Layer 3: Network Layer (hier: Internet Protocol Version 4)

Aufgabe:

Weltweite Vernetzung zum Internet. Das geschieht durch die Koppelung von einzelnen LANs (Local Area Networks) zu WANs (Wide Area Networks) und diese dann zum Internet. Die Kommunikation innerhalb eines LAN erfolgt per Layer2, d.h. durch die Verwendung von MAC-Adressen. Die IP-Adresse dient hier nur zur Feststellung, ob Sender und Empfänger im selben LAN liegen. Erst wenn Sender und Empfänger in verschiedenen LANs liegen, wird die IP-Adresse zur Weiterleitung in das Empfänger-LAN verwendet.

Neue Adressen: IPv4-Adressen

verbundene Geräte im gleichen LAN haben "ähnliche" Adressen. An den Adressen lässt sich bereits erkennen, ob die Geräte zum selben LAN gehören. Die IPv4-Adresse ist eine 32 Bit Binärzahl (in C/C++ üblicherweise: `unsigned int` bzw. besser `uint32_t`). Diese wird in 4 Bytes (1 Byte besteht aus 8 Bits) aufgeteilt und einzeln dezimal ausgegeben mit einem Punkt als Trennzeichen.

138.232.82.1	Server mat1.uibk.ac.at im WAN der Uni-Innsbruck
138.232.95.153	Vorführ-PC HSB5
138.232.95.156	Studenten-PC HSB5

Für die Ermittlung des LAN brauchen wir leider die binäre Darstellung der IPv4-Adressen. Diese kann man relativ leicht mit dem Taschenrechner von Windows 10 bekommen (Modus Programmierer!):

138.232.82.1 = (binär) 10001010.11101000.01010010.00000001
138.232.95.153 = (binär) 10001010.11101000.01011111.10011001
138.232.95.156 = (binär) 10001010.11101000.01011111.10011100

Was sind ähnliche Adressen: Idee: 138.232.82.1 ist ähnlich zu 138.232.95.153, da sie mit denselben Ziffern beginnen. Allerdings wird die Ähnlichkeit anhand der binären Ziffern und nicht der dezimalen Ziffern festgestellt! Jedes LAN besitzt eine vom Administrator festgelegte **Startadresse** und eine **Subnetzmaske**. Die Subnetzmaske legt fest, auf wie viele Binärziffern zwei IPv4-Adressen von links beginnend übereinstimmen müssen, um im selben Subnet = LAN zu liegen. Sie wird entweder als IP-Adresse oder neuerdings in der Form /n geschrieben: Sie beginnt mit n Binärziffern 1 gefolgt von Binärziffern 0, z.B.

/16 = 255.255.0.0 = (binär) **11111111. 11111111**. 0000000.0000000
/24 = 255.255.255.0 = (binär) **11111111. 11111111**. **11111111**.0000000
/23 = 255.255.254.0 = (binär) **11111111. 11111111**. **1111111**0.0000000
/26 = 255.255.255.126 = (binär) **11111111. 11111111**. **11111111.11**00000

Ob 2 IP-Adressen im gleichen LAN liegen, hängt ganz wesentlich von der Subnetzmaske ab. Verwenden wir etwa die „Default-Subnetzmaske“ **/16**, so liegen alle 3 IP-Adressen im selben LAN. Ich habe die zu vergleichenden Bits fett geschrieben:

138.232.82.1 = (binär) **10001010.11101000**.01010010.00000001
138.232.95.153 = (binär) **10001010.11101000**.01011111.10011001
138.232.95.156 = (binär) **10001010.11101000**.01011111.10011100

In dieser Konfiguration können alle 3 Geräte auf Layer-2 kommunizieren, d.h. es ist kein Layer-3 Gerät (Router) erforderlich.

Nimmt man die an der Uni Ibk effektiv verwendete Subnetzmaske **/23**, so liegen die 2 HSB5-PCs im selben LAN, der Server mat1 jedoch in einem anderen LAN.

138.232.82.1 = (binär) **10001010.11101000.0101001**0.00000001
138.232.95.153 = (binär) **10001010.11101000.0101111**1.10011001
138.232.95.156 = (binär) **10001010.11101000.0101111**1.10011100

In dieser Konfiguration werden die PCs untereinander mit Layer-2 kommunizieren. Eine Kommunikation mit dem Server mat1 ist mit Hilfe eines Layer-3 Gerätes (= Router, Gateway) möglich. Der Router muss Interfaces in beiden LANs besitzen und ist so aus beiden Richtungen erreichbar.

Mit der Subnetzmaske steuert der Administrator die Größe des LANs, je kleiner die Zahl ist, desto weniger Bits sind fixiert und desto mehr Adressen gehören zum LAN.

