Application Layer
Jean-Yves Le Boudec
2017
Contents

1. The Application Layer, TLS, QUIC
2. The Domain Name System
3. Application Layer Gateways
4. IPv4 / IPv6
5. ALG46

Textbook

Chapter 2: The Application Layer
1. The Application Layer

The application layer of TCP/IP consists of

- the distributed applications themselves – it is the topic of the courses on information systems and distributed systems
- a number of generic intermediate “layers”

You did several application layers in the socket programming lab.

In this module, we focus on

- the relationship between the application layer and the lower layers
- the generic intermediate layers
Example: World Wide Web (WWW)

HTTP (hyper text transfer protocol) uses TCP
HTTP/1 uses one TCP connection per object, HTTP/2 allows to use the same connection for several objects
A server sends us a web page with 10 embedded objects (images, scripts, etc) using http/2 over one single TCP connection. Our browser displays the elements of the page as soon as they arrive, without waiting for the page to be complete.

One packet of the first object is lost (and will be retransmitted).

A. The second object can be displayed before the lost packet is repaired because HTTP/2 multiplexes objects
B. The second object must wait for the first object to be completely received
C. It depends which version of TCP is used
D. I don’t know
Which of these port numbers are TCP server ports?

A. 20
B. 21
C. 12346
D. 20 and 21
E. 20 and 12346
F. 21 and 12346
G. all of these three
H. none of these three
I. I don’t know
Solution

• Answer B
The objects are transferred over a single TCP connection; whatever the version, TCP offers a streaming service with in-sequence delivery. Object 2 cannot be delivered to the app before object 1 (even if all packets that contain object 2 are received). This is the head-of-the-line blocking drawback of TCP

• Answer F
Port 20 used on the FTP server is not used in TCP server mode (the listen call is done on port 12346 on the FTP client side. An (application) server is not always a (TCP) server!
Transport Layer Security (TLS)

TLS adds encryption and authentication to TCP.
TLS requires a handshake to
- authenticate the server side with digital certificates
- negotiate the security parameters (crypto algorithms)
- create the (temporary) secret keys

http used over TLS and port 443 = https
With https/TLS1.2, how many handshakes are required before data transfer can occur?

A. 1  
B. 2  
C. 3  
D. 4  
E. 0  
F. I don’t know
Solution

Answer C

1 handshake for opening the TCP connection
2 handshakes (at least) for establishing the TLS session

(here we consider that a handshake is one or several messages in sequence, followed by a response of one or several messages).

With TLS 1.2 it takes at least 3 RTTs before communication starts (assuming no packet loss !)
TLS 1.3

Reduces number of handshake messages (1-RTT for full handshake)

Cached pre-shared keys can be re-used from one TLS connection to the next, as long as valid (0-RTT)

issue: replay attacks may be possible
solution: only send data that does not cause problem with replay (e.g. HTTP GET)
The following text, from [1], shows an example of replay attack on 0-RTT session establishment.

The attack is not possible with 1-RTT handshake because ServerHello contains a random number.

[1] Eric Rescorla, post on https://www.ietf.org/mail-archive/web/tls/current/msg15594.html,
**Why**? Avoid latency and head-of-the line blocking of TLS over TCP

**How**? All in one!

QUIC opens one TLS 1.3 session first

QUIC runs over UDP

Data is always secured (with TLS 1.3 by default).
QUIC versus TCP with TLS1.3 (1-RTT case)

QUIC establishes a TLS1.3 session first.
Server sends a connection id to client. Can be used to support change of IP address.
QUIC assumes MTU \( \geq 1280 \) B.
Implemented in google apps (chrome, youtube app, etc.) where it replaces TCP (try it: wireshark or chrome://net-internals)
QUIC Packets, Streams and Frames

One QUIC session (also called “connection”) has multiples streams. Stream can be reliable (like TCP) or not. For a reliable stream, data is delivered in sequence. For an unreliable stream, data may be delivered out-of-sequence (with indication of offset). Streams can be created and deleted on-the-fly.

