Transmission Control **Protocol** (TCP)

1

4

6

ТСР: Услуги

- RFCs 793, 1122, 1323, 2018, 2581, 2873, 2988, 3168, 4614
- Надеждна комуникация между два приложни процеса
 - Гарантира, че данните ще бъдат доставени без загуба, дублирания

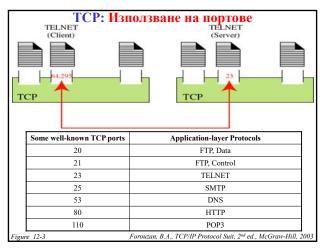
 - или грешки.
 През множество надеждни и/или ненадеждни мрежи С контролиране на потока, грешките и задръстванията
- С използване на съединение
 - Приложението изисква съединение за трансфер на данните
 - Установяване на надеждно съединение
 - 3-кратно ръкостискане гарантира надежден и синхронизиран старт на комуникацията между двете крайни точки Съединението се установява между двойка сокети

 - Сокет = IP адрес + номер на порт
 <u>1 сокет</u> може да се използва за <u>няколко съединения</u> едновременно
 - Разпадане на съединението

 - Гарантиране доставката на всички данни преди разпадане на съединението
 - 4(3)-кратно ръкостискане
- От точка до точка (Е2Е)
 - Съединението има 2 крайни точки, идентифицирани чрез сокети.
- Няма поддръжка на multicasting и broadcasting!

3

ТСР: Услуги (прод.) Пълен дуплекс Крайните точки обменят данни в двете посоки едновременно Всяка точка разполага с буфер за изпращане и получаване на данни Байтово-ориентирани Приложният процес предава своето съобщение към ТСР TĈP го разглежда като непрекъснат поток от байтове ТСР може да не запази границите на съобщението Може да изпрати съобщението веднага, или Може да го буферира с цел акумулиране на по-голям обем от данни за изпращане наведнъж Sending Receiving process process Stream of bytes Forouzan, B.A., TCP/IP Protocol Suit, 4th ed., McGraw-Hill,



5



- Данните от горния слой се разглеждат като логически <u>поток от байтове</u> Байтовете се <u>номерират по **mod** 2³²

 Започвайки от номер, произволно избран и анонсиран по време на установяване на съединението.</u>
- Данните се буферират от ТСР подателя и ТСР получателя
- Кръгов буфер Контрол на потока чрез плъзгащ се прозорец и кредитна схема
- за размера на прозореца Използване на *PUSH* флаг за принудително изпращане на всички данни, натрупани до момента в буфера (*end-of-block function*) Приложението може да укаже спешно предаване на данни (чрез URGENT флаг)

Forouzan, B.A., TCP/IP Protocol Suit, 4th ed., McGraw-Hill, 2010



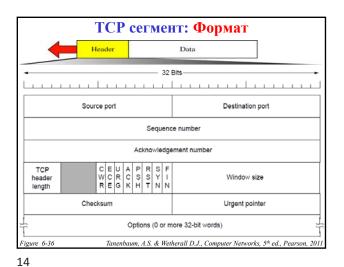
- Комуникиращите ТСР обекти могат да договорят

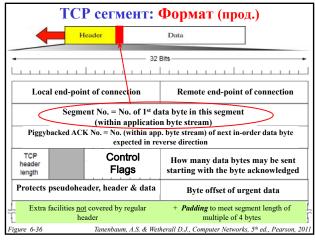
МАХ размер на сегмента (MSS)

- *MSS* ≤ 65515 B
- ТСР подателят може да реши да:
 - Акумулира данните от няколко записа в 1 сегмент, или
 - Раздели данните от 1 запис на няколко сегмента

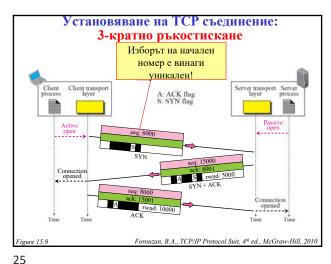
Forouzan, B.A., TCP/IP Protocol Suit, 4th ed., McGraw-Hill, 2010

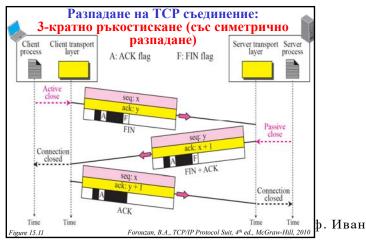
b. Иван

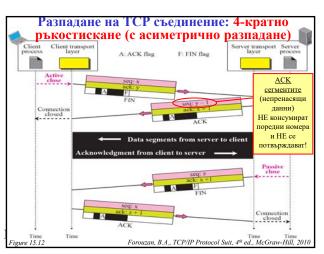












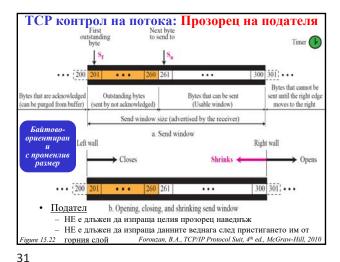
ТСР контрол на потока

29

TCP KOHTPOJI HA HOTOKA: Oбратна връзка

Application Sender Data flow Consumer Producer Segements are pushed Segements are pushed Segements are pushed Segements are pushed Flow control feedback