Die Klasseneinteilung der IPv4-Adressen

Als die IPv4-Adressen eingeführt wurden, konnte man die Subnetzmaske nicht frei wählen, sondern diese wurde durch das erste Byte der IPv4-Adresse bestimmt. Dieses legte die Klasse der Adresse und damit auch die zu verwendende Subnetzmaske fest:

A-Klasse: 0 bis 127, Subnetzmaske /8 : enthält 2^{24} Adressen (ca. 16 Mio.)
es gibt demnach genau diese 128 A-Klasse-Netzwerke
B-Klasse: 128 bis 191, Subnetzmaske /16: enthält 2^{16} Adressen (ca. 65000)
es gibt ca. 16000 solche Netzwerke (128.0.0.0/16 bis 191.255.0.0/16)
C-Klasse: 192 bis 223, Subnetzmaske /24: enthält 2^{8} Adressen (256)
es gibt ca. 2 Mio. Netzwerke (192.0.0.0/24 bis 223.255.255.0/24)
D-Klasse: 224 – 239: Multicast-Adressen, keine Subnetzmaske
E-Klasse: 240 – 255: reserviert, keine Subnetzmaske, nur 1 Adresse verwendet:
255.255.255.255 limited Broadcast-Adresse

Mittlerweile ist dieses Schema überholt! Es lassen sich für jedes LAN die Startadresse und die Subnetzmaske (fast frei) wählen. Es müssen jedoch bei der Startadresse alle Bits „nach“ der Subnetzmaske 0 sein. Man spricht daher vom „classless Internet“.

Aus einer IPV4-Adresse eines LANs und dessen Subnetzmaske ergibt sich der Adressbereich des LANs, z.B. HSB5. Zur Ermittlung der kleinsten Adresse werden die freien Bits auf 0 gesetzt. Zur Ermittlung der größten Adresse werden die freien Bits auf 1 gesetzt:

IPv4 138.232.95.153 , Subnetzmaske **/23** (255.255.254.0):

138.232.95.153	=	**10001010.11101000.0101111**1.10011001
kleinste IP binär	=	**10001010.11101000.0101111**0.00000000
größte IP binär	=	**10001010.11101000.0101111**1.11111111

kleinste IP dezimal:	138.232.94.0	Adresse des LAN bzw. Subnetzes
größte IP dezimal :	138.232.95.255	Broadcast-Adresse des LAN bzw. Subnetzes

Die kleinste und die größte Adresse im Adressbereich dürfen nicht für Geräte verwendet werden, sondern sind die LAN-Adresse und die LAN-Broadcast-Adresse.

Kommandos:

`ping`	Verbindungstest, Geschwindigkeitstest (Windows, Unix),
`tracert`	Routenverfolgung (Windows)
`traceroute`	Routenverfolgung (Unix)
`route print`	bekannte Netzwerke ausgeben (Windows)
`route`	bekannte Netzwerke ausgeben (UNIX)

```
[stix@mat1 ~]$ ping www.uibk.ac.at
```

```
PING www.uibk.ac.at (138.232.17.233) 56(84) bytes of data.
64 bytes from www1.uibk.ac.at (138.232.17.233): icmp_seq=1 ttl=62 time=0.294 ms
64 bytes from www1.uibk.ac.at (138.232.17.233): icmp_seq=2 ttl=62 time=0.279 ms
64 bytes from www1.uibk.ac.at (138.232.17.233): icmp_seq=3 ttl=62 time=0.180 ms
^C
--- www1.uibk.ac.at ping statistics ---
3 packets transmitted, 3 received, 0% packet loss, time 2000ms
rtt min/avg/max/mdev = 0.180/0.251/0.294/0.050 ms
```

Die Pingzeit misst die Zeit vom Absenden des Echo-Requests `Ping` von der eigenen Station bis zum Eintreffen des Echo-Replys `Pong` der Gegenstelle (= Round Trip Time rtt). Beim ersten Mal ergeben sich hier manchmal zu lange Werte, sodass man erst ab dem 2.

Durchgang sinnvolle Zeitwerte erhält. Durch Firewalls kann sowohl das ping-Kommando als auch die Routenverfolgung beeinträchtigt sein.

[stix@mat1 ~]$ `traceroute www.krone.at`

`traceroute` (oder `tracert` unter Windows) ermittelt alle Zwischenschritte (Hops), die ein IPv4-Paket vom eigenen Rechner zur Gegenstelle macht und gibt für diese auch ping-Zeiten aus. Sollte sich eine Zwischenstation nicht ermitteln lassen - was an einer aktiven Firewall liegen wird - werden nur Sternchen für diesen Hop ausgegeben. Die Zwischenstationen werden durch variierende Werte für die TTL (Time To Live) ermittelt.

[c80422@zid-gpl ~]$ `route -n`

Kernel IP Routentabelle

Ziel	Router	Genmask	Flags	Metric	Ref	Use	Iface
0.0.0.0	138.232.1.254	0.0.0.0	UG	100	0	0	eth0
138.232.1.0	0.0.0.0	255.255.255.0	U	100	0	0	eth0
192.168.122.0	0.0.0.0	255.255.255.0	U	0	0	0	virbr0

Obige Routingtabelle enthält die Default-Route an erster Stelle und 2 angeschlossene LANs. Die Routeradresse 0.0.0.0 in letzteren besagt nur, dass es für dieses LAN keinen Router gibt, d.h. dass es direkt an einen Netzwerk-Adapter der eigenen Maschine angeschlossen ist.