Stream 0 is for TLS1.3 and connection management.

One QUIC packet has a unique (increasing) packet number (64bits) – never wraps. A retransmitted packet has a different number. A packet contains frames for stream data and other data.

<table>
<thead>
<tr>
<th>QUIC hdr: Connection Id, Packet number</th>
<th>stream 0 crypto handshake data</th>
<th>ACK</th>
</tr>
</thead>
<tbody>
<tr>
<td>stream 1 app data</td>
<td>stream 2 app data</td>
<td>stream 3 app data</td>
</tr>
</tbody>
</table>
Reliable and Unreliable Streams

QUIC obtains reliability similar to TCP: missing packets are detected and retransmitted.

Packet loss detection incorporates the best known algorithms of TCP such as Fast Retransmit, RACK, Tail Loss Probe etc.

ACK frames contains up to 256 ack blocks (compare to TCP: 3 blocks).

Flow control is both per stream and per connection, using an explicit offset (instead of window).

Congestion control is similar to TCP Reno or Cubic, per connection (not per stream).

Unreliable streams do not have packet retransmission (in principle) but are subject to flow control and to congestion control.
2. The Domain Name System (DNS)

**Why** invented?
- support user friendly naming of resources: computers, printers, mailboxes, ...
- hide IP address changes on servers

**What** does it do?
- map DNS names (ex: ssc.epfl.ch) to IP addresses
every node on the tree represents one or a set of resources

- every node on the tree has a label `lrcsuns` and a domain name `lrcsuns.epfl.ch`

label is made of 1 to 63 characters `a-z, 0-9 or -`

- Other characters are allowed but real name is translated with *punycode*
  
  *Ex:* `www.öpfl.ch` is in fact `www.xn—pfl-rna.ch`
How Does DNS Work?

When Lisa’s machine needs to map name to IP address

- DNS resolver in Lisa’s machine contacts a DNS server
- IP address of DNS server is known to Lisa’s machine at configuration time
- DNS server may not know answer: in such a case, DNS server needs to do several iterations, as shown on next example.
- A cache is used at DNS resolver and at DNS server to avoid repeating the same requests frequently.

DNS uses UDP for queries and responses (in principle)
Lisa clicks:
http://www.zurich.ibm.com/RZ.html

1
---
query, RD=yes

2,4
---
query, RD=no

3
---
answer
answer = “”
    NS ns.austin.ibm.com.
    NS ns.almaden.ibm.com.”
additional=“watson.ibm.com. A 192.35.232.34
    ns.austin.ibm.com. A 129.34.139.4
    ns.almaden.ibm.com A 198.4.83.134”

5,6
---
answer
answer = “www.zurich.ibm.com. 7200 A 193.5.61.131”
The previous slide shows an example of name resolution.

1. an application on Ircsuns requests a name resolution (find the IP address of www.zurich.ibm.com), a request is sent to the name server configured at Ircsuns.
2. the epfl name server does not know the answer, but, as any name server, knows the IP address of root name servers.
3. a root name server knows the IP addresses of all level-2 domains. Thus, it informs Ircsuns of the IP address of the name servers responsible for the ibm.com domain.
4. the epfl name server sends the same request now to the ibm name server.
5. the ibm name server gives the IP address of www.zurich.ibm.com back to the epfl name server. The epfl name server keeps the address in its cache, this will be used if the same request comes again.
6. the epfl name server gives the IP address of www.zurich.ibm.com back to Ircsuns. End of the resolution!

