

   Yes, for several reasons. People sometimes reuse passwords.

   People sometimes follow patterns in choosing passwords so an old password may leak information about a current password. Passwords may not be as old as you think, they may be protecting data you had forgotten about or they have been used to protect communication that somebody previously captured and would now be able to read.

   (b) What is the entropy associated with a password chosen with uniform randomness from the set of length 8
   strings with symbols taken from the lowercase alphabet {a,...,z}?

   Passwords equal likely. Determine how many there are by N^L, N number of symbols in the symbol set and length L.

   Note that this won't be appropriate for determing number is different positions have different requirements.

   Emphasise that the units are important.

$$\log_2\left(26^8\right) = 8 * \log_2(26) \approx 37.60 \text{ bits.}$$

   (c) How much entropy is there associated with a typical ATM PIN?

Typically 4 digit PINS...

$$\log_2\left(10^4\right) = 4 * \log_2(10) \approx 13.29 \text{ bits.}$$

(d) Look at `http://www.datagenetics.com/blog/september32012/`

`This is an article on the distributions of PINs.`

`Noting that the distribution isn't uniform is important, and that`

`non-uniformity would be expected for passwords too.`

(e) Is `fDtk53$e3W22eSDmvfFp-4F` a good password?

`Instinctively it looks decent but in terms of the entropy measure`

`it depends on the method by which it has been chosen. If an`

`attacker knows how you choose your password and it only ever`

`in the 4F, digit Upper, at the end, it's not so good at all because`

`there aren't many options.`

(f) Without writing down your password, or the method of choosing your password, estimate the entropy associated
with the password you use most.

(g) How much confidence do you have in the method of choosing your password not being guessed?

(h) How much confidence do you have in your password under the assumption the method of choosing your password was
known by an attacker?

(i) How does considering options that are not all equally likely impact on the entropy?

`A uniform distribution maximises the entropy. If we have a non`

`uniform distribution the attacker has a better idea as to what`

`some might be so can weigh the questions they ask towards`

`needing less questions in those cases.`

2. The next set of questions relate to hashing, partially in the context of password systems:

   (a) Does taking $H(M)$, for H a cryptographic hash function, provide confidentiality for M?

   ```
   Mostly yes, in the sense that an unauthorised person is
   unlikely to be able to determine M given H(M) since the
   cryptographic hash function has pre--image
   resistance/one--way ness.
   But, if the set of possible values for M is known, it's
   possible to attempt brute force by taking each possible
   message and hashing them and seeing if they match.
   If the space isn't large it may actually be fairly easy to
   recover the original message


   Hashing is not encryption, because with encryption you
   expect authorised users to be able to decrypt and
   recover the original. But with hashing you generally
   lose information.
   ```

   (b) How might hashing be used in generating a password? How does it influence the entropy?

   ```
   Since hashing ideally hides the initial content you can
   choose something that you easily remember and then
   hash it to generate your password. You can remember
   your original and hash to recover. Notice that somebody
   can still launch an attack against your original choices
   of password.
   ```

The hashing is deterministic. It has roughly no impact on the entropy since if you had 1000 equally likely options before hashing you probably have 1000 equally likely after, assuming collisions are unlikely.

(c) What is the advantage of using a hash function like `bcrypt` rather than a classical cryptographic hash function such as MD5 or SHA1?

Scalable complexity so that it takes longer for an attacker to apply hashing.

(d) Hashing ”produces a fingerprint” of a message. In what way does this misrepresent the relationship between hash and message, relative to the relationship between human fingerprint and human?

Human fingerprints are supposed to be unique, so you won't have two people with the same fingerprint. Messages however can hash to the same hash value, so there are collisions.

(e) Look at `Trapped.gif`. What is the relevance to cryptographic hashing?

It's an eel trap representing a one--way function. The eel can go in but cannot go out.

(f) Read `Time-Oct3-2011.pdf`, not necessarily in the lab but at some point.

SCIT-EIS-UOW
CSCI262/CSCI862 Spring 2021

Laboratory Week Three: Set Two - Solutions

This lab is Unix based, students need to be connected to Capa.

