Lehrstuhl für Betriebssysteme
Fakultät für Informatik
Technische Universität München

Grundlagen Rechnernetze und Verteilte Systeme

IN0010, SoSe 2018

Übungsblatt 10

25. Juni – 29. Juni 2018

Hinweis: Mit * gekennzeichnete Teilaufgaben sind ohne Lösung vorhergehender Teilaufgaben lösbar.

Aufgabe 1 Fluss- und Staukontrolle bei TCP

Das im Internet am weitesten verbreitete Transportprotokoll ist TCP. Dieses implementiert Mechanismen zur Fluss- und Staukontrolle.

a)* Diskutieren Sie die Unterschiede zwischen Fluss- und Staukontrolle. Welche Ziele werden mit dem jeweiligen Mechanismus verfolgt?

- **Flusskontrolle:**
  Verhinderung von Überlastsituation beim Empfänger
- **Staukontrolle:**
  Reaktion von Überlastsituation im Netz

b) Ordnen Sie die folgenden Begriffe jeweils der TCP-Fluss- bzw. Staukontrolle zu:

- Slow-Start
- Empfangsfenster
- Congestion-Avoidance
- Multiplicative-Decrease

Zur Flusskontrolle gehört lediglich das Empfangsfenster, da über dieses der Empfänger dem Sender mitteilen kann, wie viel Daten er maximal auf einmal senden darf.

Die übrigen Begriffe gehören alle zur Staukontrolle, wobei Slow-Start und Congestion-Avoidance die beiden Staukontrollphasen einer TCP-Verbindung sind. Als Multiplicative-Decrease hingegen bezeichnet man das Halbieren des Staukontrollfensters bei Verlust eines Segments.

Zur Analyse der mit TCP erzielbaren Datenrate betrachten wir den Verlauf einer zusammenhängenden Datenübertragung, bei der die Slow-Start-Phase bereits abgeschlossen ist. TCP befinde sich also in der Congestion-Avoidance-Phase. Wir bezeichnen die einzelnen Fenster wie folgt:

- Sendefenster W_s, $|W_s| = w_s$
- Empfangsfenster W_r, $|W_r| = w_r$
- Staukontrollfenster W_c, $|W_c| = w_c$

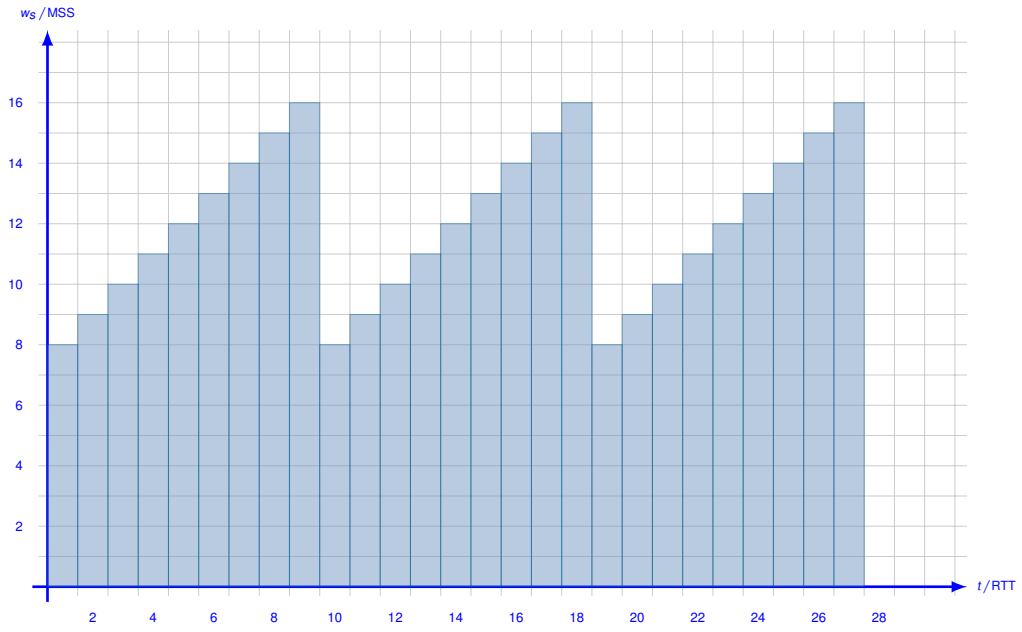
Wir gehen davon aus, dass das Empfangsfenster beliebig groß ist, so dass das Sendefenster allein durch das Staukontrollfenster bestimmt wird, d. h. $W_s = W_c$. Es treten keinerlei Verluste auf, solange das Sendefenster kleiner als ein Maximalwert x ist, also $w_s < x$.