Routing = Routen-Auswahl:

Um ein IP-Paket an eine IPv4-Adresse (die nicht eine eigene sein soll!) zu senden, muss jedes Gerät entscheiden, wer der (nächste) Empfänger des Pakets sein soll. Dieser Empfänger muss in einem verbundenen LAN liegen (d.h. mindestens eine eigene Netzwerkschnittstelle muss im gleichen LAN wie dieser Empfänger liegen). Ist die finale Zieladresse nicht direkt erreichbar, muss statt dieser ein **Gateway** adressiert werden, welches das Paket dann weiterleitet.

Dazu besitzt jedes IP-Gerät eine Tabelle mit allen bekannten Netzwerken. Für jedes Netzwerk ist die Startadresse, die Subnetzmaske und das passende Gateway zu diesem Netz eingetragen. Diese Tabelle heißt Routing-Tabelle. Für die meisten Endgeräte besitzt diese Tabelle genau 3 echte Einträge:

1) das virtuelle `Loopback` Netzwerk 127.0.0.0/8 mit Gateway 127.0.0.1 (`localhost`). Dieses „Netzwerk“ ist auf jedem Gerät vorhanden und ist praktisch die Schnittstelle des Geräts zum Internet.

2) das (die) eigene LAN(s): Für jedes aktive Netzwerkinterface dessen Netzwerk, als Gateway wird die IPv4-Adresse dieses Interfaces eingetragen.
3) der Rest der Welt: Startadresse 0.0.0.0 und Netzmaske /0: Jede IP-Adresse liegt in diesem Bereich. Das Gateway in diesem Eintrag heißt Standard-Gateway (Default Gateway).

Somit ist die Routing-Entscheidung sehr einfach:

1) Ist die finale IPv4 im eigenen LAN, so ist sie auch der nächste Empfänger. Der Layer 2 wird beauftragt, das IP-Paket an diesen zuzustellen. Das Paket wird auch im Layer 2 an diesen finalen Empfänger (an dessen MAC-Adresse) gesendet.
2) Ist die finale IPv4-Adresse nicht im eigenen LAN, so ist der nächste Empfänger das Standardgateway. Der Layer 2 wird beauftragt, das IPv4-Paket an dieses Gateway zuzustellen. Hierbei ist die Layer 3 – Zieladresse der finale Empfänger, die Layer-2 Zieladresse ist die MAC-Adresse des Standardgateways.

Wie weiß mein PC die MAC-Adressen der anderen Teilnehmer?

Zu diesem Zweck wurde das Hilfsprotokoll ARP (Address Resolution Protocol) eingeführt. Es gehört nicht direkt zu den Internetprotokollen und dient nur dem Zweck, zu einer IPv4-Adresse die zugehörige MAC-Adresse herauszufinden. Dazu sendet der eigene PC einen Layer-2 Broadcast (Zieladresse ff:ff:ff:ff:ff:ff) an alle: Wer hat die IPv4-Adresse x.x.x.x (ARP-Request). Der richtige Inhaber Antwortet daraufhin mit einem ARP-Reply und übermittelt seine MAC-Adresse. Diese wird dann für den Paketversand verwendet und auch kurze Zeit im sogenannten ARP-Cache zwischengespeichert. Solange sie dort gespeichert ist, wird sie auch weiterverwendet:

`arp –a` Gibt Inhalt des ARP-Caches aus (Windows, UNIX)

Kritik: Das ARP-Protokoll ist sehr ineffizient, denn als Layer-2 Broadcast wird es das ganze LAN und jeden Netzknoten fluten. Darüber hinaus ist es eine Sicherheitslücke, denn statt des rechtmäßigen MAC-Besitzers könnte sich auch ein Hacker melden und somit die Pakete an sich umleiten. Es werden von vielen Geräten sogar ARP-Replys akzeptiert, die gar nicht angefordert wurden. So hat der Hacker noch leichteres Spiel, da er nicht mehr schneller als der echte Besitzer sein muss.

Woher bekommt mein PC eine IPv4-Adresse?

IPv4-Adressen müssen entweder statisch konfiguriert werden (man darf keine Adresse mehr als 1x vergeben) oder können von einem DHCP-Server aus dem Netzwerk erfragt werden (siehe DHCP in Abschnitt Layer 5-7): PlugnPlay-Netzwerke

Heimrouter sind meistens schon so vorkonfiguriert, dass sie die Aufgabe des DHCP-Servers übernehmen und allen Geräten eindeutige IP-Adressen zuweisen.

Reservierte Netzwerke:

Nicht alle Netzwerkadressen sind für jeden Zweck verwendbar. Die IANA (Internet Assigned Numbers Authority) reservierte manche Bereiche der A, B, C-Klasse Netzwerke für spezielle Verwendungen, sodass diese nicht an Netzknoten im Internet vergeben werden dürfen. Die folgende Liste enthält die wichtigsten davon:

0.0.0.0/8	
10.0.0.0/8	1 A-Klasse Netz für Private Adressen
127.0.0.0/8	Loopback (die eigene Maschine ist 127.0.0.1 = localhost)
172.16.0.0/12	16 B-Klasse Netze für Private Adressen 172.16.0.0 – 172.31.0.0 /16
169.254.0.0/1	Link Local Adressen - APIPA
192.168.0.0/16	256 C-Klasse Netze für Private Adr. 192.168.0.0 – 192.168.255.0 /24

