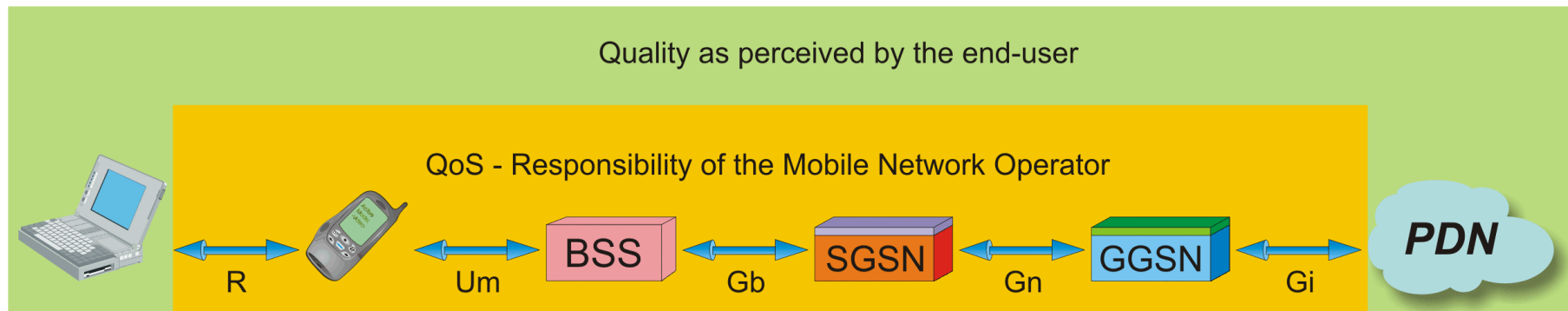


What is Quality of Service in GPRS Networks

- The QoS-Scope of the GPRS Network Operator



What is Quality of Service in GPRS Networks

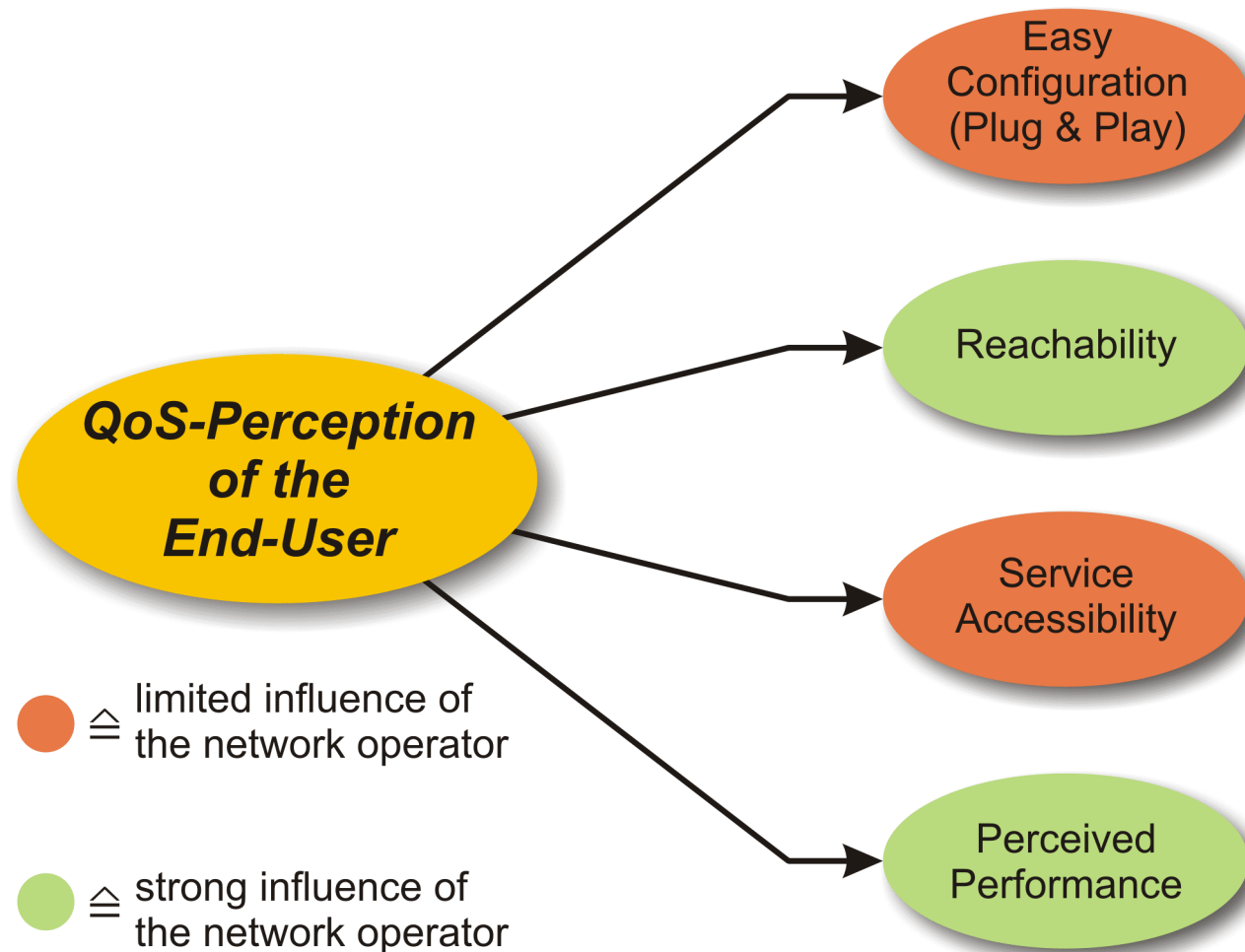
The QoS-Scope of the GPRS Network Operator

GPRS network operators are obviously eager to provide the best service possible to their customers. However, the figure already emphasizes the problem: While the end-user has an end-to-end view of performance and quality, the GPRS network operator can only be responsible for his part of the data transmission chain.

For example, delays which are caused by congestion on the internet or badly configured terminal equipment will limit the performance essentially and the customer will definitely blame it on the GPRS network.

It is particular this perception of the end-user which forces the GPRS network operator to provide an optimum performance and to continuously measure this performance by suitable measures.

The QoS-Perception of the End-User



The QoS-Perception of the End-User

Easy Configuration

Nowadays, one of the most important hurdles to use GPRS is the complicated setup of GPRS in the terminal equipment, most likely a laptop and in the mobile station itself. Even engineers have to struggle with the task of configuring GPRS. Apparently, nobody really considered AOL's advertisement slogan properly: "Oh, I'm already in ..." (the internet). In the long term, this needs to change.

Reachability

The most important feature of GPRS is the "Always On" feature. At least in theory, GPRS allows the customer to stay on the internet continuously. However, reality is different. Another issue is the lacking possibility to push IP-content to mobile subscribers. This would require the mobile terminating way of PDP-context activation.

Likewise, mobile subscribers need to stay reachable for their circuit-switched partners (voice calls) even while surfing the web. This can only be achieved through the introduction of NOM I.

Service Accessibility

Today, GPRS is primarily the domain of business users who travel abroad. Unfortunately, GPRS does still not stable operate while roaming outside the Home-PLMN. Adjusting GPRS-settings while roaming cannot be the final solution.

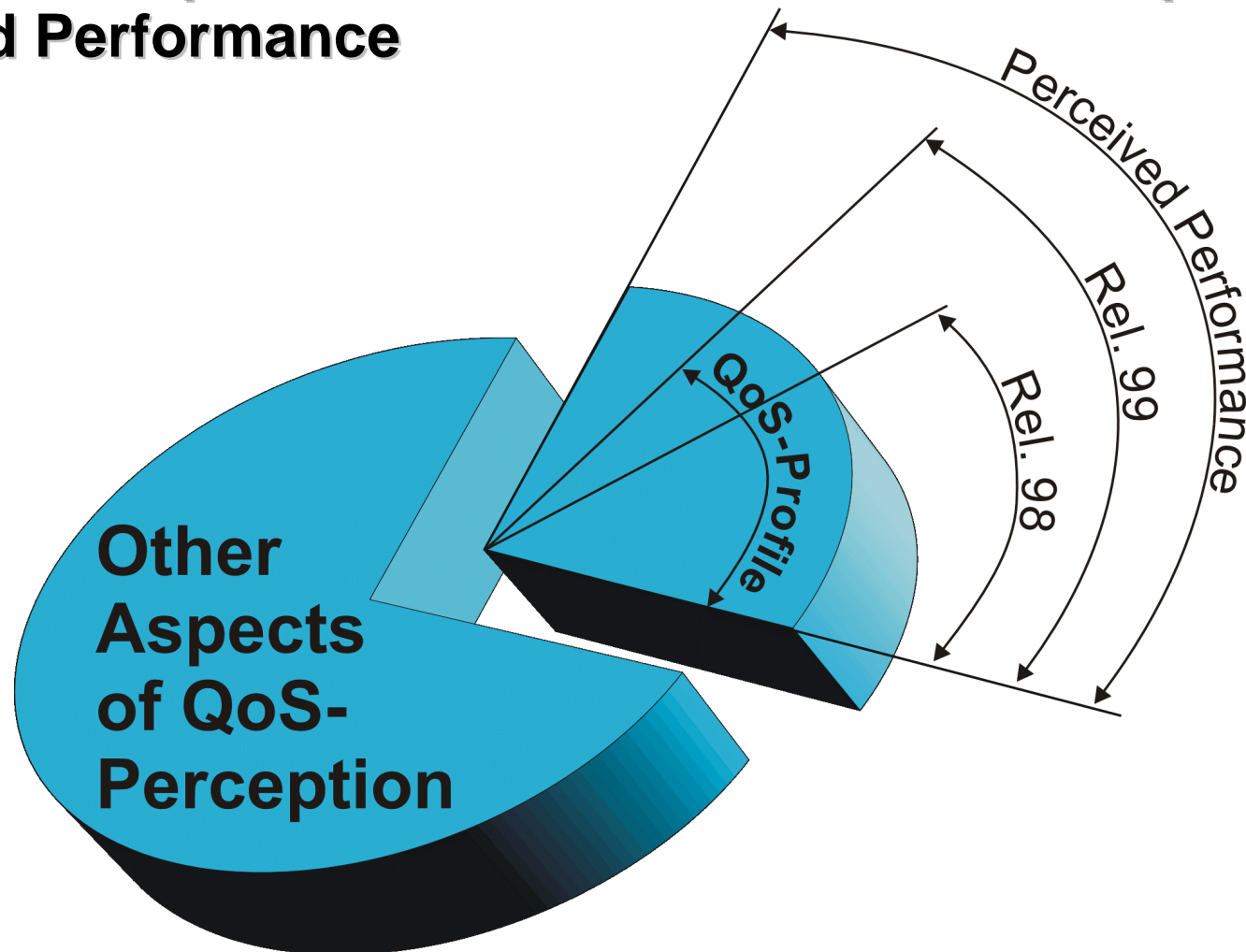
Some applications are basically out of scope for GPRS users. Surfing the internet with PDA-devices is no real pleasure because of the limited user interface (screen, keyboard) and for some applications GPRS is simply too slow.

