Computer Networks Exercise 2: DNS (2/2) & HTTP & c-Speed (1/2)

Spring Semester 2019

Answer Sheet

Q1: What is the difference between "naming" and "addressing"? Which problems does DNS solve? Why should we always use trusted DNS servers?

• Naming does not necessarily follow a strict set of rules, while the format of an address is well-defined and can easily be interpreted by a machine.

• It is much easier for users to remember domain names than IP addresses. The same service can have different addresses, which means a service can easily be relocated or accessed using different addresses (load-balancing).

• Imagine you want to go to eth-webshop.ch to buy a present. If you’re using an untrusted DNS server, it can redirect you to an attacker’s IP address instead of ETH’s. The attacker will then imitate ETH, and will serve you a "fake" ETH Webshop page, which will record and steal any information you enter (e.g. credit card information).

Q2: DNS uses BGP Anycast to distribute the traffic among DNS servers close to the destination according to some metric (e.g. lowest RTT or minimal congestion). Every name resolution request is sent using UDP to the nearest DNS server and in case of a packet loss, the UDP request is resent. Imagine a protocol which creates a TCP + TLS connection to the nearest DNS server provided by BGP Anycast and sends name resolution requests using TCP + TLS packets. What are the benefits and problems of this protocol?

• The TCP connection can time out. Either packets are sent regularly to keep the connection alive or the connection is reestablished after an inactivity period.

• The protocol requires the DNS server to keep TCP and TLS state for each client (limits scalability which is an essential property of DNS).
BGP Anycast might modify the route by changing the destination server due to network congestion, which would break the TCP + TLS connection and require reestablishing it.

The protocol would be more secure than DNS, since an attacker would not be able to inject a fake response.

Q3.1: The static "hosts.txt" approach is no longer in use. Explain why by giving at least 3 reasons with motivation.

- Scalability. It is not possible to incrementally update a non-indexed file. Updates would entail copying over the entire file again. A load-balanced hierarchical structure to issue updates is needed.
- Scalability. The file at the current size of the Internet would unnecessarily take too much space because users only visit a fraction of hostnames, which would be especially problematic in low-space machines such as in the Internet of Things.
- Management. Who is assigned which hostname becomes a management issue at a world-wide scale. For example, governments want to control their respective own top level domains.
- Consistency/availability. It is desired to have a stronger guarantee of propagation than "hoping that everyone will adopt my change by downloading an enormous file".

Q3.2: Given the DNS (caching) architecture explained during lecture, what happens from a user perspective when a hostname DNS record removal is issued? And what if the IP address of a hostname is changed?

DNS servers (and the OS/browser*) between the client and the authoritative DNS server cache records. Before the TTL expires, in the first case users will still be directed to the original IP when typing in the hostname. In the second case, they will still be redirected to the old IP address. Once the TTL expires, the records are flushed: in the first case users will no longer be sent to the original IP and will receive a notification that the hostname could not be resolved. In the second case, the users will be redirected to the new IP.

*http://kb.mozillazine.org/Network.dnsCacheExpiration (retrieved March 6th 2018)

Q3.3: Would you employ DNS in the Internet of Things (IoT)? Name an advantage and a disadvantage.

Advantage: it allows the host to be identified by hostname instead of by static IP, as such the destination of its communication can change its IP without
forcing a re-flash of the IoT devices that communicate with it. Disadvantage: too limited power, as such performing DNS queries might be too draining.

Q3.4: Would you use recursive or iterative DNS querying in IoT?

Recursive querying, since IoT devices have only limited computational resources.

Q3.5: How can you employ DNS for censorship? What are its limitations?

By blocking or spoofing DNS requests, which allows an attacker to return nothing or another website (DNS spoofing or poisoning). DNS is only a service that resolves hostnames: by only corrupting DNS queries, it is still possible to directly communicate when the user knows the IP address beforehand.

Q4. URLs

Q4.1: Which of the following URL statements are true, and which are false?