Wird ein vollständiges Sendefenster bestätigt, so vergrößert sich das aktuell genutzte Fenster um genau 1 MSS. Hat das Sendefenster den Wert x erreicht, so geht genau eines der versendeten TCP-Segmente verloren. Den Verlust erkennt der Sender durch mehrfachen Erhalt derselben ACK-Nummer. Daraufhin halbiert der Sender das

Staukontrollfenster, bleibt aber nach wie vor in der Congestion-Avoidance-Phase, d. h. es findet kein erneuter Slow-Start statt. Diese Vorgehensweise entspricht einer vereinfachten Variante von TCP-Reno (vgl. Vorlesung).

Als konkrete Zahlenwerte nehmen wir an, dass die maximale TCP-Segmentgröße (MSS) 1460 B und die RTT 200 ms beträgt. Die Serialisierungszeit von Segmenten sei gegenüber der Ausbreitungsverzögerung vernachlässigbar klein. Segmentverlust trete ab einer Sendefenstergröße von $w_s \geq x = 16$ MSS auf.

c)* Erstellen Sie ein Schaubild, in dem die aktuelle Größe des Sendefenster w_s gemessen in MSS über der Zeitachse t gemessen in RTT aufgetragen ist. In Ihrem Diagramm soll zum Zeitpunkt $t_0 = 0$ s gerade die Sendefenstergröße halbiert worden sein, also $w_s = x/2$ gelten. Zeichnen Sie das Diagramm im Zeitintervall $t = \{0, \ldots, 27\}$.



d)* Wieviel Zeit vergeht, bis nach einem Segmentverlust das Staukontrollfenster infolge eines weiteren Segmentverlusts wieder reduziert wird?

Nach einem Segmentverlust wird w_c auf $x/2$ reduziert und anschließend pro vollständig bestätigtem Fenster wieder um 1 MSS vergrößert. Da die Serialisierungszeit vernachlässigbar klein ist, können also zu einem Zeitpunkt t_0 insgesamt w_c Segmente gesendet werden, welche zum Zeitpunkt $t_0 + \text{RTT}$ bestätigt werden. Folglich erhalten wir für die Zeit bis zum erneuten Erreichen des Maximalwerts

$$T = \left(\frac{x}{2} + 1\right) \cdot \text{RTT} = 9 \cdot 200\,\text{ms} = 1{,}8\,\text{s}.$$

e)* Bestimmen Sie allgemein die durchschnittliche Verlustrate θ. Hinweis: Da das Verhalten von TCP in diesem idealisierten Modell periodisch ist, reicht es aus, lediglich eine Periode zu betrachten. Setzen Sie die Gesamtzahl übertragener Segmente in Relation zur Anzahl verlorener Segmente (Angabe als gekürzter Bruch ist ausreichend).