Perceived Performance

For the end user the performance is basically how much content can be transmitted in what time. GPRS users will also soon recognize the inherent latency of GPRS to translate a user input into a reaction, into a data transmission.

The following chapters will primarily deal with the measurement, consideration and optimization of this perceived performance.

The Relationship between QoS-Profile, QoS-Perception and Perceived Performance

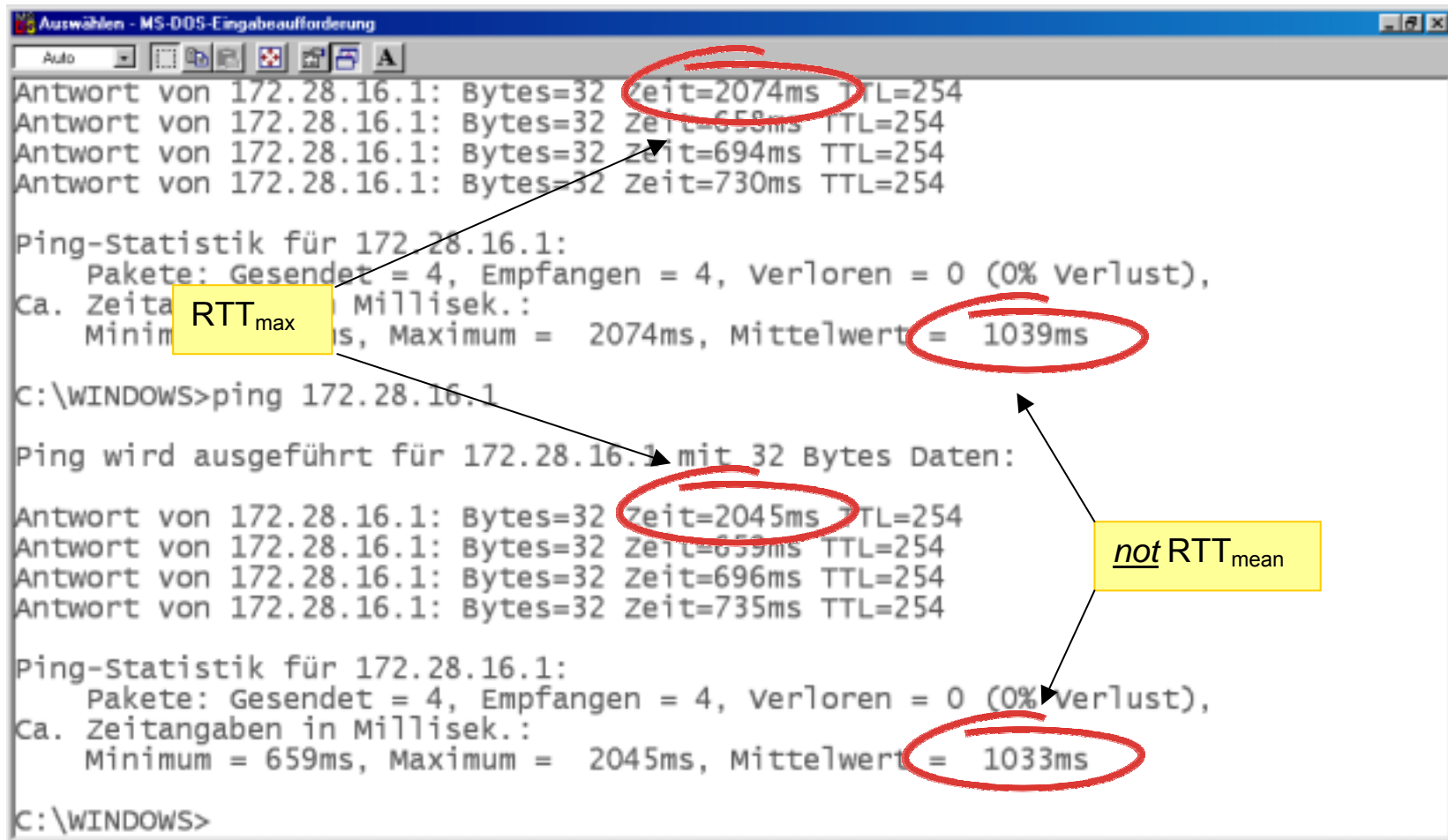


The Relationship between QoS-Profile, QoS-Perception and Perceived Performance

The diagram illustrates the relationship quite nicely:

- ⇒ Neither the QoS-profile Rel. 98 nor the QoS-profile Rel. 99 do match the perceived performance (⇔ how much content in which time).
- ⇒ And the QoS-perception of the end-user by far exceeds any definable QoS-profile.

Maximum and Mean Round Trip Time



```
Auswählen - MS-DOS-Eingabeaufforderung
Auto
Antwort von 172.28.16.1: Bytes=32 Zeit=2074ms TTL=254
Antwort von 172.28.16.1: Bytes=32 Zeit=658ms TTL=254
Antwort von 172.28.16.1: Bytes=32 Zeit=694ms TTL=254
Antwort von 172.28.16.1: Bytes=32 Zeit=730ms TTL=254

Ping-Statistik für 172.28.16.1:
Pakete: Gesendet = 4, Empfangen = 4, Verloren = 0 (0% Verlust),
Ca. Zeitangaben in Millisek.:
Minimum = 658ms, Maximum = 2074ms, Mittelwert = 1039ms

C:\WINDOWS>ping 172.28.16.1

Ping wird ausgeführt für 172.28.16.1 mit 32 Bytes Daten:

Antwort von 172.28.16.1: Bytes=32 Zeit=2045ms TTL=254
Antwort von 172.28.16.1: Bytes=32 Zeit=659ms TTL=254
Antwort von 172.28.16.1: Bytes=32 Zeit=696ms TTL=254
Antwort von 172.28.16.1: Bytes=32 Zeit=735ms TTL=254

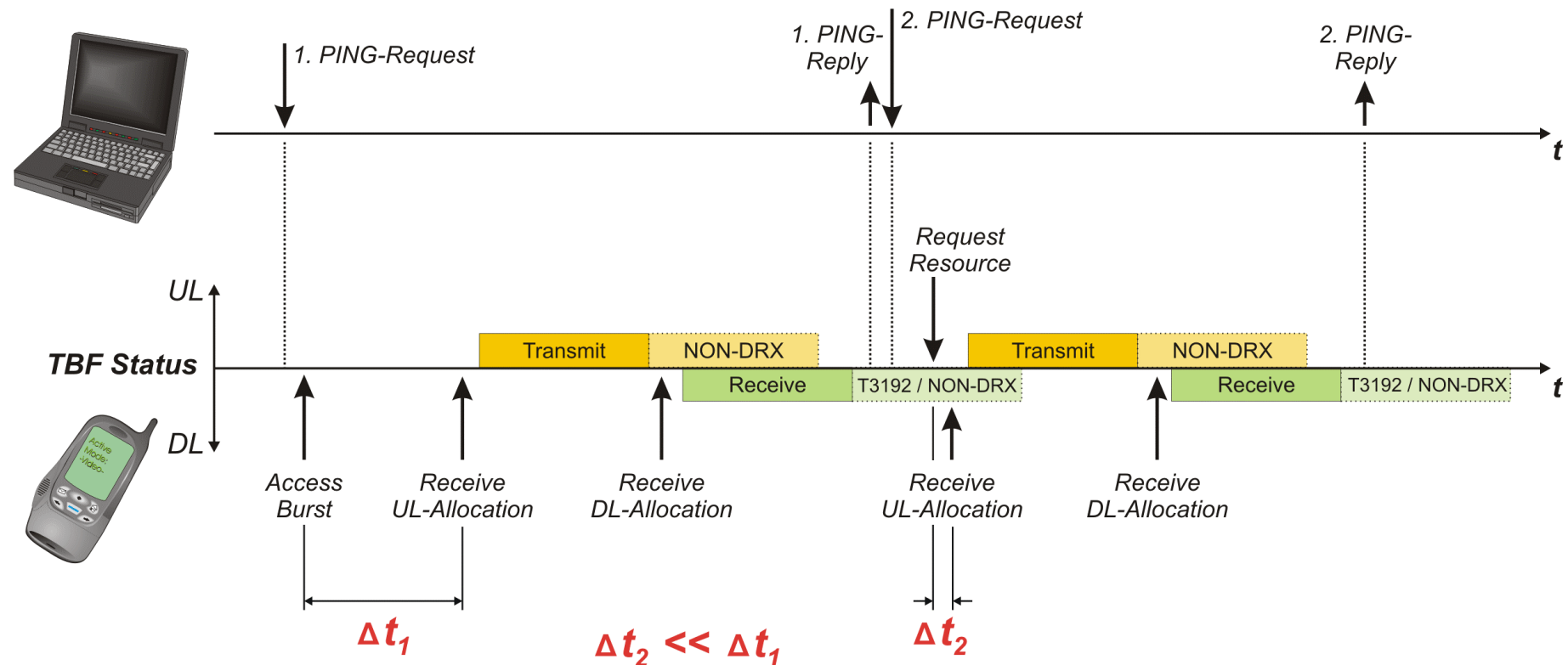
Ping-Statistik für 172.28.16.1:
Pakete: Gesendet = 4, Empfangen = 4, Verloren = 0 (0% Verlust),
Ca. Zeitangaben in Millisek.:
Minimum = 659ms, Maximum = 2045ms, Mittelwert = 1033ms

C:\WINDOWS>
```