1. The directory_path must map to the underlying file system. (true / false) false
2. Port 80 and 443 are the respective default ports for HTTP and HTTPS. (true / false) true
3. An IP address is a valid hostname. (true / false) true
4. The port numbers enable multiple services to be simultaneously active on the same host. (true / false) true
5. You always need to specify a port number in an URL. (true / false) false

Q4.2: Why are there standard port numbers?

Two answers are possible depending on interpretation of the question: (a) because of convention, it is not necessary for a user to define the port number every time: the default port associated with the protocol (i.e. 80 for HTTP) has become entrenched standard for Web access, or (b) because when a connection is established, the initiator must define which port it wants to connect to and do. The service must already be listening to receive the packets.

Q5. Critical path analysis

Within a dependency graph, the critical path decides the minimum amount of time it takes to fully load all dependencies. The next couple of questions are in relation to the dependency graph shown in Fig. [1]

Q5.1: With $A = 4$, $B = 19$: what is the critical path?
The critical path is the longest path. There are three paths: $\max\{3+5+4+10+10,3+19+10+10,3+10\} = \max\{32,42,13\} = 42$. So path FIN-N1-N3-N5-N6 is the critical path.

**Q5.2:** With $A = 5$, $B \sim \text{Poisson}(7)$: what is the probability FIN-N1-N3-N5-N6 is strictly the critical path?

Let’s call $L_1 = \text{length}(FIN-N1-N3-N5-N6)$, $L_2 = \text{length}(FIN-N1-N2-N4-N5-N6)$ and $L_3 = \text{length}(FIN-N1-N6)$. $L_1$ and $L_2$ are always longer than $L_3$ because edges cannot be negative and $L_1$ and $L_2$ are always strictly more than 13 long.

We want to know:

$$P(L_1 > L_2) = P(3 + B + 10 + 10 > 3 + 5 + A + 10 + 10) = P(B > 5 + A) = P(B - A > 5) = P(B - 5 > 5) = P(B > 10)$$

$$= 1 - P(B \leq k = 10) = 1 - e^{-\lambda} \sum_{i=0}^{\text{floor}(k)} \frac{\lambda^i}{i!}$$

$$= 1 - e^{-7} \sum_{i=0}^{10} \frac{7^i}{i!} = 1 - 0.90148 = 0.09852$$
Q6: Which are the main differences between a GET and a POST request? When should one use GET instead of POST and why?

The main difference between GET and POST methods is defined in terms of form data encoding. With GET, the query parameters are encoded in the URI. On the other hand, with POST the parameters are embedded in the body of the request. This encoding difference has a couple of consequences:

- GET requests can be cached, POST requests can’t (as the POST is not idempotent, see below).
- GET requests remain in the browser history, POST requests don’t (as their form data is not stored).
- GET requests can be bookmarked, POST requests can’t (as their form data is not stored).
- GET requests have length restrictions on its parameters (maximum URL length is enforced), POST requests haven’t.

The official recommendations say that "GET" should be used only in case of idempotent processing, meaning that the user request is a pure data retrieval that will not cause any change in the WEB application state. Furthermore, GET requests should never be used when dealing with sensitive data. This is due to the fact that servers, proxies, and user agents can log the request URI in some place where it might be visible to third parties. In other words, NEVER submit an authentication form with the GET method!

Q7: What does it mean that HTTP is a stateless protocol? How can we force it to keep a state?

HTTP is stateless in the sense that each request is handled independently with respect to the others. Some applications, anyway, need to maintain a state. This is possible through the use of the so-called Cookies. Cookies are small pieces of data, stored in the client browser in the form of key-value pairs. Cookies can also provide authentication: authentication cookies, in fact, are the most common method used by web servers in order to know whether a user is logged in or not, and with which account they are logged in.

Q8: What are the Expires and Last-Modified HTTP response headers used for and how is this achieved?