Part One: More on passwords. Collisions

1. Find where `passwd` and `shadow` are located. **It tells you in the lecture notes!**

   ```
   /etc/passwd
   /etc/shadow

   These are files not directories.
   ```

   (a) Find your own entry. Identify your userid number and native group number. `grep` and `man` are likely to help with this exercise.

   ```
   grep "hduong" /etc/passwd
   ```

   (b) How large are the `passwd` and `shadow` files?

   ```
   This can be found using
   $ ls -l /etc/passwd
   $ ls -l /etc/shadow

   Size in bytes: passwd: 4856159    shadow:  6207198

   wc /etc/passwd
   gives the number of lines, words, and characters.
   One line per user -->  64179   users
   ```

2. Some scripts are in the zip file. The perl script `2a_crypt.pl` was discussed in the lecture.

   (a) Run `2a_crypt.pl`.

   ```
   perl 2a_crypt.pl

   Usually if you run this twice in a row you will get different results.
   ```

   (b) Run `2a_crypt2.pl`. How does it differ from the previous perl program?

This one produces truncated output. It calcules the same thing as the previous script but only outputs the first two symbols of that string.

(c) What does $$ mean in the perl script? Why is it used?

It's the process id for the perl script that is running. If somebody knows when the crypt script was run they can calculate the time element, but they are unlikely to know the process id. Running the script multiple times at the same time will still generate different results because of the dependence on the process id.

(d) Increase the length of the output string from `2a_crypt2.pl` by 1, and repeat your tests. Keep on doing this. Does this make a significant difference? Why or why not? What modifications could make a difference? Make another modification and test your idea.

```
 This involves modifying
$reduce=substr($record,0,2);
to
$reduce=substr($record,0,3);
and so on...
```

Students will probably find it doesn't make a huge amount of difference but generally there will be a trend towards the longer output, the longer it will take to get a collision.

However there are collisions fairly quickly even prior to truncation, so this doesn't help much.

```
Another change that seems reasonable is to
change the salt space.
This is changing this line ...
@salts=('48' .. '57', '65' .. '70');
```

Part Two: Access Control

1. I have almost 4000 files on Capa.

(a) Assume we independently record read, write and execute permissions for every user on Banshee on each of my files. How much space would be needed to do this?

```
1 bit for each file and each user for each of the
three permissions.
So the total cost is
3*(number of users)*(number of files)
3*57966*4000
= 695592000 bits
= 86949000 bytes
~ 87 megabytes.
The number of users is taken from the
number of rows of /etc/passwd determined
earlier using wc.
```

(b) How many bits actually need to be recorded for the access control of those files?

```
Again just looking at read, write, and execute, we need
3*3*(number of files)
```

with the number of users being reduced to the number of distinct access classes on each object: owner, group, others/universe.

(c) Why does this suggest about the use of different representations?

```
It's worth putting some effort into thinking about which is most appropriate
because getting it wrong could be very costly. In a multi-user
operating system/file system we expect that a lot of the time most of
the users won't need differentiated access to objects most of the time.
```

2. Alice can read and write with respect to the file O_1, can execute the file O_2, and can read the file O_3. Bob can read O_1, and can read and write with respect to O_2. Carol can read O_3 and execute O_2.

(a) Draw up an access control matrix for this situation.

Use R, W, X to represent read, write, and execute, respectively. The first component to get is the access control matrix:

	O_1	O_2	O_3
Alice	RW	X	R
Bob	R	RW	
Carol		X	R

(b) Write the complete set of access control lists for this situation.

The access control lists are from the point of view of the objects. Not a single form of bracketing, different notations are possible.

```
O1:{(Alice,RW),(Bob,R)}
O2:{(Alice,X),(Bob,RW),(Carol,X)}
O3:{(Alice,R),(Carol,R)}
```

(c) Write the complete set of capability lists for this situation.