Maximum and Mean Round Trip Time

The screenshot illustrates the measurement of RTTmean and RTTmax using the ICMP PING-application. Please note the essential discrepancy between the initial response from the router and the following responses.

Reasons for the Huge Difference between RTT_{mean} and RTT_{max}

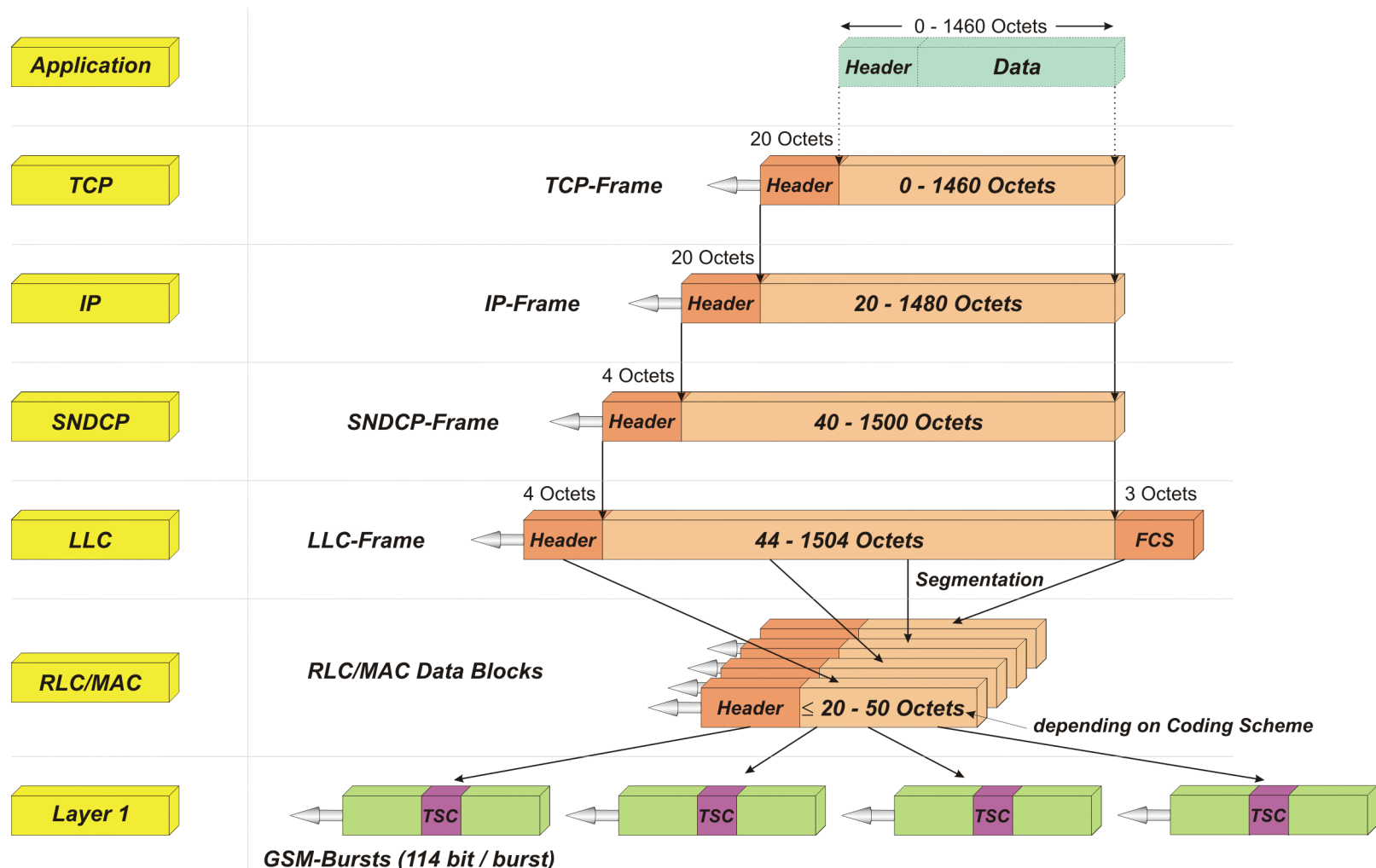


Reasons for the Huge Difference between RTT_{mean} and RTT_{max}

Before we continue let us investigate in more detail what causes the huge difference between RTT_{mean} and RTT_{max} :

- ⇒ We assume that the mobile station is currently in packet idle mode.
- ⇒ Initially, the mobile station will receive the PING-Request from the terminal equipment.
- ⇒ To suit this transmission request, the mobile station will send an access burst to the network, asking for the allocation of uplink resources.
- ⇒ As can be seen, the time Δt_1 is quite long and varies with the used access method (One-Phase / Two-Phase), with the network load and it depends on whether the access is performed using PRACH or RACH.
- ⇒ The mobile station will use the allocated uplink resource to convey the PING-Request to the PCU which will forward it through the network. Finally, the PING-Request reaches its destination (the router behind the GGSN in this case).
- ⇒ This router will respond to the ECHO-Request by sending an ECHO-Reply to the terminal equipment.
- ⇒ When this ECHO-Reply reaches finally the PCU, the uplink resources are obviously already released but the mobile station is still in NON-DRX-mode and in GMM-Ready State (depending on the setting of the respective timers).
- ⇒ Accordingly, the mobile station does not need to be paged but receives a direct resource allocation which speeds up the process. On the allocated DL-resources, the PCU will convey the ECHO-Reply to the mobile station which in turn will forward it to the terminal equipment.
- ⇒ Having received a reply for its initial request, the terminal equipment will send the next ECHO-Request.
- ⇒ Depending on whether the downlink TBF is still active or whether T3192 or the NON-DRX-timer are currently running, the mobile station will ask for uplink resources through another access procedure or by using the uplink PACCH of the downlink TBF.
- ⇒ Because of this, the mobile station may obtain the resources faster and hence Δt_2 is much smaller than Δt_1 .

Performance Leakage within the GPRS Protocol Stack



Performance Leakage within the GPRS Protocol Stack

The figure illustrates the performance leakage of the GPRS protocol stack and the performance leakage which is caused by the TCP/IP-headers.

RLCMAC

On RLC/MAC level, the maximum throughput rate depends on the available coding schemes (CS-1: 8 kbit/s x TS; CS-2: 12 kbit/s x TS; CS-3: 14.4 kbit/s x TS; CS-4: 20 kbit/s x TS). Please refer to [1] for more details.

LLC

On LLC level, the maximum throughput rate depends on the maximum size of the information field which is controlled by the parameter N201 and on the type of LLC-frame which is used for data transmission. Throughout the class we will assume that LLC-UI-frames are used (unacknowledged transmission).

SNDCP

SNDCP will add another header to the IP-frame.