The request sent by Ircsuns is recursive (RD=yes): Ircsuns will receive only the final answer. In contrast, the request sent by the epfl name server is iterative (RD=no): it receives only partial answers that help towards the solution.
A (=IPv4) and AAAA (= IPv6) records

C:\Users\leboudec> dig a lca.epfl.ch

;; global options: printcmd
;; Got answer:
;; - >>HEADER<<- opcode: QUERY, status: NOERROR, id: 652
;; flags: qr aa rd ra; QUERY: 1, ANSWER: 2, AUTHORITY: 0, ADDITIONAL: 0

;; QUESTION SECTION:
lca.epfl.ch. IN A

;; ANSWER SECTION:
lca.epfl.ch. 86400 IN CNAME lca1srv2.epfl.ch.
lca1srv2.epfl.ch.86400 IN A 128.178.156.24

C:\Users\leboudec> dig aaaa lca.epfl.ch

;; global options: printcmd
;; Got answer:
;; - >>HEADER<<- opcode: QUERY, status: NOERROR, id: 415
;; flags: qr aa rd ra; QUERY: 1, ANSWER: 2, AUTHORITY: 0, ADDITIONAL: 0

;; QUESTION SECTION:
lca.epfl.ch. IN AAAA

;; ANSWER SECTION:
lca.epfl.ch. 86400 IN CNAME lca1srv2.epfl.ch.
lca1srv2.epfl.ch.86400 IN AAAA 2001:620:618:19c:1:80b2:9c18:1
A and AAAA records map name to IP address
CNAME record maps name to name

lca.epfl.ch. 86400 IN CNAME lca1srv2.epfl.ch.

Here lca1srv2.epfl.ch is the canonical name
lca.epfl.ch is an alias
Caching and Time to Live

when receiving message 5, stisun1.epf.ch keeps answer from watson.ibm.com in a cache and does TTL = 7200 s for this cache entry. TTL decreases by 1 every second. When TTL = 0, cache entry is discarded by lrcsuns

During next 2 hours, further requests are answered directly by lrcsuns

watson.ibm.com is the “authoritative” DNS server for this record (i.e. the origin) -- decides value of TTL to be used by caching servers
Lisa clicks:
http://www.zurich.ibm.com/RZ.html

1mn later, Bart clicks:
http://www.zurich.ibm.com/RZ.html

7 query, RD=yes

8 answer
answer = “www.zurich.ibm.com. 7140 A 193.5.61.131”
Reverse DNS Lookup

Reverse Query = find name, given some IP address; used e.g. by traceroute

<table>
<thead>
<tr>
<th>Hop</th>
<th>RTT</th>
<th>RTT</th>
<th>RTT</th>
<th>Host Name</th>
<th>IP Address</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2 ms</td>
<td>&lt;1 ms</td>
<td>&lt;1 ms</td>
<td>cv-ic-dit-v151-ro.epfl.ch</td>
<td>128.178.1</td>
</tr>
<tr>
<td>2</td>
<td>&lt;1 ms</td>
<td>&lt;1 ms</td>
<td>&lt;1 ms</td>
<td>cv-gigado-v100.epfl.ch</td>
<td>128.178.10</td>
</tr>
<tr>
<td>3</td>
<td>&lt;1 ms</td>
<td>&lt;1 ms</td>
<td>&lt;1 ms</td>
<td>c6-ext-v200.epfl.ch</td>
<td>128.178.200.1</td>
</tr>
<tr>
<td>4</td>
<td>&lt;1 ms</td>
<td>&lt;1 ms</td>
<td>&lt;1 ms</td>
<td>swiel2.epfl.ch</td>
<td>192.33.209.33</td>
</tr>
<tr>
<td>5</td>
<td>18 ms</td>
<td>&lt;1 ms</td>
<td>3 ms</td>
<td>swiLS2-10GE-1-2.switch.ch</td>
<td>130.59.1</td>
</tr>
<tr>
<td>6</td>
<td>4 ms</td>
<td>4 ms</td>
<td>4 ms</td>
<td>swiE81-10GE-2-7.switch.ch</td>
<td>130.59.1</td>
</tr>
</tbody>
</table>

Lab 1a

Tracing route to read.more.at.beaglenetworks.net [216.81.59.173] over a maximum of 64 hops:
How does traceroute do DNS lookup?

