Communication Protocols and Internet Architectures
Harvard University
CSCI S-40, Summer 2018

Homework Assignment #4 Solutions

Question 1) 3 points total
The US Computer Emergency Readiness Team (US-CERT) publishes what are called Technical Cyber Security Alerts …… (Please see homework for the complete question.)

ANSWER
The answer to this question depended upon your choice of an Alert or Note. You received full credit if you presented a thorough technical analysis of the problem and the solution.

Question 2) 4 points total
The following questions all relate to email.

a). Explain the structure of the envelope, the header and the body of an SMTP message.
ANSWER
SMTP messages are in ASCII and the three parts that make up an email are the envelope, the header and the body.

The envelope of the email message is separated from the header that follows it by the DATA command (i.e., a single line with DATA on it.) The email message that follows the envelope is composed of two parts, a header and a body. A blank line separates the header from the body when the message is sent. The body (and the entire email message) is terminated by a line with nothing on it except a period.

The envelope is used by the MTAs and has at least three lines. The HELO field specifies the MTA from which the message is being sent. MAIL FROM: and RCPT TO: fields specify the source and destination of the message.

The header is used by the user agents. It contains fields such as: To:, From:, Subject:, Date: and other information useful to the user agents. The information in the header is typically presented as name:value pairs. The body is free-form ASCII text and it contains the email payload. (The header and email payload is what most people think of as the entire email message since they are not aware of the envelope.)

b). Explain how MIME is used to send non-ASCII, binary data (such as images) as an attachment.
ANSWER
MIME (Multipurpose Internet Mail Extensions) is discussed in detail in RFC 1521. Since SMTP was originally designed to use only ASCII data, MIME is used to encode non-ASCII data such as images or audio into ASCII so that it can be included in SMTP messages. The MIME structure contains a header and then the data. The MIME header contains information on the content-type and the encoding method of the data. There are many content-types such as text, multipart, message, application, image, audio, and video. There are many different encoding formats as defined in RFC 1521: 7bit, quoted-printable, base64, 8bit, and binary

As an example, suppose that a user wished to send a jpg image as part of an email message. A popular MIME encoding method to accomplish this would be Base64 encoding. Using this method, three bytes (24 bits) of binary data are encoded into four bytes (32 bits) of ASCII, where each ASCII byte contains six bits of the original data (4x6=24). Of course, the email user agent at the receiving end must be informed where the encoded data is located within the message, what encoding method is used, and the type of data that is being sent (i.e., an image.)

MIME is a standard for doing the above. MIME headers and sub-headers are inserted into the body of the SMTP message, breaking the body up into MIME-header/content pairs that are then interpreted by the user agent.

c). Give 2 examples of the differences between POP and IMAP.

ANSWER

Both POP and IMAP allow users to access and retrieve e-mail messages from a remote mail server. The user's machine does not need to be permanently connected to the Internet in order for the user to receive email. The user has an account on the mail server and on an as-needed basis, the client's machine uses POP or IMAP to connect to the mail server and manage the client's mail. The mail server does maintain a permanent connection to the Internet.

More specifically to the POP protocol, as email messages are sent to a given user, they are stored in a permanent mailbox on the mail server. When the users wish to access their e-mail, they invokes a POP client which establishes a TCP connection to the POP server, authenticates the user, then downloads the stored e-mail messages from the permanent mailbox. In such a configuration, SMTP is used as the e-mail delivery protocol from the sender's system to the user's mail server, and the client uses the POP protocol for e-mail retrieval. Using POP, the user may leave the email on the server or have it deleted. (The original POP protocol did not allow the email to be left on the server.) In either case, the email is always read locally, which means that all management of the email is done on the local machine.

The IMAP (Internet Message Access Protocol) is similar to the POP protocol. That is, both protocols implement permanent user mailboxes on mail servers that have a permanent network connection. Both protocols also use client applications which establish a TCP connection to the server when a user desires to retrieve their e-mail messages. IMAP, however, was designed from the beginning to offer more powerful functionality such as the ability to remotely manipulate e-mail messages on mail servers, the ability to rename messages and selectively download selected e-mail messages or portions of e-mail messages, and the capability to better access and manage e-mail messages stored on multiple hosts.