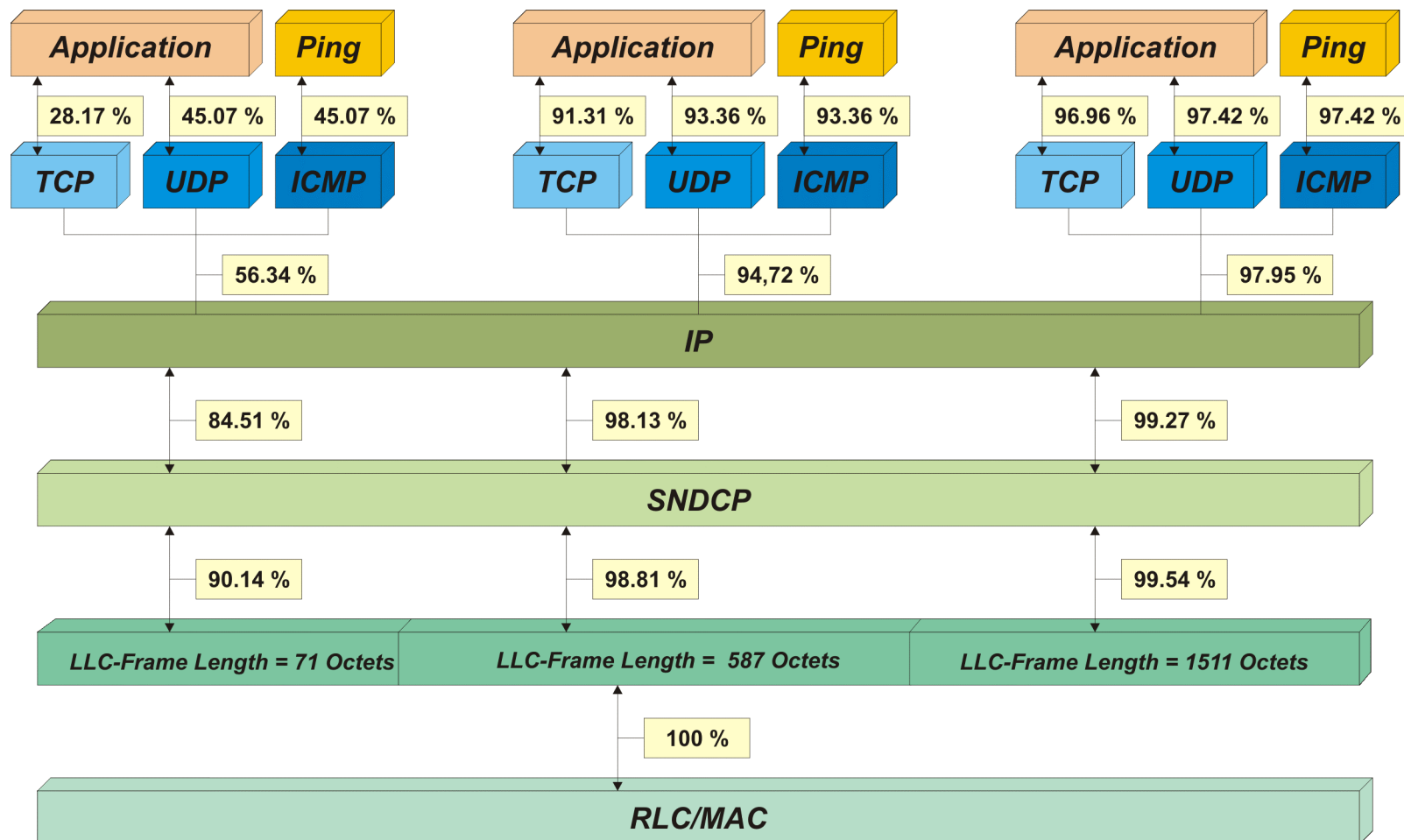
IP

For the internet protocol, we need to consider another header with 20 octets length. Optional fields of the header may possibly need to be added.

TCP

For the transmission control protocol, we need to consider another header with 20 octets length. Optional fields of the header may possibly need to be added.

Relative Performance Leakage

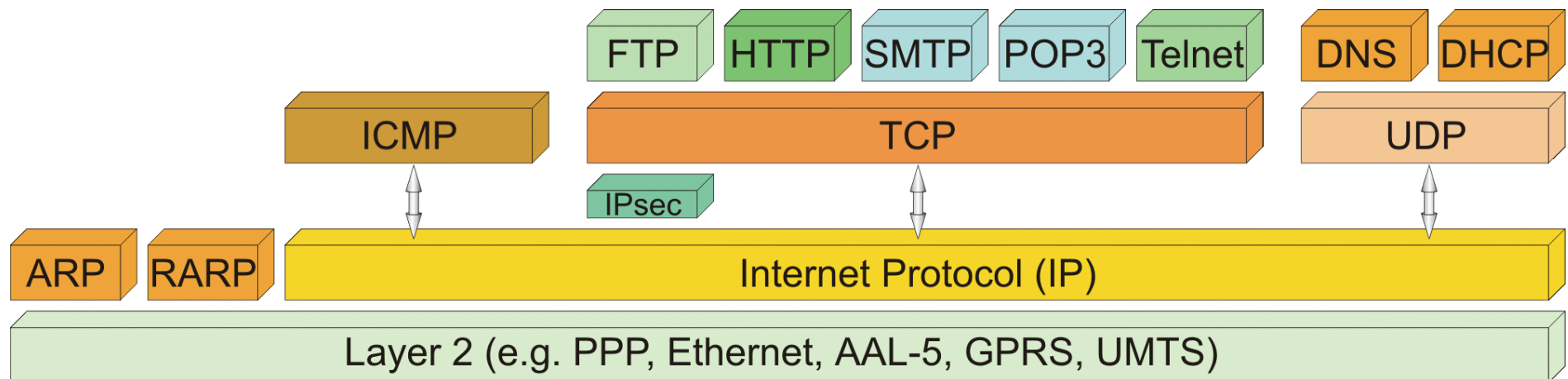


Relative Performance Leakage

The relative performance leakage considers the header lengths from the previous pages but adds a few more details:

- ⇒ The figure illustrates three columns which start at the LLC-layer and which depend on the actual length of the LLC-frame.
- ⇒ We also added UDP and ICMP as layer 4 protocols in addition to TCP (we apologize for calling ICMP a layer 4 protocol)

Introducing the IP-Protocol Stack



Introducing the IP-Protocol Stack

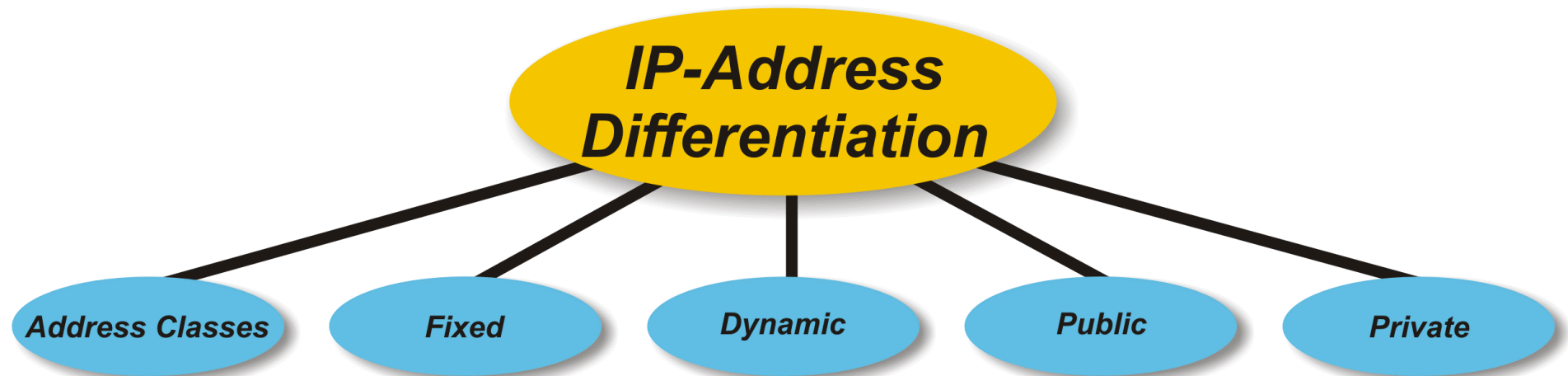
A simplified view of the IP-protocol stack is provided in the figure. The Internet Protocol (IP) itself can be located on top of almost any available layer 2-protocol like for example:

- ⇒ PPP (⇔ Point-to-Point Protocol / RFC 1661)
- ⇒ Ethernet (⇔ IEEE 802.3)
- ⇒ AAL-5 (⇔ ATM / ITU-T I.363.5 (6), Q.2110 (4))
- ⇒ or even on mobile bearers like GPRS or UMTS (⇔ 3GPP recommendations).

This flexibility of IP makes it the preferred network layer solution of the starting 21st century.

However, the term IP-protocol stack rather relates to IP and the higher layers which make use of the IP's networking capabilities. In the following sub-clauses we will take a more detailed look at the IP itself and the protocols on top of it.

IP-Addresses



IP-Addresses

One of the key issues of the internet protocol are the 32 bit long IP-addresses. The IP-address shall identify any given user of the internet uniquely at any given moment in time. Due to the enormous growth of the internet community, IP-addresses are becoming a scarce resource. Special features have been put in place to make more efficient use of the available IP-address range. IP-addresses can be distinguished by the following means:

- **Address Classes**

Already at the very beginning of the internet, five different address classes (A, B, C, D, E) were defined which imposed a hierarchical and unfortunately uneconomic structure upon all internet addressing. The ICANN (Internet Corporation for Assigned Names and Numbers) is now in charge for allocating IP-addresses to individuals and organizations. Please note that ICANN replaced IANA (Internet Assigned Number Authority) in 1999.