Zunächst bestimmen wir die Anzahl n an Segmenten, welche während eines „Sägezahns“ übertragen werden:

$$\begin{aligned} n &= \sum_{i=x/2}^{x} i = \sum_{i=1}^{x} i - \sum_{i=1}^{x/2-1} i = \frac{x \cdot (x+1)}{2} - \frac{\left(\frac{x}{2}-1\right) \cdot \frac{x}{2}}{2} \\ &= \frac{x^2+x}{2} - \frac{x^2}{8} + \frac{x}{4} \\ &= \frac{3}{8}x^2 + \frac{3}{4}x \\ &\overset{x=16}{=} 108 \end{aligned}$$

Pro „Sägezahn“ geht genau ein Segment verloren. Wir erhalten also für die Verlustrate

$$\theta = \frac{1}{\frac{3}{8}x^2 + \frac{3}{4}x} = \frac{1}{108} \approx 9{,}30 \cdot 10^{-3}$$

f) Bestimmen Sie mit Hilfe der Ergebnisse aus den Teilaufgaben (c) und (e) die in der betrachteten TCP-Übertragungsphase durchschnittlich erzielbare Übertragungsrate in kB/s.
Hinweis: Verwenden Sie den exakten Wert (Bruch) aus Teilaufgabe e).

Für die Datenrate erhalten wir

$$\begin{aligned} r_{TCP} &= \frac{n \cdot \mathrm{MSS}}{T} \cdot (1-\theta) \\ &= \frac{108 \cdot 1460\,\mathrm{B}}{1{,}8\,\mathrm{s}} \cdot \frac{107}{108} \\ &= \frac{107 \cdot 1460\,\mathrm{B}}{1{,}8\,\mathrm{s}} \\ &= \frac{1562200}{18}\,\mathrm{B/s} \\ &\approx \frac{1562}{18}\,\mathrm{kB/s} \approx 86{,}79\,\mathrm{kB}. \end{aligned}$$

g)* Bis zu welcher Übertragungsrate könnten Sie mit UDP maximal über den Kanal senden, ohne einen Stau zu erzeugen? Berücksichtigen Sie, dass der UDP-Header 12 B kleiner als der TCP-Header ohne Optionen ist.

Offenbar lassen sich 15 MSS verlässlich übertragen. Zusätzlich trägt ein Segment bei UDP 12 B mehr Nutzdaten als bei TCP. Wir erhalten also

$$\begin{aligned} r_{UDP} &= \frac{15 \cdot (\mathrm{MSS} + 12\,\mathrm{B})}{\mathrm{RTT}} \\ &= \frac{15 \cdot (1460\,\mathrm{B} + 12\,\mathrm{B})}{0{,}2\,\mathrm{s}} \\ &= \frac{15 \cdot 1472\,\mathrm{B}}{0{,}2\,\mathrm{s}} \\ &\approx 110{,}40\,\mathrm{kB/s}. \end{aligned}$$

Aufgabe 2 TCP und Long Fat Networks

In dieser Aufgabe betrachten wir sog. *Long Fat Networks*. Darunter versteht man Verbindungen, welche zwar eine hohe Übertragungsrate aber insbesondere auch eine hohe Verzögerung aufweisen. Beispiele dafür sind u. a. Satellitenverbindungen in Folge der hohen Ausbreitungsverzögerungen. Wir wollen insbesondere die Auswirkungen auf die TCP-Staukontrolle untersuchen.

a)* Bei TCP wird das Sendefenster in Abhängigkeit des Empfangsfensters sowie des Staukontrollfensters gewählt. Wie lautet der genaue Zusammenhang?

$$w_s = \min(w_r, w_c)$$

Zwei Nutzer seien nun über einen geostationären Satelliten an das Internet mit hoher Übertragungsrate angebunden. Die RTT zwischen beiden Nutzern betrage 800 ms, die Übertragungsrate sei $r = 24\,\text{Mbit/s}$.

b)* Wie groß muss das Sendefenster (gemessen in Byte) gewählt werden, damit kontinuierlich gesendet werden kann?