As an example, IMAP allows users to download partial messages or only message headers, while POP only allows the download of complete messages. Therefore, if you get a large email message with a header that says "EASY MONEY!!!!!" and a large file attached to it, you can delete the email from the server without ever loading it into your client.

d). What is SMTP relaying and why is it not a good idea? How can it be used maliciously?

ANSWER

SMTP relay is when a user from one domain is able to send email through an SMTP server located in a second domain, to an intended recipient, who is located in a third domain. In other words, the intended recipient is not located in the domain of the SMTP server through which the email message is relayed. This is a problem because it consumes resources on the SMTP server that is used as the relay and obscures the true identity of the original sender. SMTP relay is often used by spammers for these reasons.

Question 3) 2 points total

Web browsers have a configuration field for a Proxy Server…. (Please see the homework for the complete question.)

ANSWER

A proxy server acts as an intermediary between the browser client and the web server the client is communicating with. This means that the proxy server receives the request from the client, interprets it (based on some policy) and then acts on it. A proxy server caches web pages that have been requested and this means that if there was a previous request for a specific page, and the page has not expired from the cache, the local copy of the page held by the proxy is returned to the client. This reduces the number of times the same page is requested from a remote web server thereby reducing the traffic on the access line between the company and the Internet. For example, caching the sports page of Boston.com on a

local proxy server would reduce the network traffic to that website. A proxy server can also prevent users from receiving pages from certain web sites based on some policy (set by the administrator of the proxy) by simply not forwarding the request to the web server. The proxy would return some predefined page to the client in place of the requested page. In addition, a proxy server can filter out certain types of content such as streaming video and audio. (Once again, the administrator sets the policy for this.)

Question 4) 4 points total

Assume that you are using a PC at Harvard to connect to a host at MIT and that no machine or DNS at Harvard has ever…. (Please see the homework for the complete question.)

ANSWER

When you enter "www.mit.edu" into your web browser, the browser software in combination with the operating system (OS) will make a request to the OS to see if the IP address for this website is available locally in a cache within the OS. (We state in this question that this is not the case, but local lookups will be done by default by the OS and all the caching DNS servers.) Given it is not in a local cache, the next step will be for the OS to contact the DNS server that has been configured in the OS by the system admin or via DHCP. This request is for an "A" record and on an IPv6 system it would be for a "AAAA" record.

The DNS server that is asked first is typically a local DNS server for the organization and here at Harvard it would be maintained by the local IT group in the department, or the centralized IT group. This DNS server will be responsible for finding the answer to your computer's question and providing a single reply back to your machine with the correct answer. It will do so by making several requests to "zero in" on the actual DNS nameserver that knows the correct answer.

The first step for the local DNS will be to ask one of the root nameservers; the list of these servers is maintained by IANA and changes very rarely. The root nameservers will reply with an "authority section" rather than the answer to the specific query: in short, it will specify what machine to ask next. In this case, it specifies a list of the root nameservers for the EDU TLD.

For example:

```
;; AUTHORITY SECTION:
edu.            172800  IN      NS      d.edu-servers.net.
edu.            172800  IN      NS      l.edu-servers.net.
edu.            172800  IN      NS      f.edu-servers.net.
edu.            172800  IN      NS      a.edu-servers.net.
edu.            172800  IN      NS      g.edu-servers.net.
edu.            172800  IN      NS      c.edu-servers.net.
```

Notice that unlike a caching nameserver, these authoritative nameservers do not continue the search for you. The local DNS then makes a further query to one of those EDU nameservers and that nameserver returns the authority information for MIT.

For example:

```
;; AUTHORITY SECTION:
mit.edu.        172800  IN      NS      bitsy.mit.edu.
mit.edu.        172800  IN      NS      strawb.mit.edu.
mit.edu.        172800  IN      NS      w20ns.mit.edu.
```

The local DNS then contacts one of MIT's nameservers and they return the correct A record for "www.mit.edu":

```
;; ANSWER SECTION:
www.mit.edu.        60      IN      A       18.9.22.169
```


This result is finally provided back to your computer by the local DNS (which is typically a caching nameserver) and the OS on your machine then encapsulates the HTTP request from the browser and sends it to the web server at MIT. Your system is unaware of how many requests were required to find the answer.