- **Fixed and Dynamic IP-Addresses**

Any IP-address is either fixed or dynamic. In the “fixed” case, each user has at least one IP-address which cannot be used by anybody else even when this user has switched off the computer. Opposed to that, when dynamic addressing is used, any user of a system will obtain another IP-address any time the host computer is switched on. Switching off the computer also means to return the IP-address, used, and therefore making it available again to another user. We will later get back to the issue of dynamic IP-addressing.

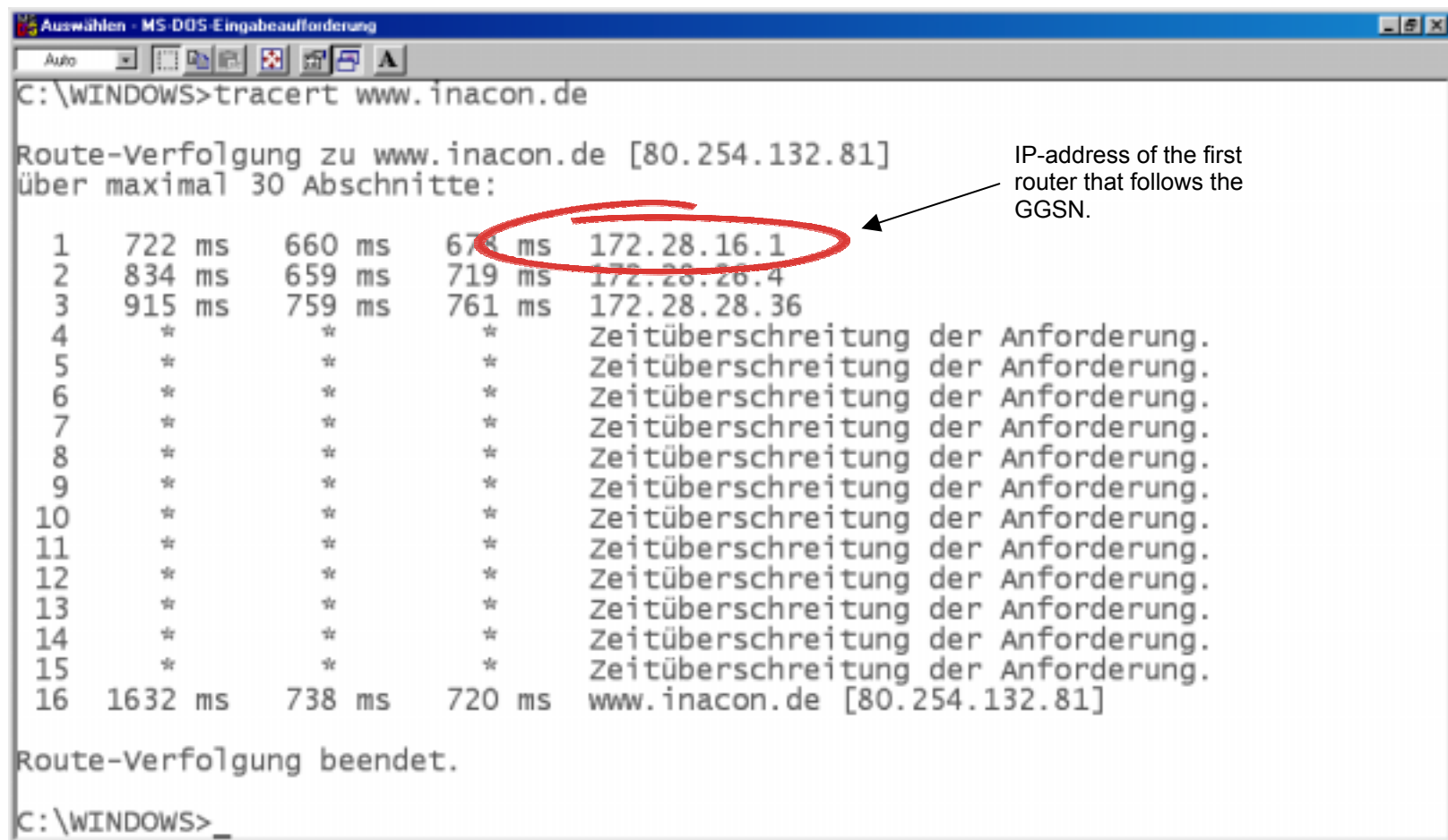
- **Public and Private IP-Addresses**

Yet another means to distinguish IP-addresses is to have either public or private IP-addresses. Private IP-addresses are unique only and exclusively in the very network where they are used. However, such an IP-address cannot be used for communication purposes towards hosts or servers outside this network. Special means are required to allow such users access to services outside the local network.

Opposed to that, public IP-addresses are unique worldwide and can be used for any access to services inside and outside the local network. We will later see more details about private and public IP-addressing.

In this context we will focus on the 32 bit long IPv4-addresses. In addition, the IETF and major industry leaders are preparing the roll-out of IPv6 [RFC 1883], also referred to as IPnG (\Leftrightarrow for next generation). IPv6 comes with the major advantage of using IP-addresses with a length of 128 bit compared to 32 bit in IPv4.

Use Trace Route to Determine the IP-Address of the 1st Router



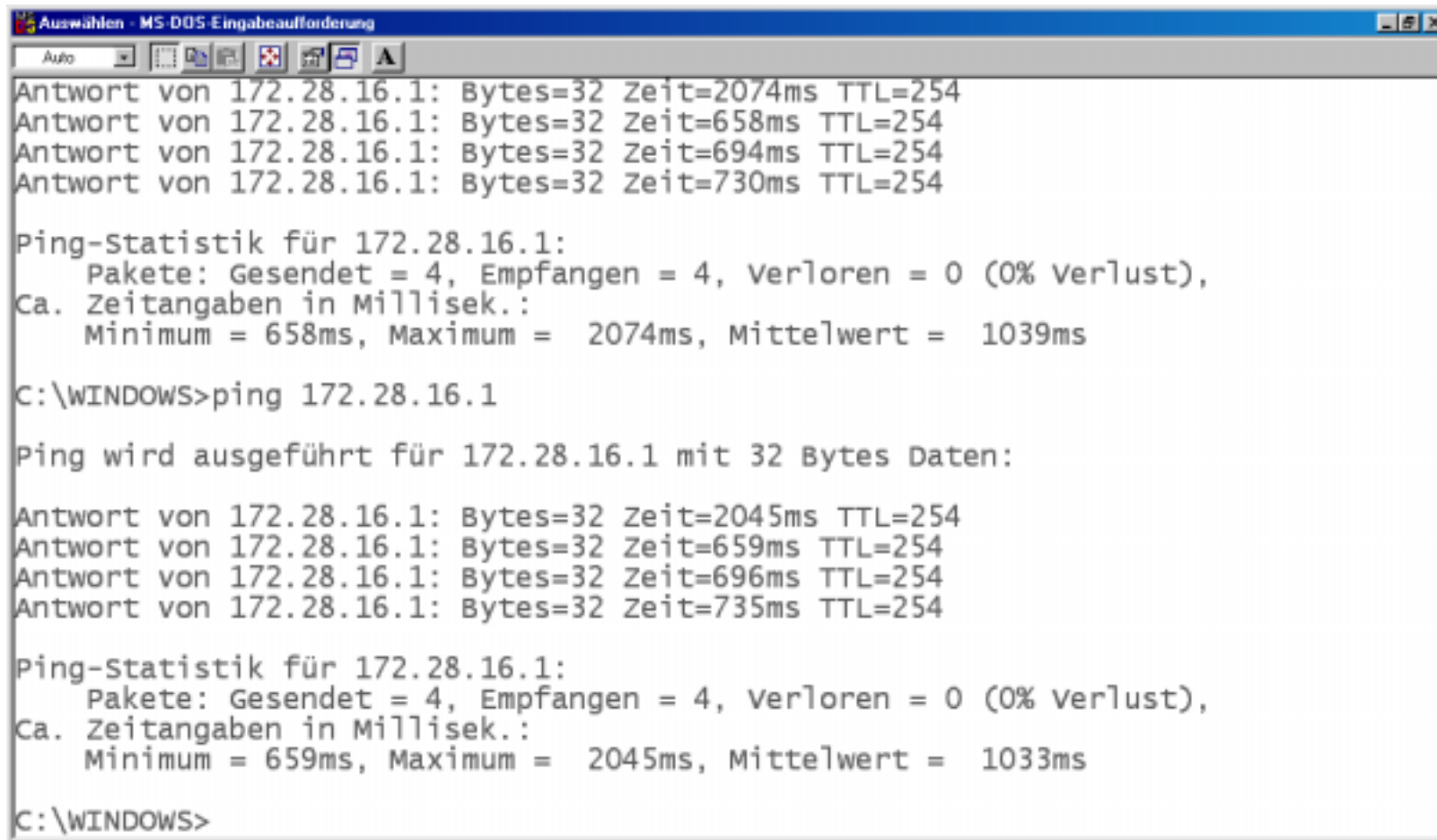
```
Auswählen - MS-DOS Eingabeaufforderung
C:\WINDOWS>tracert www.inacon.de

Route-Verfolgung zu www.inacon.de [80.254.132.81]
über maximal 30 Abschnitte:

 1  722 ms    660 ms    678 ms    172.28.16.1
 2  834 ms    659 ms    719 ms    172.28.28.4
 3  915 ms    759 ms    761 ms    172.28.28.36
 4  *         *         *         Zeitüberschreitung der Anforderung.
 5  *         *         *         Zeitüberschreitung der Anforderung.
 6  *         *         *         Zeitüberschreitung der Anforderung.
 7  *         *         *         Zeitüberschreitung der Anforderung.
 8  *         *         *         Zeitüberschreitung der Anforderung.
 9  *         *         *         Zeitüberschreitung der Anforderung.
10  *         *         *         Zeitüberschreitung der Anforderung.
11  *         *         *         Zeitüberschreitung der Anforderung.
12  *         *         *         Zeitüberschreitung der Anforderung.
13  *         *         *         Zeitüberschreitung der Anforderung.
14  *         *         *         Zeitüberschreitung der Anforderung.
15  *         *         *         Zeitüberschreitung der Anforderung.
16 1632 ms    738 ms    720 ms    www.inacon.de [80.254.132.81]

Route-Verfolgung beendet.
C:\WINDOWS>_
```