Das erste ACK kann frühestens nach einer RTT eintreffen, sofern man die Serialisierungszeiten vernachlässigt. Es ergibt sich also für das Sendefenster

$$w_s \geq \text{RTT} \cdot r = 800 \cdot 10^{-3}\,\text{s} \cdot 24 \cdot 10^6\,\frac{\text{bit}}{\text{s}} = 100 \cdot 24 \cdot 10^3\,\text{B/s} = 2{,}4\,\text{MB}.$$

c)* Warum ist die Situation in Teilaufgabe b) ein Problem für die TCP-Flusskontrolle?

Da das Sendefenster als Minimum aus Empfangs- und Staukontrollfenster gewählt wird und der Empfänger dem Sender sein Empfangsfenster über das Receive-Window-Feld mitteilt, welches auf 16 bit beschränkt ist, ist auch das Sendefenster auf einen Maximalwert von $(2^{16} - 1)\,\text{B} = 65535\,\text{B}$ beschränkt. Wir bräuchten laut Teilaufgabe b) allerdings ein Sendefenster der Größe $2,4 \cdot 10^6\,\text{B}$.

d)* Lesen Sie Sektion 2 von RFC 1323 (`http://www.ietf.org/rfc/rfc1323.txt`, siehe Anhang). Beschreiben Sie die Lösung für das Problem aus Teilaufgabe c).

Wir benötigen die Option *TCP-Window-Scaling*, welche dafür sorgt, dass das Receive-Window mit 2^x skaliert wird. Das Feld „shift.cnt“ der TCP-Window-Scaling-Option gibt den Exponenten x an.

e) Bestimmen Sie den minimalen Wert für das `shift.cnt`-Feld der TCP-Window-Scaling-Option.

Es muss gelten:

$$\begin{aligned}(2^{16} - 1) \cdot 2^x &\geq 2,4 \cdot 10^6 \\ x &\geq \text{ld}\left(\frac{2.4 \cdot 10^6}{2^{16} - 1}\right) \\ &= \text{ld}\,(36{,}62) \\ &\approx 5{,}19 \\ &\Rightarrow x = 6\end{aligned}$$

Erklärung:

Wir suchen den kleinsten Exponenten x, so dass das maximale Receive-Window größer als der in Teilaufgabe b) berechnete Wert von $2.4 \cdot 10^6\,\text{B}$ ist. Das Receive-Window ist 16 bit breit, kann also maximal den Wert `0xffff` $= 2^{16} - 1$ annehmen. Dieser Wert muss also nun mit 2^x skaliert werden.

Ein kurzer Blick auf die Größe des Sequenznummernraums von TCP ($|\mathcal{S}| = 2^{32}$, da SEQ- und ACK-Nummern 32 bit lange Felder sind) zeigt, dass wir kein Problem wie in der Aufgabe “Schiebefensterprotokolle” bekommen.

f) Geben Sie den Header des ersten TCP-SYN-Pakets an, welches die Verbindung aufbaut. Verwenden Sie dazu die konkreten Zahlenwerte aus der Angabe. Ein TCP-Header ist zur Erinnerung nochmals in Abbildung 1 dargestellt. Dort finden sich auch zwei Vordrucke zur Lösung.

Hinweis: Es ist nicht notwendig, den Header binär auszufüllen. Machen Sie aber bitte deutlich, ob es sich um hexadezimale, dezimale oder binäre Darstellung der Zahlen handelt.

Siehe Vordruck. Die $(6)_{10}$ des Data Offsets sagt aus, dass der Header eine Länge von $6 \cdot 32$ bit hat, die Nutzdaten also nach jener Anzahl von Bits beginnen.

Angenommen die Größe des Staukontrollfensters betrage derzeit die Hälfte des in Teilaufgabe b) berechneten Werts. Die MSS betrage 1200 B und die TCP-Verbindung befinde sich derzeit in der Congestion-Avoidance-Phase.

g) Wie lange dauert es, bis das Fenster die Leitung komplett ausnutzen kann?

Hinweis: Das Staukontrollfenster wird durch TCP-Window-Scaling nicht beeinflusst.