They are used for caching. The Expires field tells us until when we can cache the page before we have to discard it. The value of Last-Modified can be used in the If-Modified-Since header to tell the server to only return a webpage if it has been modified since Last-Modified.
Q9: The speed of light places a fundamental limit on how fast we can transmit information over the internet. What is the only way to "beat" this speed-of-light-limit?

Place servers closer to the client. Smaller distance means lower latency. This is what forward proxies are used for.

Q10: DNS

Dig is a command line utility that allows you make DNS requests using the command line. In order to solve this exercise you will need to use dig, you can install it through any package manager under linux. If you don’t have access to a linux machine or are unable to install it you can also use the online tool/gui: [https://www.diggui.com](https://www.diggui.com).

Q10.1 To start with, make a simple query to find the IP address of the domain www.ethz.ch using the command `dig www.ethz.ch`. What is the result of the command? What type of record is it and what is this type of record used for? What is the TTL of this record and what does it indicate?

We get back the DNS A record for the domain www.ethz.ch. The record has a TTL of about 5 min (300 sec), but can be lower depending on the time that you requested the DNS entry. It indicates that the DNS record of www.ethz.ch has to get reloaded/requested again every 300 sec, which is how long the local cached copy is considered valid.

Q10.2 What kind of other record types does DNS support? How can we request these through dig? HINT: Use the man-page on the linux system or go to [https://linux.die.net/man/1/dig](https://linux.die.net/man/1/dig) to look up the different possible flags of dig.

DNS supports many types of records: A, AAAA, MX, TXT, CNAME, NS. These are used for the following purposes:

A returns a IPv4 address to access web services.

AAAA returns an IPv6 address, used like the A record for IPv4 addresses.

MX returns the mail server address, used to find the server to send email corresponding to that domain.

CNAME returns an alias domain, the query continues with a lookup of the returned domain.

NS returns the DNS server responsible for that domain.

TXT returns arbitrarily settable, human readable text. This is often used to return authentication information for mail servers, such as DKIM, DMARC,...
These types of records can be requested by simply adding the type at the end of the dig request, like this: `dig www.ethz.ch NS` for the name-servers record.

Q10.3 There are two different types of DNS queries, what are they called and how do they differ from each other? Try out using `+trace` before the domain name. What kind of query type does this show?

There are recursive and iterative lookups. In recursive lookups you ask the DNS server to look up the rest of the domain (again recursively), if it doesn’t have it cached. While in the iterative lookup, the DNS server returns the IP of a DNS server responsible for the next part of the domain. An example of this would be that a DNS server for the TLD `ch` would return the IP address of the `ethz` DNS server for a lookup of `www.ethz.ch`. Recursive DNS queries are often blocked by the servers because they put additional load on them.

The query returned by adding `+trace` is an iterative query, where the result to each of the servers is output.

Q10.4 The `+norecurse` flag also allows the user to indicate that he doesn’t want to request a recursive query. We can now send this query to a specific DNS server by using the `@dns.server` flag. If we add these two to our simple query (`dig @a.root-servers.net +norecurse www.ethz.ch`) and send it to the first root-server, we don’t get an IP in return, why is this? What can we do to resolve this? HINT: `ns1.ethz.ch` is one of ETH’s nameservers.

If we don’t request recursive queries and request the domain from a root server, we don’t get back a result because of the hierarchical structure of the DNS system. The root server just refers us to the next lower level DNS server, the one for the `ch` TLD. We can resolve this by requesting the IP from the responsible DNS server in the first place, in this case the ETH DNS server. In order to achieve this we use the flag `@ns1.ethz.ch` instead. The total query is: `dig @ns1.ethz.ch +norecurse www.ethz.ch`.

Q10.5 `Dig` can also be used to request the inverse query, requesting the domain name from an IP-address, which is called reverse IP lookup and is achieved with the `-x` flag. Which domains do the IP `1.1.1.1` and `8.8.8.8` belong to?

They belong to the Cloudflare DNS Service One.One.One.One (1.1.1.1 or 1.0.0.1) and the Google Public DNS (8.8.8.8 or 8.4.4.8) respectively.