The capabilities are from the point of view of the subjects.

```
Alice:{(O1,RW),(O2,X),(O3,R)}
Bob:{(O1,R),(O2,RW)}
Carol:{(O2,X),(O3,R)}
```

SCIT-EIS-UOW
CSCI262/CSCI862 Spring 2021

Laboratory Week Four: Set Three - Solutions

1. In a lattice based system a security level is composed of a clearance/classification and a composite category. Given classifications Top Secret, Secret, Confidential, and Unclassified (ordered from highest to lowest), and the categories A, B, and C, we have security levels like (Top Secret, {A,C}). Such a level implies Top Secret clearance in {A} AND in {C} as well, and that the composite {A,C} dominates both {A} and {C}. The composite category part of the clearance {A,C}, is higher than having both A and C clearance independently.

   For each of the following specify what type of access (read, write, read & write or none) is allowed, assuming the discretionary access matrix permits it. Use BLP rules to determine access. Assume the people have operated at the specified clearance.

   (a) Alice, cleared for (Top Secret, {A,C}), wants to access a document classified (Secret, {B,C}).

   ```
   None. No domination relation between the levels because
   of the lack of a domination relation between the categories.
   ```

   (b) Bob, cleared for (Confidential, {C}), wants to access a document classified (Confidential, {B}).

   ```
   None. No domination relation between the levels because
   of the lack of a domination relation between the categories.
   ```

   (c) Chris, cleared for (Secret, {C}), wants to access a document classified (Confidential, {A}).

   ```
   None. No domination relation between the levels because
   of the lack of a domination relation between the categories.
   ```

   (d) Dan, cleared for (Top Secret, {A,C}), wants to access a document classified (Confidential, {A}).

   ```
   Read only. Top Secret dominates Confidential and {A,C} dominates {A}.
   ```

   (e) Eve, who has no clearances (and so works at the Unclassified level), wants to access a document classified (Confidential, {B}).

   ```
   Write only. Confidential dominates Confidential and {B} dominates {}.
   ```

2. In general, you may have a scenario where you need to identify the subjects, objects and actions. How would you do so for the following statements? Can you think of any general implications from handling the case?

   ```
   Alice kicks Bob.
   Bob kicks himself.
   Alice and Bob can lift the table together.
   ```

Alice kicks Bob.	Objects can be subjects too.
Bob kicks himself.	Self-action.
Alice and Bob can lift the table together.	Collaborative actions.

3. Suppose Bob wishes to edit the file F_1 in a capability based system.

   (a) How can Bob check to see what other files he can access using the editing program?

   The capability based system stored in the information with the user, so it should be easy for Bob to locally check.

   (b) Could this be done in an ACL based system? If so, how? If not, why not?

   Possible to check but Bob would need to go through every object in the system and check the ACL's that are stored with each of those objects :(

4. RBAC questions:

   (a) Consider a role–based access control system where users are classified in accordance with a BLP model, so each user is given a clearance in a lattice level. Describe a potential problem with a user having access to multiple roles with different clearances?

   It's possible to read something from a sensitive level under a role that can read it, and then write back at a less sensitive level using a role that is allowed to write there.

   (b) Page 15 of lecture set Week3 contains an example of a role hierarchy.

      i. Could this be a lattice? Why or why not?

      Assuming we define the operator appropriately it can be. There is a least upper bound on any pair of models and a greatest lower bound on any pair of models.

      ii. If it is, how might the operation $\leq$ be interpreted, defined, or described?

      One of the models dominates another if it has all the functionality of the other.

      Note that a model would be dominated by the other if it contains a subset of all functionality of the other.

      iii. If it's not a lattice, could it be turned into one? If so, how so?

      It can be thought of as a lattice so nothing to do hear really :)

      iv. Would a role hierarchy always have to be a lattice, or be able to be turned into a lattice in the way you may have described?

      No. We may need to add an additional level, or levels, if there isn't both a least upper bound and a greatest lower bound for any pair of levels.

5. There are various important principles relevant to access control. This question looks at them.

   (a) The Principle of Least Privilege:

      i. What is it?