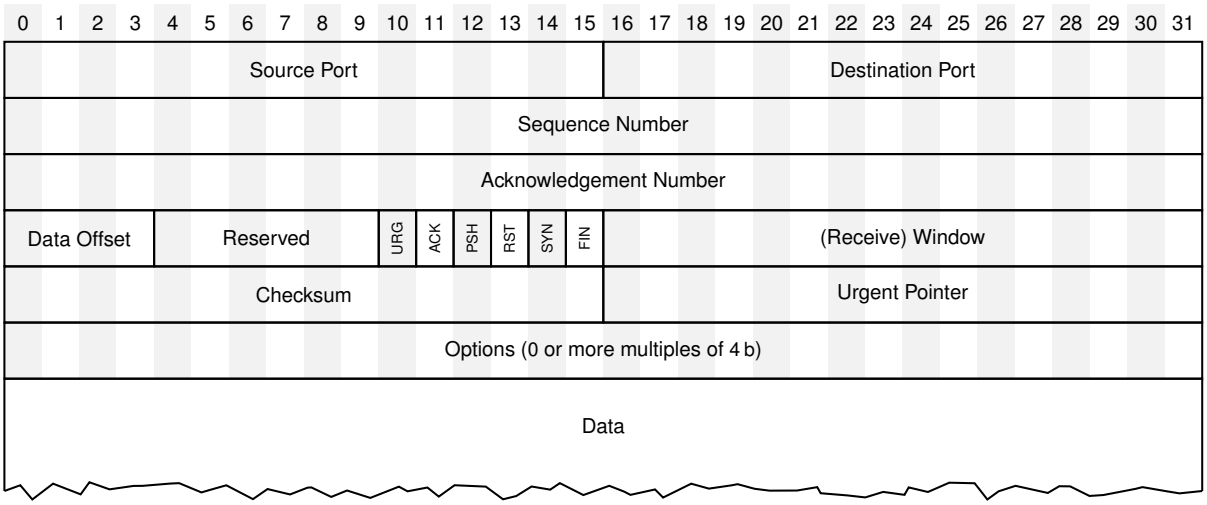
Das Fenster wird pro RTT um 1 MSS vergrößert. Folglich werden

$$\frac{1,2 \cdot 10^6\,\mathrm{B}}{1200\,\mathrm{B}} \cdot 0,8\,\mathrm{s} = \frac{96 \cdot 10^2}{12}\,\mathrm{s} = 800\,\mathrm{s}$$

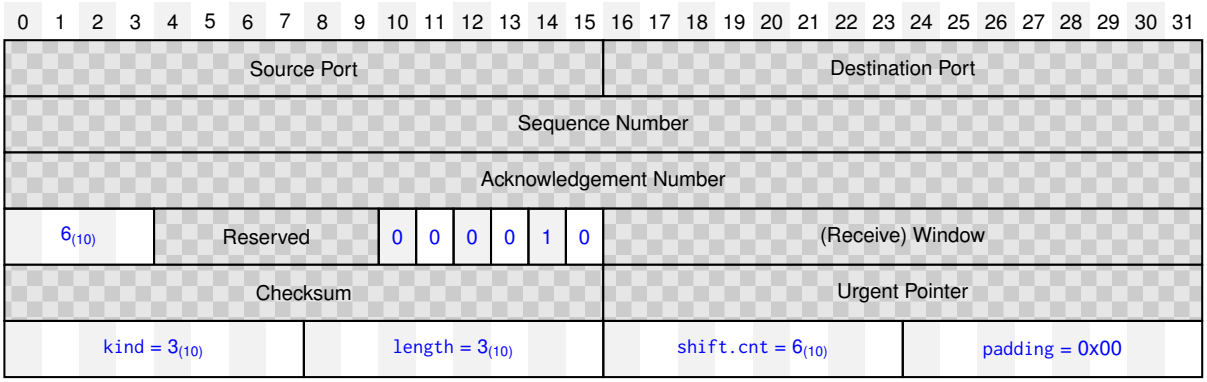
benötigt.

h) Ergibt sich aus dem Ergebnis von Teilaufgabe g) ein Problem?