Private Adressen dürfen im Internet nicht verwendet werden sondern sind für nicht direkt mit dem Internet verbundene Netze reserviert (z.B. Heimnetze). An der Übergangsstelle zum Internet muss hier eine „Übersetzung“ der privaten Adressen in offizielle Internet-Adressen erfolgen : **NAT (Network Address Translation)**

Das IPv4-Paket:

Das IP-Paket reist zur Gänze als Nutzlast in einem Layer-2 Frame mit dem Datentyp IPv4. Der Aufbau ist hochkomplex, da man ursprünglich an viele Möglichkeiten gedacht hatte, die letztendlich nie zum Einsatz kamen. Die wichtigsten Teile sind:

Absender IPv4-Adresse
Empfänger IPv4-Adresse
TTL - Time To Live
Typ der Nutzlast
Nutzlast

Eine kurze Erklärung zur TTL: Jeder „Zwischenempfänger“ (= Router) zieht 1 von dieser Zahl ab. Er schickt das Paket nur weiter, wenn die TTL dadurch nicht 0 wird. Ist die TTL 0, sollte der Absender (sofern es eine Unicast-Adresse ist) über diesen Umstand benachrichtigt werden: (TTL-Exceeded in Transit). Durch eine geeignete Wahl der TTL kann man die

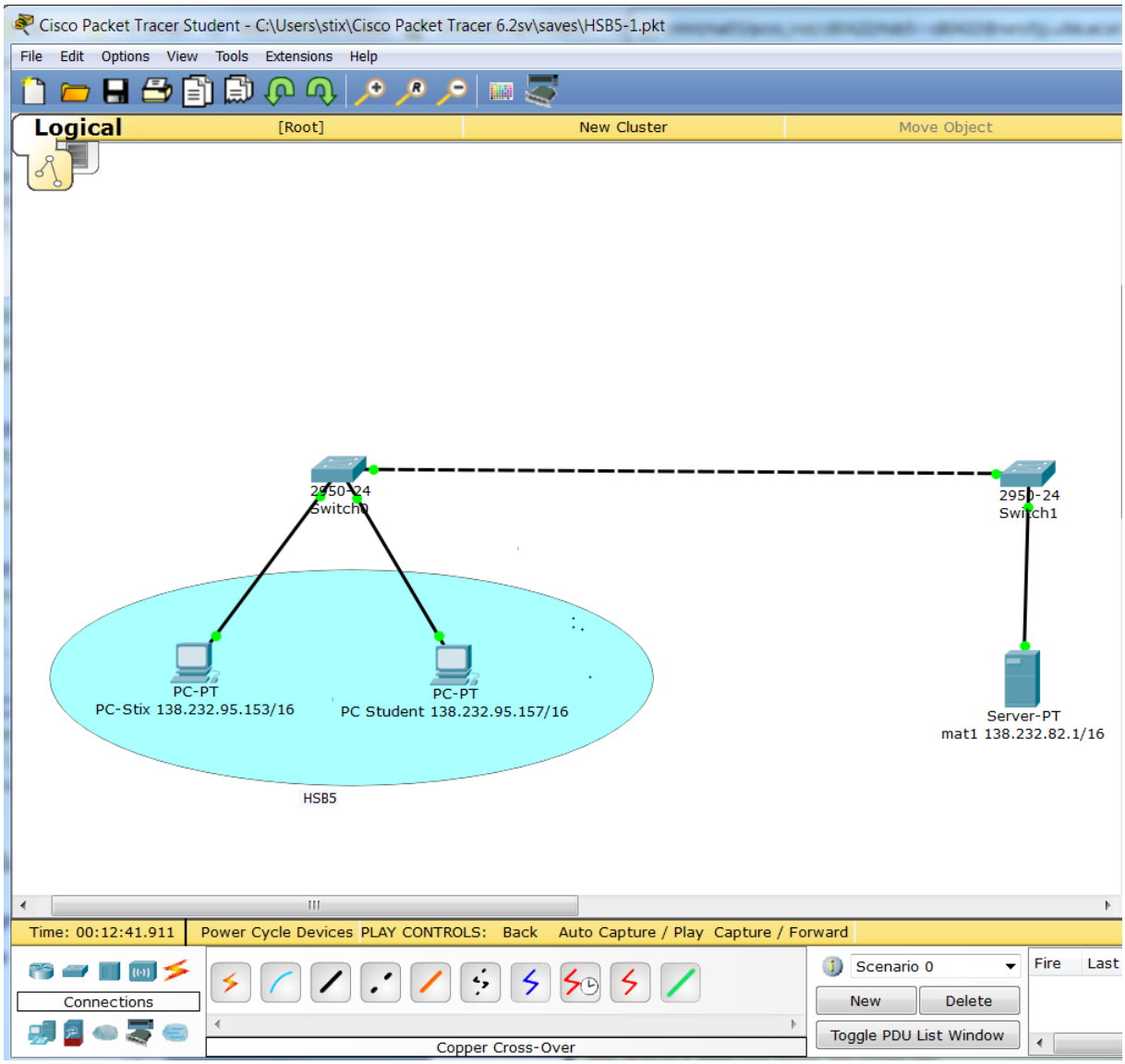
Reichweite eines Paketes beeinflussen: TTL == 1: Das Paket verbleibt im lokalen LAN. Das Traceroute-Kommando benutzt die TTL zum Herausfinden der Zwischenstationen. Pakete werden mit TTL 1, 2, 3 … an die Endstelle gesendet. Falls es dort nicht ankommt, kann man anhand der „TTL Exceeded“ Meldungen die Zwischenstellen identifizieren. Die wichtigste Aufgabe der TTL ist es jedoch, Schleifen im Internet, in denen Pakete endlos kreisen, zu entschärfen, weil jedes Paket garantiert nach maximal 255 Hops (das ist der Maximalwert) vernichtet wird.