A. DNS client tries all possible DNS names and verifies each of them with DNS server

B. DNS client asks DNS server: *what is the name corresponding to this IP address?* and DNS server searches its database to find an answer

C. Traceroute asks intermediate routers: *what is your DNS name?*

D. None of the above

E. I don’t know
Solution: Reverse DNS Lookup

Reverse Query = find name, given some IP address; used e.g. by traceroute
DNS servers do not do reverse search on A and AAAA records
Instead, reverse records (PTR) are used
IP addresses are first translated to domain names
128.178.151.1 -> 1.151.178.128.in-addr.arpa

in-addr.arpa = top level domain reserved for IPv4 addresses
ip6.arpa = top level domain reserved for IPv6 addresses

the reverse and direct records are written independently, consistency is expected but not enforced
we are looking for the name corresponding to the IPv4 address 128.178.151.1

we verify that the IPv4 address for this name is as expected

it is as expected, i.e. reverse record (PTR) and direct record (A) are consistent
we are looking for the name corresponding to the IPv4 address 206.214.251.89

the name is found from a PTR record in DNS

we verify that the IPv4 address for this name is as expected

this DNS name is bogus, there is no A (nor AAAA) record for it. The PTR record is bogus.
DNS authority

DNS authority is hierarchical.

Top level authority allocates top-level domains (e.g. .ch)
The top level is currently run by PTI (Public Technical Identifiers – formerly ICANN)
The 13 top-level DNS servers (root servers) are run by multiples companies

National registry allocates .ch domain names (in Switzerland: SWITCH)
Registrar is a commercial organizations accredited by registry to sell names (e.f. infomaniak.ch, register.ch, switchplus.ch)
DNS servers are run by ISPs or other organizations

namecoin is an attempt to provide peer-to-peer ( = without authority ) DNS names for the domain .bit. It uses the blockchain technology.
DNS and Load Balancing

Why?
sovkom.com is very popular and hosted on many servers. Traffic should be split between sites

How?
One solution is to use a load balancing DNS server:

- sovkom.com is mapped to many different addresses (or CNAMEs)
- load balancing DNS server chooses one address (based on load, response time...)
- TTL is very short to allow for frequent changes

```
sovkom.com 300 AAAA 2001: babe::b0b2
sovkom.com 300 AAAA 2001: b88b::b0bc
sovkom.com 300 AAAA 2001: b55b::b0ba
...```
DNS alone is unsecure even if servers are trusted anyone can send (incorrect) answers to DNS queries

DNSSEC solves this problem
► Records are signed, using public key cryptography
► Chain of trust is initialized with root servers

► Some domains (such as .se) implement it systematically
► Some DNS servers (Google’s 8.8.8.8 and 8.8.4.4) implement DNSSEC systematically

Alternative: DNS over TLS / over QUIC
Bonjour = Server-less DNS

Why invented?
For names of local importance only: e.g. your printer at home

hp-laser2719.local

top level domain local is reserved for names visible in subnet

Uses multicast address 224.0.0.251/ff02::fb, UDP port 5353
Works only in one single (possibly bridged) LAN
We change the IPv4 address of lca.epfl.ch. How long does it take for the whole world to be informed of the change?

A. less than a minute
B. Up to 6 hours
C. Up to 1 day
D. Up to 2 days
E. None of these
F. I don’t know

C:\Users\leboudec> dig a lca.epfl.ch
<<< DiG 9.3.2 <<< a lca.epfl.ch
;; global options: printcmd
;; Got answer:
;; ->>>HEADER<<- opcode: QUERY, status: NOERROR, id: 6634
;; flags: qr aa rd ra; QUERY: 1, ANSWER: 2, AUTHORITY: 0, ADDITIONAL: 0

;; QUESTION SECTION:
;lca.epfl.ch.

;; ANSWER SECTION:
lca.epfl.ch. 86400
lca1srv2.epfl.ch.86400

EPFL names have TTL= 86400 secs, i.e. 1 day
Solution

EPFL names have TTL= 86400 secs, i.e. 1 day
It may take up to one day for the change to take effect

```bash
C:\Users\leboudec> dig a lca.epfl.ch
<<< DiG 9.3.2 <<< a lca.epfl.ch
; global options: printcmd
; Got answer: 
; ->>HEADER<<- opcode: QUERY, status: NOERROR
; flags: qr aa rd ra; QUERY: 1, ANSWER: 4, AUTHORITY: 0, ADDITIONAL: 0

; QUESTION SECTION:
;lca.epfl.ch.