         `Give subjects as much privilege as they need to do their job and no more.`

      ii. Why is it important?

         `Reduces scope for damage, malicious or accidental.`

      iii. Is it applied in practice?

         `Often only approximately because the fine scale of differences can be problematic to manage in terms of availability and request complexity. Often there will be levels of access priviliges and you might give someone the lowest than allows them to do their job but they may still be able to access things they don't need. Academics being able to access records on any students is a good example. Logging/intrusion detection systems may detect policy breaches if there isn't a blocking mechanism.`

   (b) The Principle of Attenuation of Privilege:

      i. What is it?

         `You cannot give a privilege that you don't have yourself to someone else.`

      ii. Why is it important?

         `Restricts addition of rights. Note it doesn't mean that anyone with a permission can pass it to someone else.`

      iii. How does it apply to the owners of objects?

         `Owners can generally give themselves read/write/execute etc. on their own files so the check isn't typically applied when it's the owner passing something.`

6. The Chinese Wall model of access is about confidentiality but there is a meta property governing confidentiality, conflict of interest.

   (a) Explain what is meant by a conflict of interest.

      `It's where there is a competing interest or responsibility, maybe personal or professional. Somebody who should be neutral may not be anymore.`

      `Somebody who works for the University and owns shares in a fleet of rental cars should declare a conflict of interest if it was going to be them deciding which rental company was to be used by the University.`

(b) We can think of something like the Chinese Wall model as implying dynamic access control, so an access control matrix would be changed by virtue of an action taking place. Give examples illustrating how adding triplets to the current access set can change the access control matrix.

```
In the context of an example, such as the one in
the lecture notes, somebody carrying out access
within a conflict class precludes them from subsequent
actions within that conflict class.
```

7. We didn't go into a great deal of detail about ABAC in the lecture notes. Here goes an example of an ABAC policy, taken from the textbook. The policy R1 describes when a user u can access a movie m.

$$\begin{aligned} R1 \;:\; & \text{can\_access}(u, m, e) \leftarrow \\ & (\text{Age}(u) \geq 17 \wedge \text{Rating}(m) \in \{R, PG13, G\}) \vee \\ & (\text{Age}(u) \geq 13 \wedge \text{Age}(u) < 17 \wedge \text{Rating}(m) \in \{PG13, G\}) \vee \\ & (\text{Age}(u) < 13 \wedge \text{Rating}(m) \in \{G\}) \end{aligned}$$

(a) What are the subjects, actions, and objects of relevance here?

```
Subjects: Users.
Actions: Access. Kind of vague ... view.
Objects: Movies.
```

(b) What are the relevant attributes here?

```
Users have the attribute Age.
Movies have the attribute Rating.
```

(c) Draw a table explaining the access control based on the attributes.

Movie rating	Age to view
R	17 or older
PG13	13 or older
G	Any age

(d) Provide a simplified policy representing the same access but based on your table.

$$\begin{aligned} R1 \;:\; & \text{can\_access}(u, m, e) \leftarrow \\ & (\text{Age}(u) \geq 17 \wedge \text{Rating}(m) \in \{R\}) \vee \\ & (\text{Age}(u) \geq 13 \wedge \text{Rating}(m) \in \{PG13\}) \vee \\ & (\text{Rating}(m) \in \{G\}) \end{aligned}$$

Part Two: CAPTCHA

These questions all relate to CAPTCHA.

1. `ReCaptcha`: What is it?

```
This is a project used to help identify humans,
thus Captcha, but also to help in digitising books.
```

Users are given a typical Captcha challenge and a scanned word from an old text that optical character recognition (OCR) might have failed on. The same scanned word can be given to multiple users to give a reliable transcription.

2. NuCaptcha: What is it? You should go to the website of the company and have a look at some examples. Look at the article in the lab directory too and see what claims they make.

   Use animated video technology to make puzzles.

3. DeCaptcha: What is it?

   Supposed to be for solving captcha challenges.

4. In groups of size greater than one, discuss alternative forms of CAPTCHA.