Geräte, die Layer 1-3 können: Router

Router besitzen fast immer mehrere Netzwerkschnittstellen in verschiedenen LANs, sodass sie die Pakete zwischen diesen Netzwerken vermitteln können (packet forwarding). Der einzige Unterschied zu den Endgeräten liegt darin, dass ihre Routing-Tabelle oft nicht statisch (d.h. unveränderlich) ist, sondern sich durch Kommunikation mit Nachbarroutern ändert. Die Router lernen also von ihren Genossen und werden unterrichtet über unterbrochene Verbindungen (Bagger kappen Kabel) und wiederhergestellte oder neue Verbindungen. Ihre Hauptaufgabe ist es, für jedes Paket die „beste“ Verbindung zum Ziel zu berechnen und das Paket an die nächste Station dieser Route weiterzuleiten.

Allerdings ist nicht immer klar, was die „beste“ Verbindung ist: die schnellste, die kostengünstigste, die zuverlässigste? Gute Router behandeln nicht alle Pakete gleich sondern können durchaus (durch tiefere Inspektion der Pakete oder des IPv4-Headers) unterschiedliche Optimierungsziele ausmachen (QoS: Quality of Service): Echtzeit-Daten brauchen eine möglichst schnelle Weiterleitung, Massendaten nehmen die Billigschiene.

Praktisches Beispiel: HSB5-1 (vereinfachte Konfiguration des HSB 5)

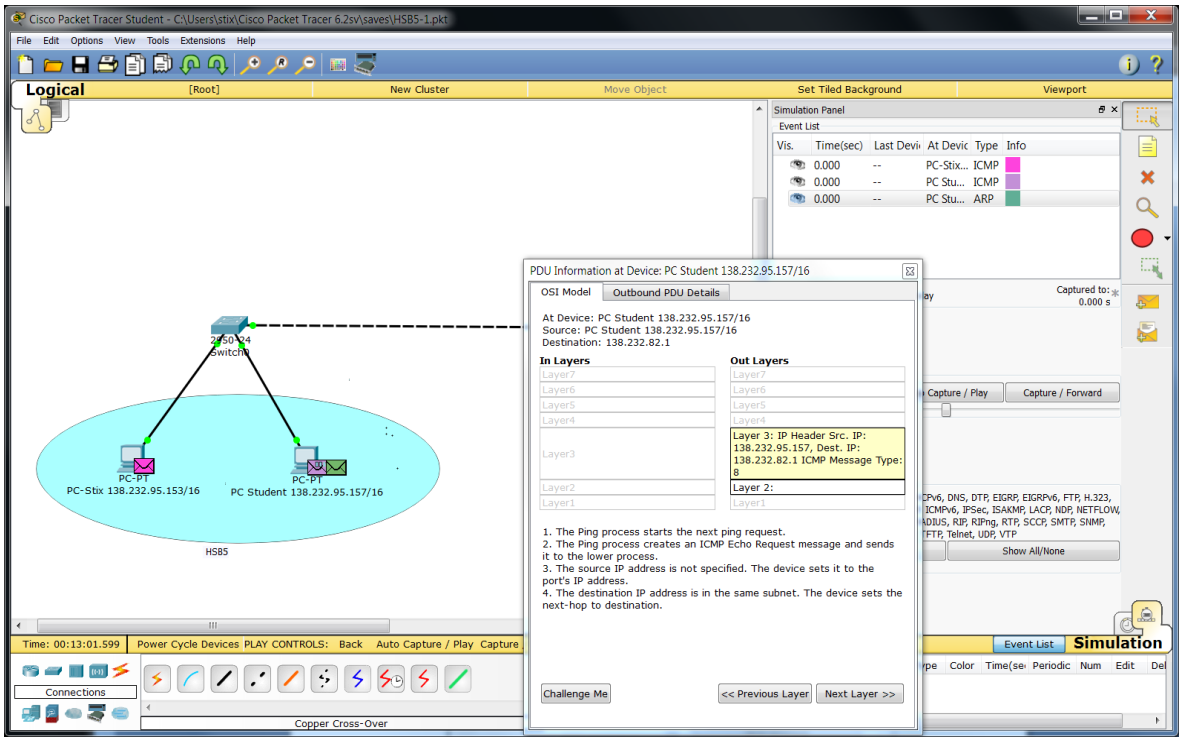


An diesem Beispiel kann man unterschiedliche Twisted-Pair Kabelarten austesten: Straight-Through sind durchgezogen, Crossover sind strichliert. Eine falsche Wahl verhindert, dass die Anschlüsse hochfahren (LEDs bleiben rot). Anschließend konfigurieren wir die 2 PCs auf die 2 angegebenen IPv4-Adressen, ebenso den Server mat1 und lassen den Rest auf Default stehen. Alle 3 Geräte sollten sich erreichen können (ping-Kommando). Startet man auf dem PC einen Web-Browser, sollte dieser Webseiten des Servers abrufen können.