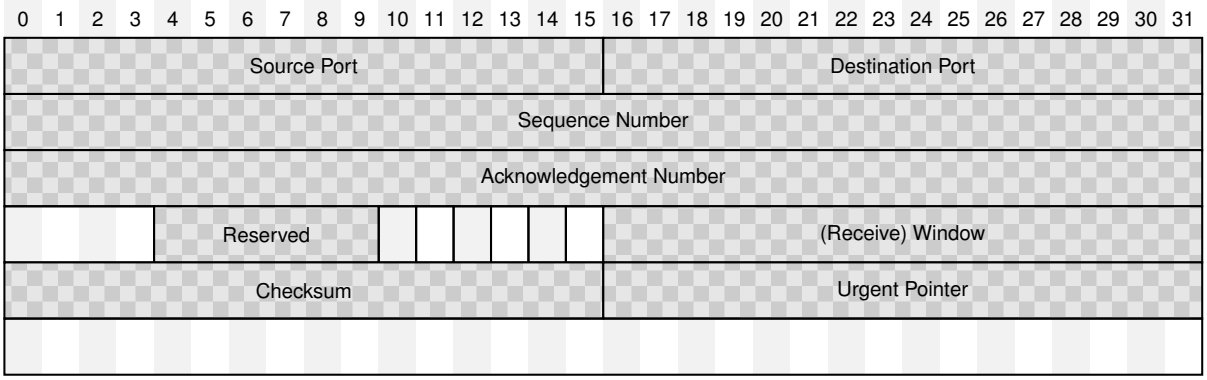
Ja. Es dauert mehr als 10 min bis TCP das Receive-Window wieder vollständig ausnutzt – viel zu lange.



(a) TCP-Header



(b) Vordruck



(c) Noch ein Vordruck, falls man sich vermalt hat

Abbildung 1: TCP-Header und Vordrucke zur Lösung von Aufgabe 2

2. TCP WINDOW SCALE OPTION

2.1 Introduction

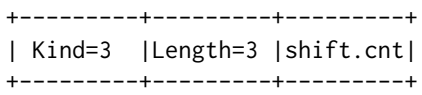
The window scale extension expands the definition of the TCP window to 32 bits and then uses a scale factor to carry this 32- bit value in the 16-bit Window field of the TCP header (SEG.WND in RFC-793). The scale factor is carried in a new TCP option, Window Scale. This option is sent only in a SYN segment (a segment with the SYN bit on), hence the window scale is fixed in each direction when a connection is opened. (Another design choice would be to specify the window scale in every TCP segment. It would be incorrect to send a window scale option only when the scale factor changed, since a TCP option in an acknowledgement segment will not be delivered reliably (unless the ACK happens to be piggy-backed on data in the other direction). Fixing the scale when the connection is opened has the advantage of lower overhead but the disadvantage that the scale factor cannot be changed during the connection.)

The maximum receive window, and therefore the scale factor, is determined by the maximum receive buffer space. In a typical modern implementation, this maximum buffer space is set by default but can be overridden by a user program before a TCP connection is opened. This determines the scale factor, and therefore no new user interface is needed for window scaling.

2.2 Window Scale Option

The three-byte Window Scale option may be sent in a SYN segment by a TCP. It has two purposes: (1) indicate that the TCP is prepared to do both send and receive window scaling, and (2) communicate a scale factor to be applied to its receive window. Thus, a TCP that is prepared to scale windows should send the option, even if its own scale factor is 1. The scale factor is limited to a power of two and encoded logarithmically, so it may be implemented by binary shift operations.

