ETH zürich

Networked Systems
ETH Zürich — seit 2015

# Communication Networks

Prof. Laurent Vanbever

**Solution:** Exercise 12 – DNS, Web, and Email

## DNS

### 12.1 Local DNS server

On Linux and Mac computers you can use the command line tool `dig` to perform DNS lookups. The corresponding tool for Windows is `nslookup`. First, perform a lookup for `nyu.edu` using your default DNS server by running the command `dig nyu.edu` or `nslookup nyu.edu`.

- What is the IP address of the server behind `nyu.edu`?

  **Solution:** Note that the actual IP address can depend on the local DNS server you use. We got the following answer with `dig`:

  ```
  ;; ANSWER SECTION:
  nyu.edu.    60    IN    A    216.165.47.10
  ```

  Note that the format is slightly different for `nslookup`:

  ```
  Non-authoritative answer:
  Name:    nyu.edu
  Address: 216.165.47.10
  ```

Now, perform the same lookup, but use one of the DNS root servers (e.g., `a.root-servers.net`) by running[a]

```
dig @a.root-servers.net nyu.edu
nslookup nyu.edu a.root-servers.net
```

[a] If the `nslookup` command does not yield helpful output for you, try adding `-type=soa` after `nslookup`. (For some, it may be the other way round—adding this option may hide the relevant part in the output.)

- Why does the answer differ compared to the one from your local DNS server?

  **Solution:** The request is not sent to an open DNS resolver, but to a DNS server that only provides answers about its own zone. Therefore, the root DNS server only points you to the name servers responsible for the next zone in the hierarchy, the edu zone.

  We got the following answer with `dig`:

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

  For `nslookup`, the output had this format (without `-type=soa`):

  ```
  Served by:
  - a.edu-servers.net
            192.5.6.30
            2001:503:a83e::2:30
            edu
  - b.edu-servers.net
            192.33.14.30
            2001:503:231d::2:30
            edu
  ...
  ```

- How would you proceed with this answer to find the IP address behind `nyu.edu`?

  **Solution:** Now that we know which servers are responsible for the edu zone, we can continue step-by-step just like your local DNS server would. Next, we would send a request to one of the edu name servers:

  ```
  dig @a.edu-servers.net nyu.edu
  ```

  The reply points us to the name servers in charge of the zone of NYU. By sending a request to them, we finally get the IP address behind the URL `nyu.edu`.

## 12.2 Name it or Route it: pick one

In the course, we saw two ways to replicate and load-balance content: *(i)* using Anycast routing; or *(ii)* using DNS.

**a)** List and briefly justify three pros and cons of each;

**Solution:** DNS

(i) Advantages
- i. Simplicity: DNS-based load-balancing can be implemented by any CDN.
- ii. (potentially) Fine-grained: Decisions can be particularized on a per-source IP basis.
- iii. Near real-time: Assuming small TTL values (modulo the problem of load, see below), DNS load-balancing enables frequent load adaption.

(ii) Disadvantages
- i. Infrastructure cost: Good load-balancing requires small TTL values which induce a high load at the DNS server level, which in turn requires to dimension the DNS infrastructure accordingly.
- ii. Location: The source IP seen by the server and on which the load-balancing decision is made is the one from the resolver (e.g. Swisscom's one), not the direct client, meaning the resolver and the client can actually be far away from each other (think of Google's open DNS resolver).
- iii. (potentially) Coarse-grained: The DNS resolver source IP can actually serve a huge amount of clients which will therefore share the same load-balancing decisions.




Any clear winner?

Anycast routing

(i) Advantages
- i. Simplicity: The required infrastructure is simple as the load-balancing is done by the network directly.
- ii. Reliability: Whenever a route fails, the network will simply converge to another replica (route).
- iii. Expandablility: It is easy to add an additional replica, as simply an additional location needs to start advertising the respective prefix.

(ii) Disadvantages
- i. Broken Connections: As routing changes can happen at anytime, packets of the same TCP flow might arrive at different replicas effectively breaking the connection.
- ii. Content and performance: The load-balancing is not aware of the utilization of the replicas. Hence, it is not possible to offload work from heavily utilized replicas.
- iii. Location: Packets may not use the nearest replica due to routing policies.

**b)** We saw that CDNs often rely on DNS for distributing their load. Could they also use Anycast routing instead? Explain why or why not.

**Solution:** Yes, you can use Anycast routing instead to distribute the load. However, the load-balancing is neither content nor performance aware. You have to make sure that the same content is available on all replicas and none of the replicas is fully utilized.

# Web and Email

## 12.3 HTTP host header

Perform a DNS lookup for `google.ch` and open `http://172.217.168.35` in your browser. What do you observe?

Now try to repeat the same process for `nsg.ee.ethz.ch` and `comm-net.ethz.ch`. Open the websites in your browser using the IP(s) from the DNS lookup. Do you see the expected websites?

Normally, one machine can host multiple websites at the same time. To distinguish which website has to be provided by the server, clients can add a so called "host header" in their HTTP request which specifies the website they want to access. You can try that yourself with the two websites from above. For example with the following commands:

```
telnet comm-net.ethz.ch 80

GET / HTTP/1.1
Host: comm-net.ethz.ch
```

Do you see another way how you could host multiple websites on the same machine? Can you see potential problems with this approach compared to the host header?

**Solution:** You could assign the server multiple IP addresses and link each IP address to a single website. The biggest drawback of this solution is the need for multiple IP addresses. IPv4 addresses are limited and hence expensive.

## 12.4 E-mail

Answer the following questions about e-mail with True or False and justify your choice.