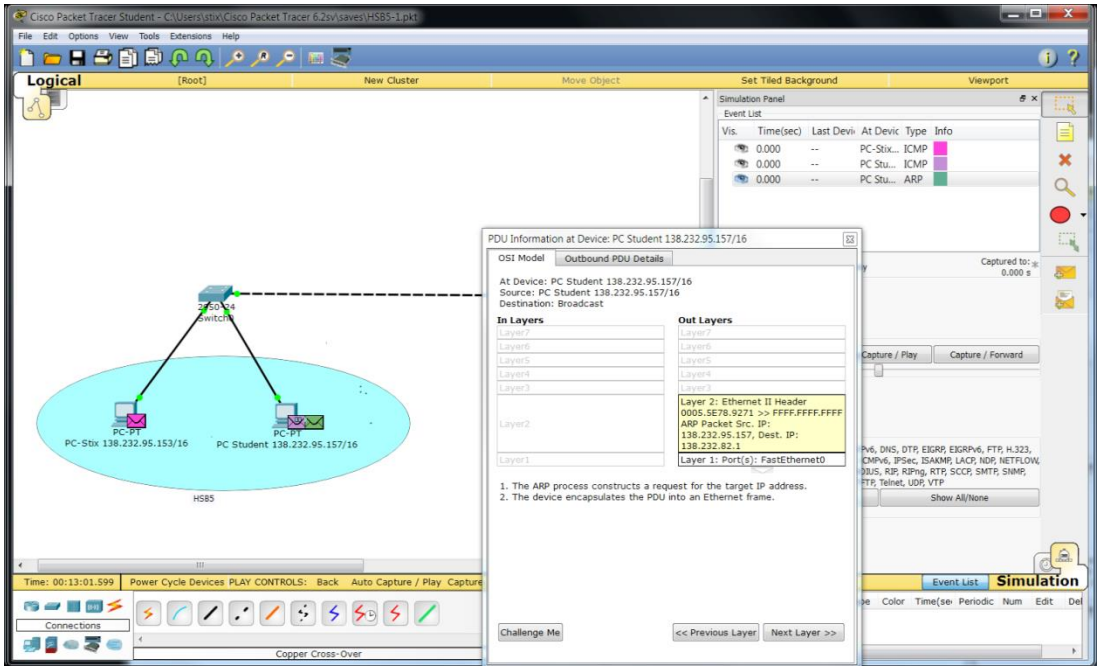
Im Simulationsmodus ist eine genaue Verfolgung aller Pakete möglich. Vorher muss man die ganze Konfiguration neu starten, damit alles zurückgesetzt wird:

PC0: `ping 138.232.82.1` (den Server):

1) Es wird ein Layer-3 ping-Paket (vom Typ ICMP) erzeugt, das aber zu diesem Zeitpunkt unvollständig ist, da die MAC-Adresse des Ziels (Server) nicht bekannt ist! Man sieht, dass die entsprechenden Felder im Layer 2 noch leer sind.



2) Um die MAC-Adresse des Servers zu bestimmen, wird das ARP-Protokoll (Address Resolution Protocol) verwendet (Address schreibt man im Englischen tatsächlich mit Doppel-d!!). Dies ist ein Layer-2 Protokoll, das an die L2-Broadcastadresse `ff:ff:ff:ff:ff:ff` sendet (an alle) mit der Bitte, dass sich der richtige Empfänger (der die richtige IP-Adresse besitzt) melden möge (ARP-Request). Der Host mit der gesuchten IP-Adresse soll sich beim Anfrager melden (ARP-Reply) und diesem seine MAC-Adresse mitteilen.