; ANSWER SECTION:
lca.epfl.ch.     86400
lca1srv2.epfl.ch.86400
```
3. Application Layer Gateways (ALGs)

Definition: an *application layer gateway* is an application layer intermediate system. It terminates the TCP connections (if the application layer uses TCP) and does “store and forward” for the application layer data.

**Example:** HTTP gateway, also called “web proxy”

- A sends HTTP request to gateway, gateway sends another HTTP request to server, server sends objects to gateway, gateway sends objects to A
- Gateway terminates the TCP connections and does “store and forward”
Web Proxy can be deployed to reduce traffic

HTTP Intermediate Systems can keep frequently asked documents close to user

- requested files are kept in a cache
- similar systems deployed in content distribution networks
The “End-to-end” Principle

The “end-to-end” principle of the Internet says that any layer above the network layer should avoid intermediate systems.

Thus: Application Layer Gateways should be avoided

Why?

- Simplify the network. The network is independent of applications and can be run more safely.
- Allow easy deployment of applications. Ex: the web was deployed in 1994 in a few months. Before that, TCP/IP existed, but not HTTP.
- Performance is better – no store and forward
The “End-to-end” Principle for Email

Q. what would a strict application of the end-to-end principle on the figure give?
Q. what would a strict application of the end-to-end principle on the previous figure give?

A. PC A should open a TCP connection directly to B and transfer the email over the connection. This is not possible here since PCs are not expected to be always available for service, as an email server is. However, one could require that A directly opens a TCP connection to email server Y instead of going through email server X.
The End-to-end Principle is not always Applicable

Application layer gateways are still desirable in some cases.

Q. Can you mention three good reasons for desiring an application layer gateway?
For Example

1. **Mobility** (or partial connectivity). On the previous figure, this is why we send email to Y and not to B.

2. **Security / Access Control.** X connects to the internet through web proxy S. S can deny access to some web sites.

3. **Interworking.** Interworking between IPv4 and IPv6.

4. **Performance.** Counter TCP deficiencies on wireless links, Load balancers
Server-side Load balancer are application-layer gateways: they terminate TCP or QUIC connections and re-direct user request to one of the application servers.

Load balancers are replicated for reliability.
How is user traffic split among load balancers?
Solution

DNS load balancing is possible but is perhaps better adapted to splitting traffic across multiple sites.

Equal Cost Multipath Routing is typically used in the access routers: with this solution, the IP address of the service is allocated to both Load Balancers 1 and 2. The access routers route randomly to one of the two load balancers using ECMP and per-flow load balancing.
Firewalls use ALGs

Firewall = a system that separates Internet from intranet: all traffic must go through firewall; only authorized traffic may go through; firewall itself cannot be penetrated (as one thinks)

Firewall =
one dual homed application layer gateway

Firewall =
several gateways + “sacrificial” subnet separated by filtering routers
How many TCP connections are there on the figure (if TCP is used by Lisa and Bart)?

A. 1
B. 2
C. 3
D. 4
E. 5
F. 6
G. I don’t know
Answer D

There are 2 TCP connections for Lisa and 2 for Bart. The load balancer breaks the TCP connection and acts as an Application Layer Gateway.
4. The Application Layer has to select IPv4 or IPv6 sockets

Application layer programs must be explicitly choose IPv4 or IPv6 sockets
Old programs can use only IPv4; newer ones can use IPv4 and IPv6 sockets