Ping with 32 Octets of Data (no Segmentation)

A screenshot of a Windows command prompt window titled "Auswählen - MS-DOS Eingabeaufforderung". The window shows the results of a ping command to the IP address 172.28.16.1. The output is as follows:

```
Auswählen - MS-DOS Eingabeaufforderung
Auto
Antwort von 172.28.16.1: Bytes=32 Zeit=2074ms TTL=254
Antwort von 172.28.16.1: Bytes=32 Zeit=658ms TTL=254
Antwort von 172.28.16.1: Bytes=32 Zeit=694ms TTL=254
Antwort von 172.28.16.1: Bytes=32 Zeit=730ms TTL=254

Ping-Statistik für 172.28.16.1:
    Pakete: Gesendet = 4, Empfangen = 4, Verloren = 0 (0% Verlust),
    Ca. Zeitangaben in Millisek.:
        Minimum = 658ms, Maximum = 2074ms, Mittelwert = 1039ms

C:\WINDOWS>ping 172.28.16.1

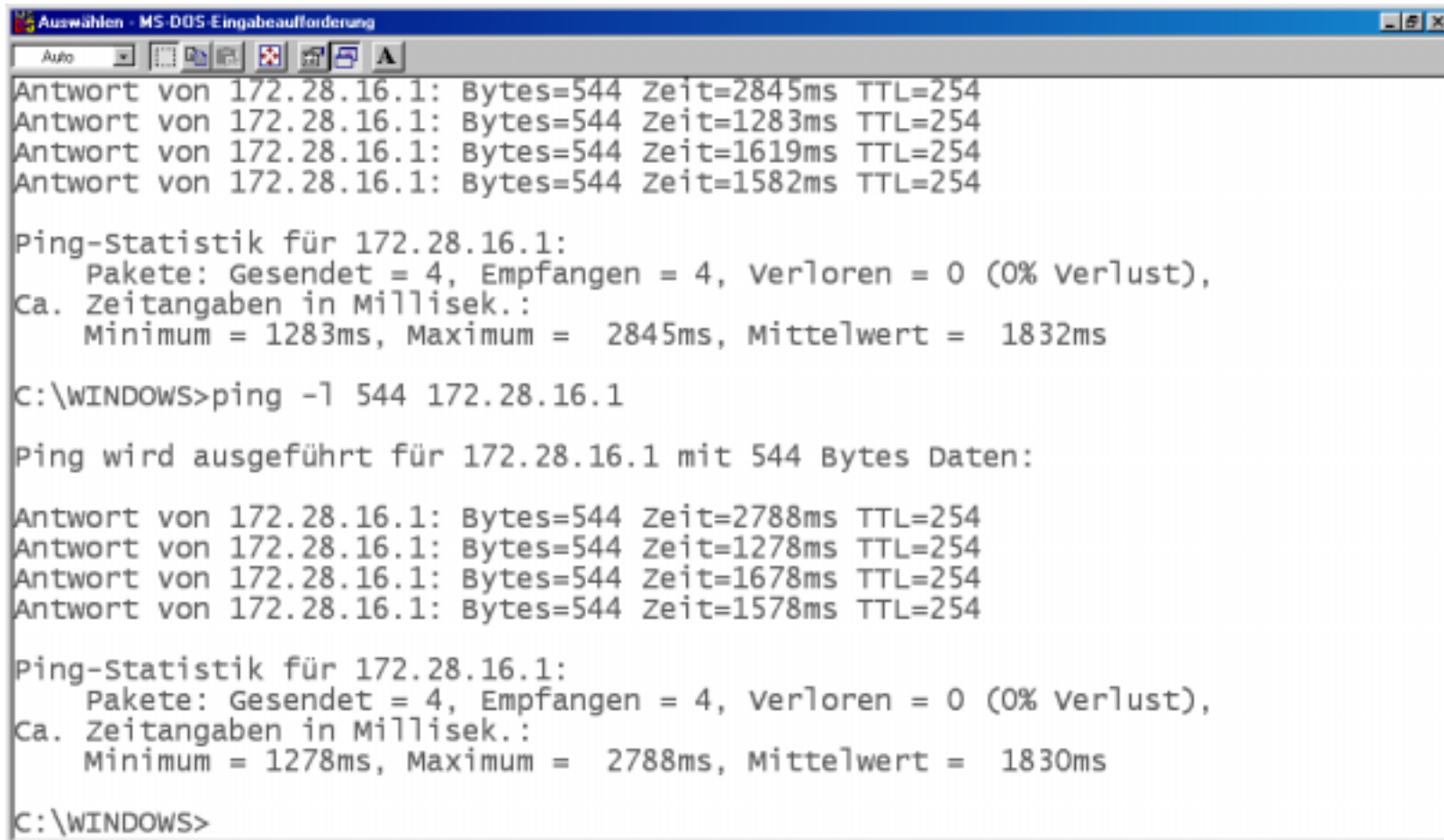
Ping wird ausgeführt für 172.28.16.1 mit 32 Bytes Daten:

Antwort von 172.28.16.1: Bytes=32 Zeit=2045ms TTL=254
Antwort von 172.28.16.1: Bytes=32 Zeit=659ms TTL=254
Antwort von 172.28.16.1: Bytes=32 Zeit=696ms TTL=254
Antwort von 172.28.16.1: Bytes=32 Zeit=735ms TTL=254

Ping-Statistik für 172.28.16.1:
    Pakete: Gesendet = 4, Empfangen = 4, Verloren = 0 (0% Verlust),
    Ca. Zeitangaben in Millisek.:
        Minimum = 659ms, Maximum = 2045ms, Mittelwert = 1033ms

C:\WINDOWS>
```

Ping with 544 Octets of Data (still no Segmentation)