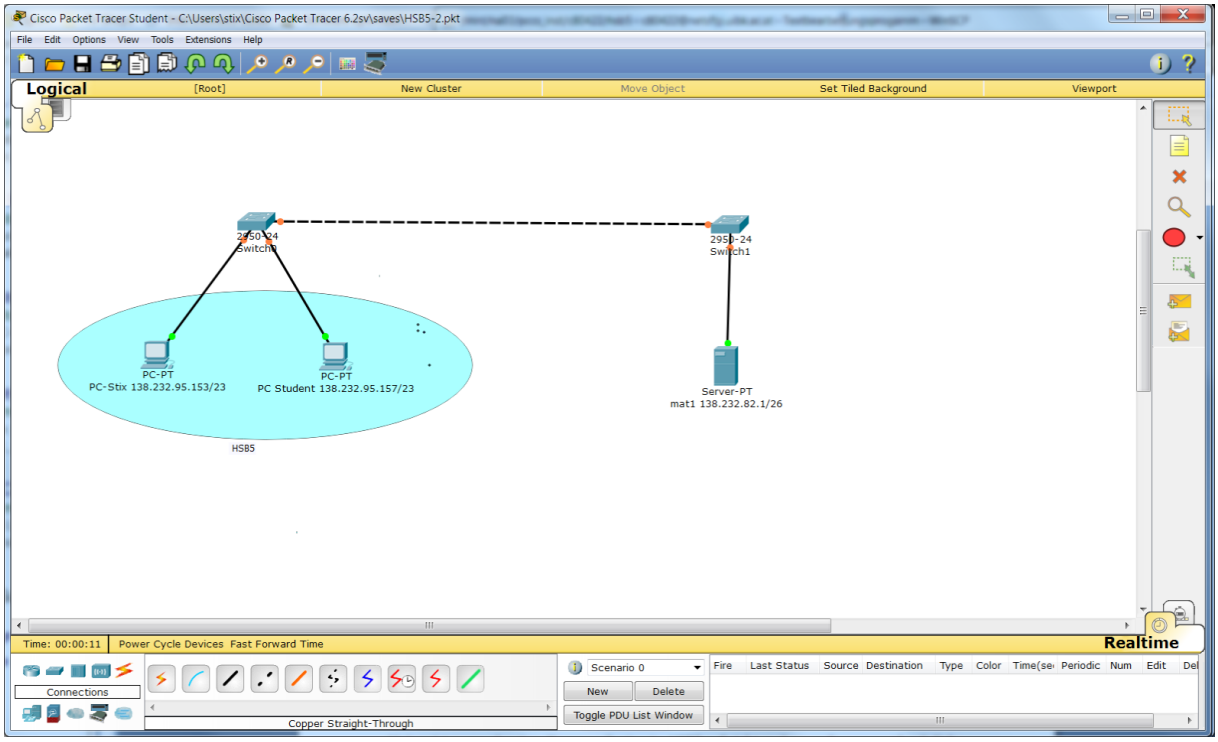
3) Der ARP-Request geht an den Switch und wird auf allen Leitungen (da Broadcast) weitergesendet. Der angesprochene Server sendet daraufhin seine MAC-Adresse in einem ARP-Antwortpaket an den anfragenden PC0. Dieses Paket ist ein Unicast

(siehe die Destination-MAC) und wird vom Switch nur an die richtige Leitung weitergeleitet.

4) Nach dem Eintreffen der MAC-Adresse wird das ICMP-ping-Paket (ICMP Message Type 8) fertig aufgebaut und abgesendet. Es erreicht auf direktem Weg den Server, der ein Antwortpaket (pong: Type 0) zurücksendet. Der PC gibt am Bildschirm eine Empfangsbestätigung aus mit zusätzlichen Timing-Informationen (Ping-Zeit). Ein niedriger Wert bedeutet eine gute, latenzarme Verbindung. Das ganze wird unter Windows noch 3 Mal wiederholt, UNIX versendet bis zum Programmabbruch laufend ping-Pakete.

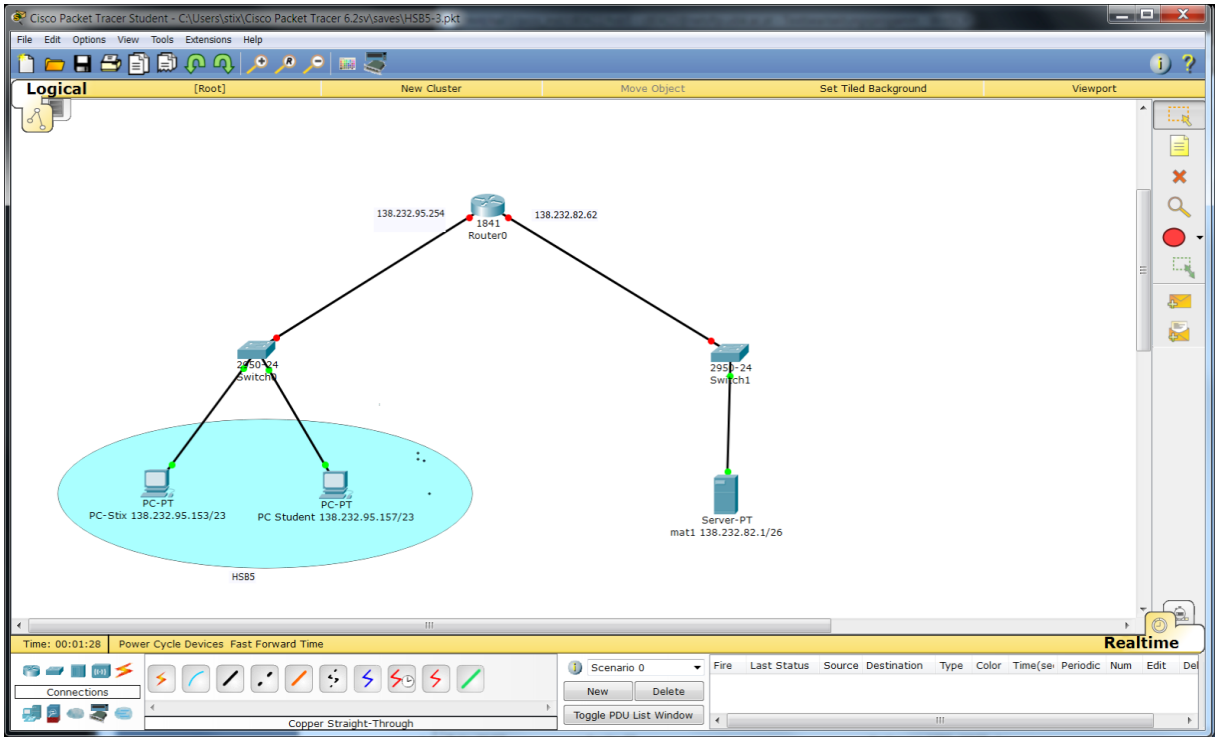
Beispiel: HSB5-2 (genauere Konfiguration des HSB 5)

Ändert man die Subnetzmaske (Subnet Mask) bei allen 3 Knoten auf den echten Wert der Uni, so ist nur mehr eine Kommunikation innerhalb des HSB5 aber nicht mehr zum Server `mat1` möglich.