Note that application layer protocols (such as HTTP) are independent of the version of IP (IPv4/IPv6)
A Dual Stack Host with up-to-date application code can use both IPv4 and IPv6

Will Lisa use IPv4 or IPv6 to connect to lca.epfl.ch?
Solution

Rules:
- If both are v4/v6, choose v6 (except when 6to4 addresses are used – see later)
- Otherwise do what is possible
  - v4only to v6only is impossible
How does Lisa know whether other end is v4 or v6?

A. Lisa’s PC sends a ping to newstuff.epfl.cn
B. Lisa’s PC tries http over both IPv4 and IPv6 in parallel and sees what works
C. None of the above
D. I don’t know
**Solution:** DNS is used by application layer to know if other end is IPv4 or IPv6

Lisa’s PC DNS resolver asks for both AAAA and A records
If AAAA record obtained, use IPv6
Else if A record obtained, use IPv4
Else error (unknown host)
**DNS can be accessed via IPv4 or IPv6**

Lisa’s PC is configured to talk to a DNS server using IPv4

Can ask A as well as AAAA questions

Lisa’s PC could be configured to use v6 for speaking to DNS server, if a v6 DNS servers is accessible to Lisa’s PC
IPv4 and IPv6 are incompatible

- v4 only host cannot handle IPv6 packets
- v6 only host cannot handle IPv4 packets

What needs to be solved:

Interworking (h4 to h6): allow IPv6-only hosts and IPv4-only hosts to communicate

- example: IPv6 host connects to an IPv4 web server

like to like access

- 6 to 6 over IPv4 infrastructure; ex: IPv6 host at home connects to IPv6 server at EPFL
- 4 to 4 over IPv6; in a distant future

In this module we study interworking;
In a later module we will study like-to-like
Dual Stack Application Layer Gateways (ALG46s) can be used to solve h4 to h6 Interworking

Application layer gateway (e.g. web proxy) relays HTTP questions / answers.
ALG must be dual stack
How does Homer’s PC know it should go to the ALG instead of the final web server?
Solution: via DNS

Homer's PC

Application
TCP
IPv6

www.epfl.ch

www6.epfl.ch

ALG46
TCP
IPv6
IPv4

DNS server

QUESTION
www.epfl.ch
type = AAAA

ANSWER
www.epfl.ch
AAAA

www.epfl.ch
128.178.50.12

Application
TCP/IP
IPv4
Bart’s PC will not use the ALG
A. Yes
B. No
C. It depends on the configuration
D. I don’t know
Solution

What the system does depends on its configuration.

Some systems (Windows/Linux allow to set the preference - by default: prefer IPv6). If this is the case, Lisa will connect to via IPv6 (and is treated less well!). Some other systems (some versions of OSX) try to dynamically select the best choice based on response time.
Pros and Cons of ALG46

+
Transparent to end-users (no software change)
No change in server software – easy deployment

−
This is store and forward... therefore the performance is not good / or the cost is high (disks, CPUs)
One ALG logic for every service (http, email, etc)
Will Joe’s PC use IPv4 or IPv6 to connect to lca.epfl.ch?

A. IPv4  
B. IPv6  
C. It depends on the configuration of Joe’s PC  
D. I don’t know
Solution

Joe’s PC can use either IPv4 or IPv6; DNS queries will provide both an A and an AAAA answer. What the system does depends on its configuration, as discussed earlier with Lisa.
Does Joe’s PC use IPv4 or IPv6 to contact the DNS server?

A. Always IPv4
B. Always IPv6
C. It depends on the configuration of Joe’s PC
D. I don’t know

QUESTION:
lca.epfl.ch.
type = AAAA
type = A

ANSWER:
lca.epfl.ch.
A 128.178.156.24
Facts to Remember

Application layer runs on hosts, not routers
Application layer gateways (ALGs) should be avoided whenever possible (“end to end principle”) but are deployed e.g. for load balancing / security
Application chooses UDP or TCP and must be able to use both IPv4 and IPv6; old apps use IPv4 only
DNS is a worldwide distributed data base used for mapping names to IP addresses (and vice versa)
DNS requires DNSSEC to be secure
DNS is used by apps on dual stack machines to know whether to use v4 or v6
ALG46s can be used to solve the h4 to h6 interworking problem