(Note that there might be a number of local DNS servers that would be queried in some administratively defined order for the cached A record before any of the root servers were queried, be we don't describe that configuration in this question.)

B) As you know DNS provides the translation between the human-readable names and the associated IP address. Because the same host names and websites are used very frequently by the users within a group or organization, a DNS cache significantly improves the performance of the system since a lookup is not necessary if the name to IP address mapping is cached locally. This is important since it keeps the load on the DNS servers down. Without a cache, the core DNS servers would be overwhelmed by the load of millions of repetitive requests.

DNS caching is controlled by the use of a TTL. This TTL is in seconds and it is set by the authoritative nameserver for the zone and is included in the DNS response. When a caching nameserver receives that response, it will store the IP address mapping and the TTL and begin a count down. Until a point in time when the count reaches zero (the TTL expires), it will return the stored result without querying upstream again. In the event there are multiple levels of caching nameservers, the TTL will be seen by each of them as part of the response to the upstream DNS query. This will ensure that the record expires properly across all servers in the chain at the correct time.

Question 5) 3 points total

5.) Determine the public IP address of your Internet connection at your home, school, or office, and then identify the Autonomous System Number (ASN) that corresponds to your public network address . . . (Please see the homework for the complete question.)

ANSWER

The specific details of your answer will depend upon your IP address and the AS of the ISP that provides your Internet service. For example, I have Comcast at home and my IP address is 24.aaa.bbb.ccc. Comcast's ASN at my location is AS7015. Two upstream providers to AS7015 are AS7922, which is another AS owned by Comcast, and AS3356, which is an AS owned by Level 3 Communications.

Question 6) 4 points total

6.) In February 2015 it became known that Lenovo had pre-installed adware software on some specific models of their computers (running Windows 8) that made the machines very vulnerable to TLS/SSL spoofing attacks without a warning to the user…. (Please see the homework for the complete question.)

ANSWER

The adware software installed by Lenovo was intended to make it possible for third parties to monitor and inject advertisements into encrypted web connections; this was possible because the adware software installed a software web proxy on the machine that allowed a man-in-the middle (MITM) attack. The proxy software intercepted SSL/TLS traffic sent between a user's browser and an intended website, decrypted it, examined it, and then re-encrypted it.

Such MITM attacks would typically result in a browser warning message which would notify the user that they are not communicating directly with the intended website. To prevent this from happening, the adware software needed a method to convince the browser that it is actually communicating with the intended website and not the proxy. To accomplish this, the adware software also installed a root

CA certificate in the system trust store on the affected machines. This allowed it to create fake certificates for the intended websites in real-time. Since the adware software cannot predict all websites that a user might visit, it also needs to sign the fake certificates in real-time in order to make them appear to be trusted. (In other words, to appear to be created by the intended website.) To accomplish this, the adware software also included a private key that was used to sign the certificates that were being created by the proxy. As a result, the certificate created by the adware software would appear to be from the intended website and properly signed, and the user would not be able to tell that their communications had been intercepted by the proxy.

In addition to inserting this proxy, the private key (used for signing the certificates) that was distributed with the adware software was identical on all machines, and was protected by a simple password (and that some reports said was also included within the adware software.) The simple password was quickly broken and widely communicated. Since the private key is now widely known and the proxy is present, affected machines are at risk for several very serious security vulnerabilities, such as: trusting fake websites, downloading infected software which appear to have a valid certificate, and connection eavesdropping and hijacking via rogue WiFi hotspots. This was possible because attackers are able generate trusted certificates for any Internet site that might be accessed by the affected machines. For example, WiFi hotspots typically provide DNS services for machines that are connected to them and a rouge WiFi hotspot could easily send a user that is attempting to connect to their bank's website to a hacker's website instead. The hacker's website would be able to masquerade as the proper website since the hacker's website's certificate would be signed by the well-known private key. This means that the browser on the Lenovo laptop would trust the certificate and the user would be sending personal passwords and data to the hacker's site.