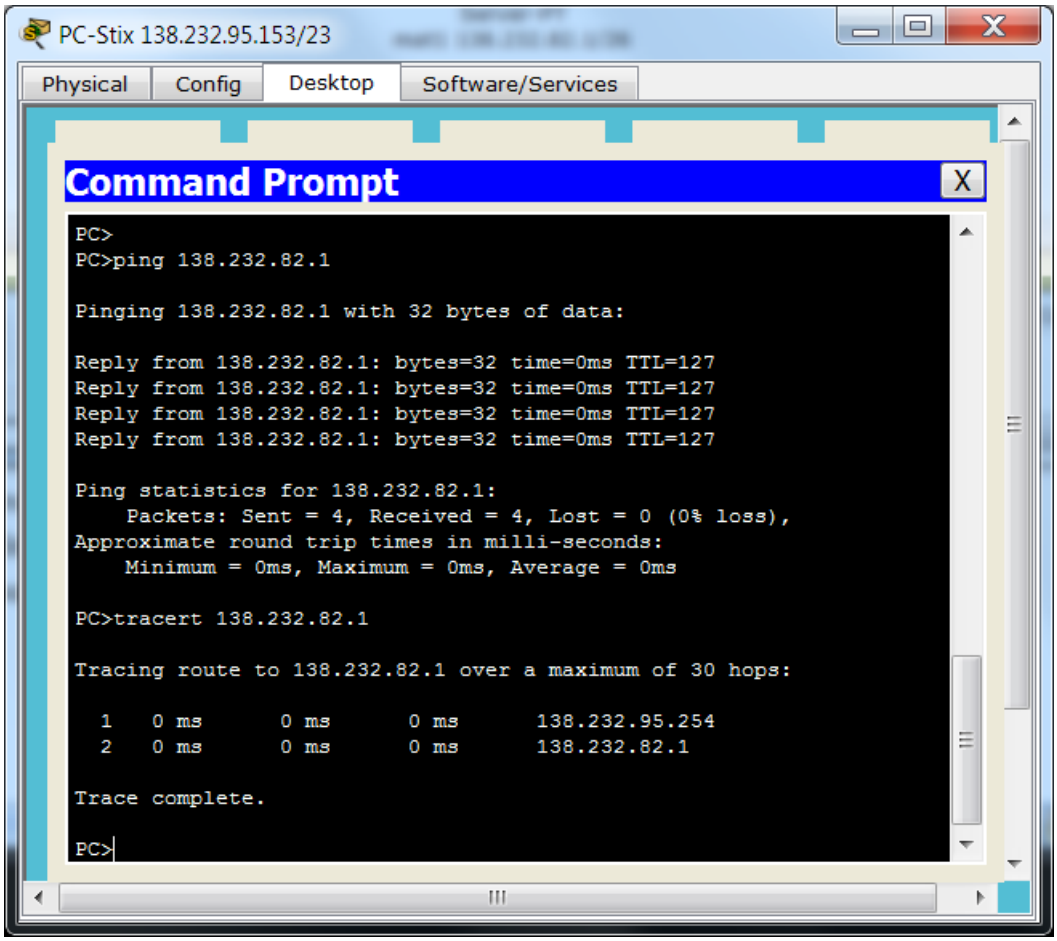
Beispiel: HSB5-3 (noch viel genauere Konfiguration des HSB 5)

Da in Beispiel 2 keine Kommunikation zwischen HSB5 und `mat1` möglich war (diese liegen in verschiedenen Subnetzen), koppeln wir hier die Subnetze mit einem Router. Statt der Switch-Verbindung schleifen wir den Router ein:



Der Router besitzt 2 Netzwerk-Interfaces, mit denen er an die 2 Subnetze (HSB5-Subnetz und Mathematik-Subnetz) angeschlossen ist und diese verbindet. Die IP-Adressen der Interfaces sind jene, die auf den (echten) PCs und dem Server als Standardgateway eingetragen sind. Auch in unserer Konfiguration müssen wir diese Gateways eintragen: im HSB5 `138.232.95.254`, auf der `mat1 138.232.82.62`

Natürlich müssen wir auch am Router diese 2 Adressen einstellen (mit den dazugehörigen Subnetzmasken), wobei wir diese Adressen auch den richtigen Interfaces zuweisen müssen. Ab jetzt lassen sich wieder alle Geräte von allen erreichen. Der folgende Screenshot zeigt die erfolgreiche Kommunikation PC-Stix <-> mat1:



Insbesondere erkennt man am Traceroute-Kommando, dass die Kommunikation in der Tat über den Router erfolgt.