A screenshot of a Windows command prompt window titled "Auswählen - MS-DOS Eingabeaufforderung". The window shows the results of a ping command with a size of 544 bytes. The first four lines show individual ping responses with times ranging from 1283ms to 2845ms. The fifth line shows the ping statistics for 172.28.16.1, indicating 4 packets sent, 4 received, and 0% loss, with an average time of 1832ms. The sixth line shows the command "C:\WINDOWS>ping -l 544 172.28.16.1". The seventh line shows the message "Ping wird ausgeführt für 172.28.16.1 mit 544 Bytes Daten:". The next four lines show individual ping responses with times ranging from 1278ms to 2788ms. The final line shows the ping statistics for 172.28.16.1, indicating 4 packets sent, 4 received, and 0% loss, with an average time of 1830ms. The command prompt ends with "C:\WINDOWS>".

```
Auswählen - MS-DOS Eingabeaufforderung
Auto
Antwort von 172.28.16.1: Bytes=544 Zeit=2845ms TTL=254
Antwort von 172.28.16.1: Bytes=544 Zeit=1283ms TTL=254
Antwort von 172.28.16.1: Bytes=544 Zeit=1619ms TTL=254
Antwort von 172.28.16.1: Bytes=544 Zeit=1582ms TTL=254

Ping-Statistik für 172.28.16.1:
    Pakete: Gesendet = 4, Empfangen = 4, Verloren = 0 (0% Verlust),
    Ca. Zeitangaben in Millisek.:
        Minimum = 1283ms, Maximum = 2845ms, Mittelwert = 1832ms

C:\WINDOWS>ping -l 544 172.28.16.1

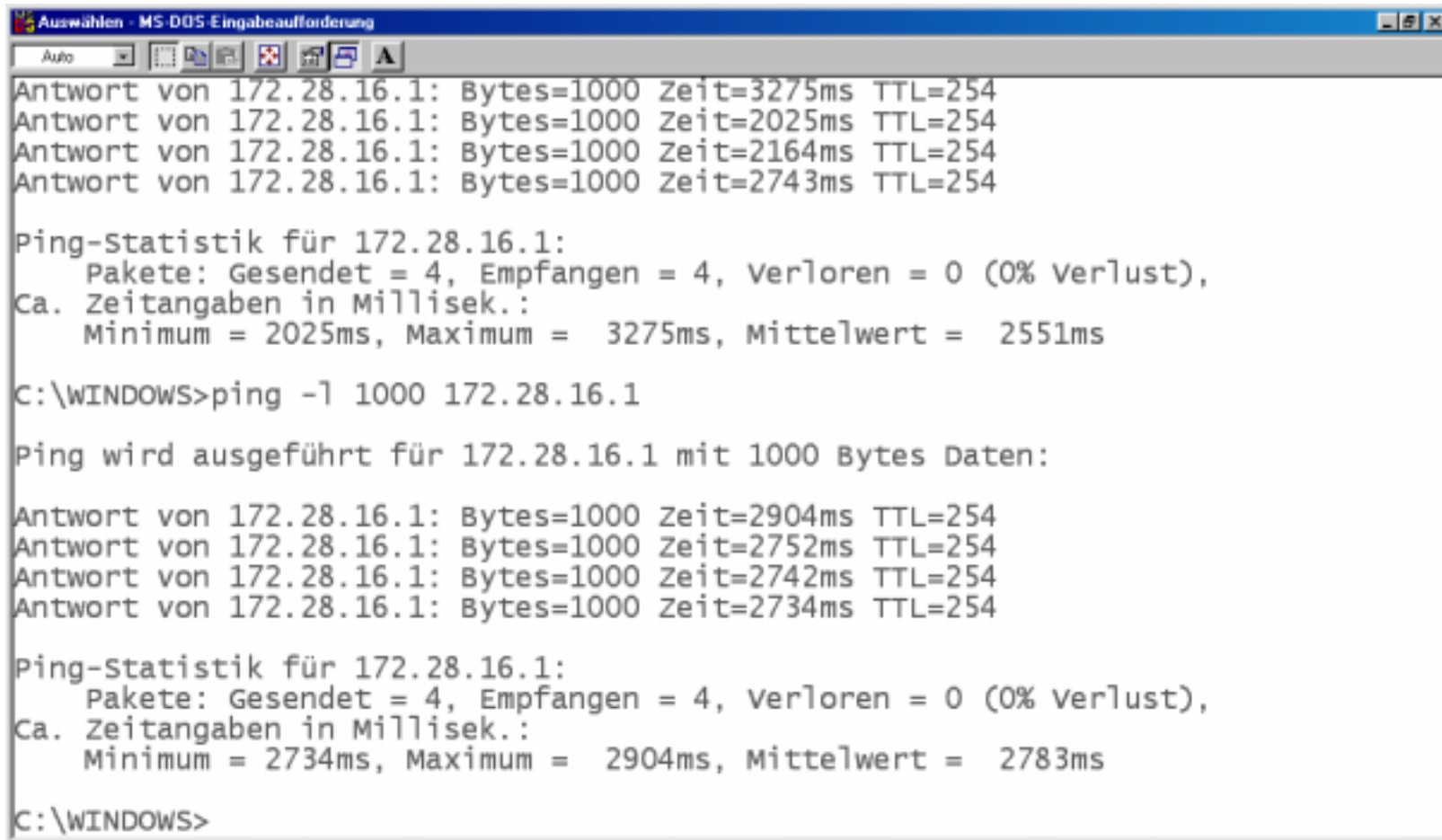
Ping wird ausgeführt für 172.28.16.1 mit 544 Bytes Daten:

Antwort von 172.28.16.1: Bytes=544 Zeit=2788ms TTL=254
Antwort von 172.28.16.1: Bytes=544 Zeit=1278ms TTL=254
Antwort von 172.28.16.1: Bytes=544 Zeit=1678ms TTL=254
Antwort von 172.28.16.1: Bytes=544 Zeit=1578ms TTL=254

Ping-Statistik für 172.28.16.1:
    Pakete: Gesendet = 4, Empfangen = 4, Verloren = 0 (0% Verlust),
    Ca. Zeitangaben in Millisek.:
        Minimum = 1278ms, Maximum = 2788ms, Mittelwert = 1830ms

C:\WINDOWS>
```

Ping with 1000 Octets of Data (Segmentation)

A screenshot of a Windows command prompt window titled "Auswählen - MS-DOS Eingabeaufforderung". The window shows the results of a ping command with a segment size of 1000 bytes. The first four lines show individual ping responses with times ranging from 2025ms to 3275ms. The fifth line shows the ping statistics for 172.28.16.1, indicating 4 packets sent, 4 received, and 0 lost. The sixth line shows the command being executed: "C:\WINDOWS>ping -l 1000 172.28.16.1". The seventh line shows the message "Ping wird ausgeführt für 172.28.16.1 mit 1000 Bytes Daten:". The next four lines show individual ping responses with times ranging from 2734ms to 2904ms. The final line shows the ping statistics for 172.28.16.1, indicating 4 packets sent, 4 received, and 0 lost.

```
Auswählen - MS-DOS Eingabeaufforderung
Auto
Antwort von 172.28.16.1: Bytes=1000 Zeit=3275ms TTL=254
Antwort von 172.28.16.1: Bytes=1000 Zeit=2025ms TTL=254
Antwort von 172.28.16.1: Bytes=1000 Zeit=2164ms TTL=254
Antwort von 172.28.16.1: Bytes=1000 Zeit=2743ms TTL=254

Ping-Statistik für 172.28.16.1:
    Pakete: Gesendet = 4, Empfangen = 4, Verloren = 0 (0% Verlust),
    Ca. Zeitangaben in Millisek.:
        Minimum = 2025ms, Maximum = 3275ms, Mittelwert = 2551ms

C:\WINDOWS>ping -l 1000 172.28.16.1

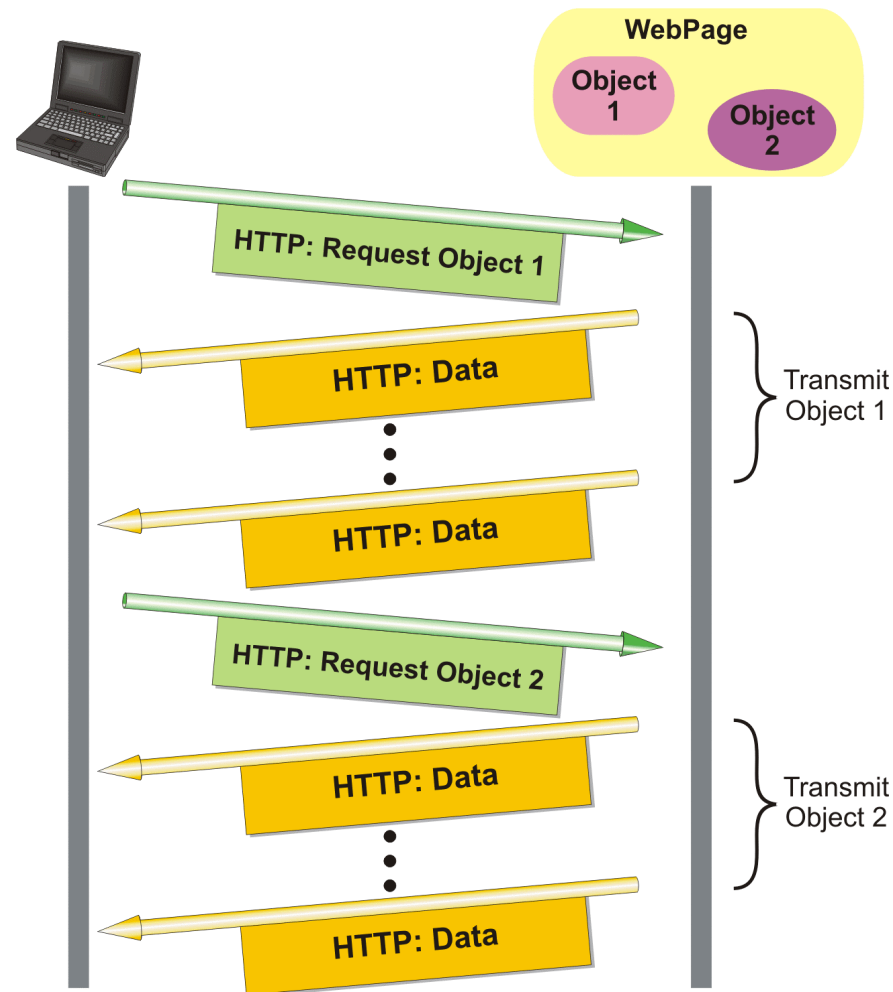
Ping wird ausgeführt für 172.28.16.1 mit 1000 Bytes Daten:

Antwort von 172.28.16.1: Bytes=1000 Zeit=2904ms TTL=254
Antwort von 172.28.16.1: Bytes=1000 Zeit=2752ms TTL=254
Antwort von 172.28.16.1: Bytes=1000 Zeit=2742ms TTL=254
Antwort von 172.28.16.1: Bytes=1000 Zeit=2734ms TTL=254

Ping-Statistik für 172.28.16.1:
    Pakete: Gesendet = 4, Empfangen = 4, Verloren = 0 (0% Verlust),
    Ca. Zeitangaben in Millisek.:
        Minimum = 2734ms, Maximum = 2904ms, Mittelwert = 2783ms

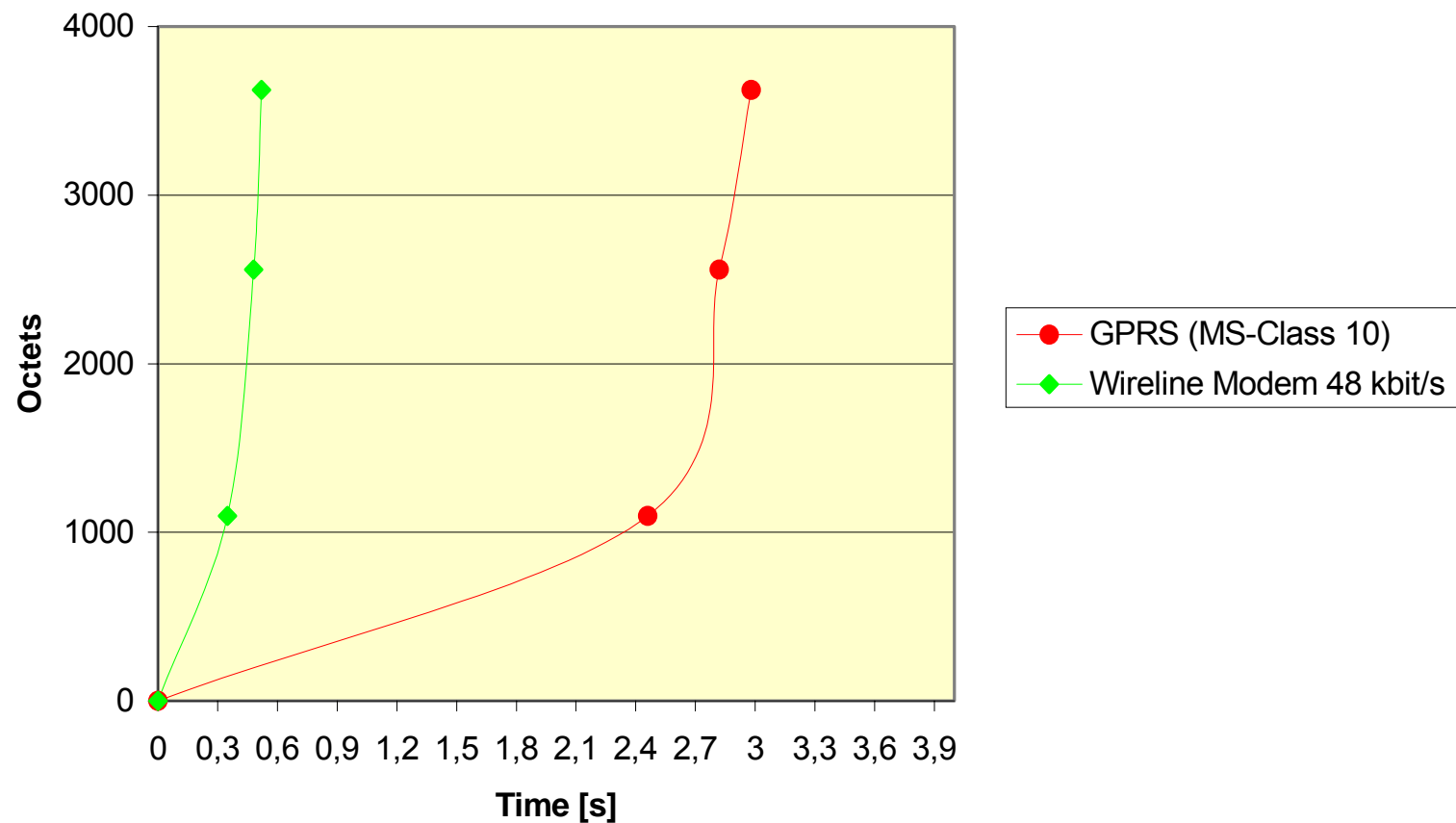
C:\WINDOWS>
```