**a)** SMTP and IMAP can be used to forward e-mails from one e-mail server to another one.

**Solution:** False, the Simple Mail Transfer Protocol (SMTP) is mainly used to forward e-mails from the e-mail client to the server or between servers. The Internet Message Access Protocol (IMAP) is one possible protocol for the client to retrieve e-mails from the server.

**b)** Looking at the header of a received e-mail, you can reconstruct through which e-mail servers the message was forwarded.

**Solution:** True, every e-mail server adds a received entry to the header.

**c)** IMAP is the encrypted counter-part of POP.

**Solution:** False, IMAP is like POP a protocol for a client to retrieve e-mails from the server. IMAP has more features than POP. For example, it allows to download e-mails partially and to connect multiple clients to the same mailbox.

**d)** It is not possible to verify that the e-mail was actually sent by the given FROM address.

**Solution:** True, no checks are performed to verify that the sender is authorized to send e-mails on behalf of that address.

**e)** The IP address of the mail server of a domain can be found by issuing a DNS query asking for the A record of that domain.

**Solution:** False, a mail server is identified using a DNS query asking for MX records (e.g., `dig MX ethz.ch`).

**f)** Images attached to an e-mail are transformed to text for transmission.

**Solution:** True, as e-mail relies on 7-bit U.S. ASCII, all non-English text and binary files have to be encoded in 7-bit U.S. ASCII. For this purpose MIME is used.

## 12.5 E-Mail analysis (Exam Style Question)

You received an email with the raw content shown in Figure 2.

```
Received: from edge20.ethz.ch (82.130.99.26) by CAS10.d.ethz.ch
 (172.31.38.210) with SMTP Server (TLS) id 14.3.408.0; Thu, 2 Aug
 2018 11:17:27 +0200
Received: from phil2.ethz.ch (129.132.65.3) by edge20.ethz.ch (82.130.99.26)
 with SMTP Server id 14.3.408.0; Thu, 2 Aug 2018 11:17:23 +0200
Received: from filter.spam.ch ([5.152.185.154]  helo=filter.spam.ch)
 by phil2.ethz.ch with esmtps  (TLSv1:AES128-SHA:128)   (Exim 4.69)
 (envelope-from <john.doe@anonymous.ch>)        id  1fl9jO-0004C9-7T
 for lvanbever@ethz.ch; Thu, 02 Aug 2018 11:17:15 +0200
X-Note: This Email was scanned by filter.spam.ch
Received: by filter.spam.ch with PIPE id
 93122453; Thu, 02 Aug 2018 11:17:13 +0200
Received: from [10.40.0.131] (HELO smtp.ch.exg7.mailhost.com)  by
 filter.spam.ch with ESMTPS id 93122443
 for lvanbever@ethz.ch; Thu, 02 Aug 2018 11:17:10 +0200
Received: from exg7.mailhost.local (192.168.40.105) by
 exg7.mailhost.local (192.168.40.107) with SMTP Server
 (version=TLS1_2, cipher=TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA256_P256) id
 15.1.1531.3; Thu, 2 Aug 2018 11:17:09 +0200
From: Anonymous Student <john.doe@anonymous.ch>
To: Laurent Vanbever <lvanbever@ethz.ch>
Subject: Exam solutions
Date: Thu, 2 Aug 2018 09:17:09 +0000
Message-ID: <11A5442F-4D6E-436F-A873-2E3DA3656C06@anonymous.ch>
Accept-Language: de-CH, en-US
Content-Language: en-US
Content-Type: text/plain; charset="us-ascii"
Content-ID: <F43C7219ADA84040984B4640587C2B70@fwd7.mailhost.com>
Content-Transfer-Encoding: quoted-printable
MIME-Version: 1.0

Hey, can you give me the solutions for the exam?
```

Raw content of a received email

**a)** According to Figure 2, what are the e-mail addresses of the sender and the receiver of this message?

**Solution:**

- Sender: john.doe@anonymous.ch
- Receiver: lvanbever@ethz.ch

**b)** List the IP addresses of all servers that have seen this email according to Figure 2 in chronological order starting with the server that saw the email *first*.

**Solution:**

- 192.168.40.105 (exg7.mailhost.local)
- 192.168.40.107 (exg7.mailhost.local)
- 10.40.0.131 (smtp.ch.exg7.mailhost.com)
- 5.152.185.154 (filter.spam.ch)
- 129.132.65.3 (phil2.ethz.ch)
- 82.130.99.26 (edge20.ethz.ch)
- 172.31.38.210 (CAS10.d.ethz.ch)

**c)** According to the header in Figure 2, the email passed a spam filter (`filter.spam.ch`). Could one of the other servers have added this entry without the email actually passing `filter.spam.ch`? If yes: why and which server(s) could have done it? If no: why not?

**Solution:** Yes other servers could have added the entry as the header is not authenticated. All the servers that see the email after `smtp.ch.exg7.mailhost.com` can add the entry (i.e. `phil2.ethz.ch`, `edge20.ethz.ch`, `CAS10.d.ethz.ch`)

**d)** Which servers (according to Figure 2) could modify the email message ("Hey, ..."))? Why?

**Solution:** All of the servers could have modified the message as it is not encrypted or signed.

**e)** Assume you have `telnet` access to an open SMTP server that does not appear in Figure 2 and you want to fake the email shown in Figure 2. That is, your goal is that the receiver of the email in Figure 2 receives the same email again (with the same sender). Which parts of the email in Figure 2 can you replicate in your email and which parts will be different? Use the line numbers in Figure 2 to list parts that are equal or different in your email and briefly explain the reasons why they are equal or different.

**Solution:** You can replicate everything from line 20 and below but not the headers because one cannot influence where the server will send the email next to.