Lab 2: SSL Certificates
Due at 11:59PM on April 11 as a PDF via websubmit (submit as HW5).

March 28, 2013

The purpose of this lab is to explore the structure of the SSL certificate hierarchy, and look for complications and/or security vulnerabilities. For some background on the certificates used in SSL, search for information on "X.509 certificates". In this lab, we will look through a snapshot of the SSL observatory, which, roughly, is a "download of all the SSL certificates", as described here:

`https://www.eff.org/files/DefconSSLiverse.pdf`

Objective of this assignment. Your objective in this lab is to explore the SSL observatory using SQL and look for interesting observations. "Interesting observations" is an open-ended requirement. In your writeup, for each observation, please include (1) a description of the observation, (2) its relevance to SSL security/functionality, and (3) the sql queries (or other means) that you used to find it.

Grades not be based on quality, not quantity, but as rough guideline please include at least 6 observations in your report. For example, a result that says "there are XXX valid certificates" is much less interesting than "There are XXX CA certs issued by parties in the US, YY CA certs issued by parties outside the US. [Insert explanation of what a CA cert is, here, and explain how it can be used to sign any certificates that can be trusted by any browser.] The relevance of this is that YYY CA certs are outside US jurisdiction, which means that if they are compromised and used to sign a "malicious" certificate for a domain owned by an American company, it might be difficult to have legal recourse under American law. [Insert SQL query here.]"

Another interesting thing to look at is how many CA certs are signed by issuers that are in a different country than the signer. You can look at my queries below for some more ideas, or the following:

`https://www.eff.org/files/DefconSSLiverse.pdf`
`https://www.eff.org/files/colour_map_of_CAs.pdf`
`https://factorable.net/`

Your assignment can include some of the results I suggested here, or appearing the papers linked above. (Indeed, reproducing some of the results in the papers I reference here is far from trivial!) You can and should also try to add some ideas of your own.

Accessing the SSL observatory. We have downloaded all 17GB of a 2010 snapshot of the SSL observatory to the cs servers. If you have an account on the CS research servers, log in, and run the following command:

```
mysql -h csr-db.bu.edu -D cs558 -u cs558student -p
```

You will then be prompted for a password. The password will be distributed during lecture. If you do not have an account on the cs servers, you should email `support@cs.bu.edu` and CC `goldbe@cs.bu.edu`, indicating that you are a student in CS558 and need access to the database

mysql. You can query the database using mysql. If you have no experience writing SQL queries, do some reading online to learn how to use `SELECT`, `WHERE`, `COUNT`, `GROUPBY`, and `JOIN` to form simple queries. To help get you started, here are some examples.

- `show tables;`
  Shows you the tables in the database. The "roots" table has the root certificates, for example.
- `describe roots;`
  Shows all the fields in the "roots" table.
- `select count(*) from roots;`
  Counts the number of rows in the "roots" table. This tells you how many root certificates where found by the SSL observatory folks.
- `select count(*) from roots where roots.ms_valid='Yes';`
  Counts the number of root certificates that are trusted by "ms" browsers (*i.e.,* by Internet Explorer).
- `select * from roots limit 1;`
  Prints out the entire first row of the "roots" table. That is, print out all the information about the first certificate in the "roots" table.
- `select 'Serial Number' from roots limit 1;` get the serial number of the first cert in the "roots" table.
- `select DISTINCT('X509v3 extensions:X509v3 Basic Constraints:CA') FROM valid_certs;`
  Look at all the different values that 'basic constraints' can be set to in the "valid_certs" table. This field tells us if a certificate is an EE cert or CA cert, and also about pathlen constraints.
- `select count(*), 'X509v3 extensions:X509v3 Basic Constraints:CA' FROM valid_certs GROUP BY 'X509v3 extensions:X509v3 Basic Constraints:CA';`
  This uses `GROUP BY` to report the number of rows (*i.e.,* certs) in the valid_certs table that have a given value in their 'Basic Constraints' field. Using the above two queries, we found that there is a (weird) certificate that has basic constraints "critical) TRUE, pathlen:12", which means they can issue certificate chains of length 12!
- `select * from valid_certs where 'X509v3 extensions:X509v3 Basic Constraints:CA'='(critical) TRUE, pathlen:12';`
  This seemed weird to us, so we print out the cert with above query, so we can see who issued it, and which browsers trust it.
- `select count(*), 'Subject Public Key Info:RSA Public Key:Exponent' FROM valid_certs GROUP BY 'Subject Public Key Info:RSA Public Key:Exponent';`
  This query tells us how many valid certificates have a given RSA exponent.

- ```
  select count(), valid_certs.‘Issuer‘ from valid_certs, roots
  where valid_certs.‘Issuer‘=roots.‘Subject‘
  GROUP BY valid_certs.‘Issuer‘
  HAVING count() > 200;
  ```
  This tells use the number of # valid certificates signed by each root cert.

I look forward to seeing your results!

Submission policy. Every submitted assignment MUST include the following information:

1. List of collaborators
2. List of references used (online material, course nodes, textbooks, wikipedia, etc.)
3. Number of late days used on this assignment
4. Total number of late days used thus far in the entire semester

If any of this information is missing, at least 20% of the points for the assignment will automatically be deducted from your assignment. See also discussion on plagiarism on the course syllabus.

Ack. Thanks to Ben Fuller for helping prepare this assignment.
  ```