TCP Window Scale Option (WSopt):
Kind: 3 Length: 3 bytes



This option is an offer, not a promise; both sides must send Window Scale options in their SYN segments to enable window scaling in either direction. If window scaling is enabled, then the TCP that sent this option will right-shift its true receive-window values by 'shift.cnt' bits for transmission in SEG.WND. The value 'shift.cnt' may be zero (offering to scale, while applying a scale factor of 1 to the receive window).

This option may be sent in an initial <SYN> segment (i.e., a segment with the SYN bit on and the ACK bit off). It may also be sent in a <SYN,ACK> segment, but only if a Window Scale op- tion was received in the initial <SYN> segment. A Window Scale option in a segment without a SYN bit should be ignored.

The Window field in a SYN (i.e., a <SYN> or <SYN,ACK>) segment itself is never scaled.

Aufgabe 3 Code Demos – Beispiele für mögliche Klausurfragen (Hausaufgabe)

Am 18. und 19. Juni fanden in der Vorlesung Code-Demos zu UDP und TCP statt. Dabei wurden auch einige mögliche Fragestellungen für die Klausur genannt, die aus diesen Vorlesungseinheiten stammen. Da jedoch offensichtlich grade am 19. Juni viele Teilnehmerinnen / Teilnehmer verhindert waren, möchten wir an dieser Stelle beispielhafte Aufgabenstellungen für die Klausur vorstellen.

Ausführliche Informationen zu den behandelten Syscalls finden sie in den Manpages unter Linux bzw. OS X, welche selbstverständlich auch ohne den Umweg einer lokalen Linux-Installation oder der Nutzung der von uns bereitgestellten VMs direkt im Internet verfügbar sind (‘`man <syscall>`‘ auf Google).

Bitte beachten Sie, dass es zu diesen Fragen **keinen** Lösungsvorschlag geben wird.

a)* Erläutern Sie in 1–2 Sätzen oder Stichpunkten die Funktion des Syscalls `socket()`.

b)* Erläutern Sie in 2–3 Sätzen oder Stichpunkten die Funktion des Syscalls `bind()`.

c)* Erläutern Sie in 1–2 Sätzen oder Stichpunkten die Funktion des Syscalls `connect()`.

d)* Erläutern Sie in 2–3 Sätzen oder Stichpunkten die Funktion des Syscalls `accept()`. Gehen Sie dabei insbesondere auf Argument und Rückgabewert ein.

e)* Erläutern Sie in 2–3 Sätzen oder Stichpunkten die Funktion des Syscalls `select()`.

f)* Erläutern Sie in 2–3 Sätzen oder Stichpunkten die Funktion des Syscalls `listen()`.

g)* Erläutern Sie in 3–4 Sätzen oder Stichpunkten die Unterschiede zwischen den Syscalls `recv()` und `recvfrom()`.

h)* Erläutern Sie in 3–4 Sätzen oder Stichpunkten die Unterschiede zwischen den Syscalls `send()` und `sendto()`.

i)* Nennen Sie die notwendigen Syscalls zum Öffnen eines passiven Sockets in der richtigen Reihenfolge (auch bekannt als Listening Socket).

Machen Sie sich mit den in der Vorlesung programmierten bzw. vorbestellen Programmen `udpchat` und `tcpchat` vertraut. Diese stehen auf `https://grnvs.net` zum Download zur Verfügung.

j)* Weswegen **müssen** beim `udpchat` (ohne Server) Absender-Port, Ziel-IP und Ziel-Port angegeben werden, während beim TCP-Client die Angabe von Ziel-IP und Ziel-Port ausreichen?

k)* Weswegen konnte der UDP-Relay-Chat (`udp_server`) einem Client erst dann Antworten, **nachdem** dieser eine Textnachricht an den Server geschickt hatte?

l)* Weswegen kann der `udp_server` nicht ohne Weiteres erkennen, ob ein Client offline ist?

m)* Beschreiben Sie mindestens zwei Vorteile, die der TCP-Chat gegenüber dem UDP-Chat bietet.