Download of a given Web Page



Impact of GPRS Specific Delays on HTTP-Performance

HTTP-Download Times (\Leftrightarrow File Size: 3625 Octets)



■ **Index:**

A

ACCESS_BURST_TYPE	250
ACK (TCP-flag)	136
Acknowledgement Number (TCP)	134
Active Open	116
Address Classes (IP)	50
Address Mask Reply (ICMP-Message)	102
Address Mask Request (ICMP-Message)	102

B

Background Applications	22
Bc	276
Be	276
BS_CV_MAX	254
BSSGP	272

C

Checksum (ICMP)	96
Checksum (TCP)	136
CIR	276
Class A, B, C, D, E (IP)	50
Code (ICMP)	96
coding scheme	36
Committed Information Rate	276
congestion avoidance algorithm (TCP)	148, 154

congestion window	152
connection establishment (TCP)	116
Conversational Applications	22
Cost (IP-Header)	80
cwnd	152

D

DE-bit	276
Delay (IP-Header)	80
Destination Address (IP-Header)	90
destination port (TCP)	134
Destination Unreachable (ICMP-Message)	98
DF-flag = "Do not Fragment"	86
DHCP	56
DHCPACK-message	58
DHCPDISCOVER-message	58
DHCPOFFER-message	58
DHCPRELEASE-message	58
DHCPREQUEST-message	58
Differentiated Services	82
Differentiated Services Code Points	82
Dotted Decimal Notation	50
DS	82
DSCP	82
Dynamic Allocation	56, 264
Dynamic Host Configuration Protocol	56

E

Echo Reply (ICMP-Message)	98
Echo Request (ICMP-Message)	98
EnablePMTUDiscovery	168
encryption	290
Extended Dynamic Allocation	264

F

fast recovery algorithm (TCP)	148
fast retransmit algorithm (TCP)	148, 158
FIN (TCP-flag)	136
Fixed Allocation	264
Flags (IP-Header)	86
Flow Control	114
flow control (BSSGP)	280
Fragment Offset	86

H

half-close	124, 126
Header Checksum (IP-Header)	90
Header Length (TCP)	136
Host ID	50

I

IANA	48
ICANN	48
ICMP	96
Identification (IP-Header)	86
IHL (IP-Header)	78
Information Reply (ICMP-Message)	102

Information Request (ICMP-Message)	102
Initial Sequence Number (TCP)	116
Interactive Applications	22
Internet Assigned Number Authority	48
Internet Corporation for Assigned Names and Numbers	48
Internet Header Length (IP-Header)	78
Internet Timestamp (IP-Option)	92
IP-Header	74
IPnG	48
ISN (TCP)	116

K

Key Performance Indicators	40
KPI	40

L

LLC	284
Loose Source Route	92

M

Maximum Round Trip Time	12
maximum segment size (TCP)	116, 138
Maximum Transmit Unit	88
Mean Round Trip Time	12
MF-flag = "More Fragments"	86
MSS (TCP)	116, 138, 288
MTU	88
MTU Probe (IP-Option)	92
MTU Reply (IP-Option)	92

N

N200	292
N201	288
N201-I	288
N201-U	288
N3102	262
Nagle algorithm(TCP)	148
NAT	66
Net ID	50
Network Address Translation	66
Network Control Order	258
Network Service	272
NON-DRX-Timer	260

O

Options (IP)	92
Options (TCP)	138

P

PACK_CTRL_ACK-messages	256
PAN_DEC	262
PAN_INC	262
PAN_MAX	262
Parameter Problem on a Datagram (ICMP-Message)	100
Passive Open	116
performance leakage	36
Ping-application	98
Precedence (IP-Header)	80
Protocol (IP-Header)	90
PSH (TCP-flag)	136
pull services	68

push services	68
---------------------	----

R

Raw Throughput Rate	12
Ready State	300
Ready Timer	300
Record Route (IP-Option)	92
Redirect (ICMP-Message)	98
Reliability (IP-Header)	80
resumption (GPRS)	278
RLC/MAC	246
Router Advertisement (ICMP-Message)	100
Router Alert (IP-Option)	92
Router Solicitation (ICMP-Message)	100
RST (TCP-flag)	136
RTO	144
RTT	144
RTT _{max}	12

S

Sequence Number (TCP)	134
sliding window mechanism (TCP)	142
slow start algorithm (TCP)	148, 154
slow start threshold	152
socket	114
Source Address (IP-Header)	90
source port (TCP)	134
Source Quench (ICMP-Message)	98
special IP-addresses	52
SRTT	146
ssthresh	152
Streaming Applications	22

Strict Source Route (IP-Option)	92
suspension (GPRS)	278
SYN (TCP-flag)	136

T

T200	292
T3168	266
T3192	268
T3312	302
T3314	300
Tc	276
TCP	114
TcpDelAckTicks	150
three way handshake (TCP)	116
Throughput (IP-Header)	80
Throughput Rate	12
Time Exceeded for a Datagram (ICMP-Message)	100
Time To Live (IP-Header)	90
Timestamp Reply (ICMP-Message)	102
Timestamp Request (ICMP-Message)	100
TOS	80

Total Length (IP-Header)	78
Traceroute (IP-Option)	92
traffic classes	22
TTL (IP-Header)	90
Type (ICMP)	96
Type of Service field (IP-Header)	80

U

URG (TCP-flag)	136
Urgent Pointer	136

V

Version (IP-Header)	78
---------------------------	----

W

Window Size	136
Windows Parameter	
TcpInitialRTT	146