30



ТСР контрол на потока: Прозорец на получателя Next byte to be pulled by the proces ••• 200 201 301 ••• Bytes that can be ecceived from sender ytes that have already ulled by the process Receive window size (rwn ceived from send Allocated buffer Байтовоa. Receive window and allocated buffer ориентиран и Left wall размер Closes • • • 200 201 301 ... Получател b. Opening and closing of receive window - НЕ е длъжен да изпраща веднага потвърждение Може да изпрати потвърждение по всяко време Forouzan, B.A., TCP/IP Protocol Suit, 4th ed., McGraw-Hill, 2010 Figure 15.23

32



Не е пряко обвързано с АСК механизъм
 Специфична особеност на протоколите от транспортния слой

 Размерът на прозореца може да се увеличава или намалява от получателя динамично

 Въз основа на размера на прозореца.

 Нодателят спира да предава сегменти

 2 изключения
 Специален сегменти

 2 изключения
 Специален породо се изграща периодично за да се предявата перкупера получателя, че подателят все още очаква анопсиране на ненулев размер на прозореца.

 Подателят все още очаква напостране на ненулев размер на прозореца.

 Тиве прозореца получателя, че подателят все още очаква анопстране на ненулев размер на прозореца.

 Тиве прозореца получателя, че подателят все още очаква анопстране на ненулев размер на прозореца.

 Тиве прозореца получателя, че подателят все още очаква анопстране на ненулев размер на прозореца.

 Тиве прозореца прозореца проделение на ненулев размер на прозореца.

 Тиве прозореца прозореца проделение на ненулев размер на прозореца.

 Тиве прозореца проделение на ненулев размер на прозореца проделение на ненулев размер на прозореца.

 Тиве прозореца проделение на ненулев размер на прозореца проделение прозореца проделение проделени

ТСР: Синдром на глупавия прозорец Възниква, когато данните се обработват бавно от приложенията. Синдром, създаден от подателя. Когато предаващото приложение генерира данните бавно • Т.е. само по няколко байта наведнъж. Решение: Алгоритм на Nagle • Цел: подателят да не изпраща малки сегменти Синдром, създаден Receiver's buffer is full от <u>получателя.</u> Когато получаващото приложение Room for one more byte консумира данните бавно Header 2 решения: Алгоритм на Clark Забавено Цел: получателят да не иска малки сегменти 35

Синдром, създаден от подателя:

Алгоритм на Nagle (прилаган от подателя)

- Стъпка 1
 - Изпращане на 1. порция данни, получени от предаващото приложение.
 - Дори, ако това е само 1 байт.
- Стъпка 2
 - Акумулиране на данни в изходящия буфер докато:
 - Получателят изпрати обратно потвърждение (АСК), или
 - Се съберат достатъчно данни за запълване на:
 - Сегмент с максимален размер (MSS), или
 - $-\frac{1/2}{2}$ от прозореца на подателя
 - Изпращане на сегмент
- Стъпка 3
 - Повторяне на стъпка 2 до края на комуникацията

36

Синдром, създаден от получателя:

Алгоритм на Clark (прилаган от получателя)

- Изпращане на потвърждение, веднага след пристигане на сегмент.
 - За да се подпомогне точното изчисление/актуализиране на RTT от страна на подателя
- Но се анонсира нулев размер на прозореца, докато във входящия буфер не се освободи достатъчно място, равно на:
 - <u>1 MSS</u>, или
 - $\frac{1}{2}$ от буфера

available buffer space \geq MIN $\left(\frac{\text{buffer size}}{2}, \text{maximum segment size}\right)$

Синдром, създаден от получателя: Забавено потвърждение (прилагано от получателя)

- Забавя се изпращането на потвърждение
- Докато не се освободи достатъчно място във входящия буфер
- Предимство
 - Намален трафик
 - Получателят не е длъжен да потвърждава всеки сегмент (особено, ако няма данни за предаване в обратната
- Недостатък
 - Забавено потвърждение може да принуди подателя да изпрати отново някой сегмент
 - Затова потвържденията се бавят не повече от 500 ms

38

37

контрол на грешките

ТСР: Контрол на грешките

Базиран на <u>PAR</u> схема

Положителна квитанция с повторно предаване Positive Acknowledgement with Retransmission

Всеки ТСР сегмент съдържа контролна сума

в заглавната си част

За защита от грешки на целия ТСР сегмент

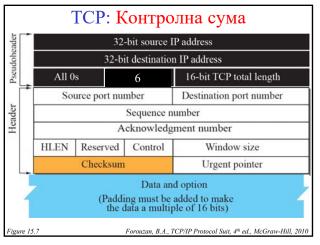
(вкл. псевдозаглавната част)

Получател

- Ако сегментът не е повреден, го потвърждава с АСК.
 - Кумулативни потвърждения
 - Piggybacking
- Сегментите, използвани само за потвърждение (т.е. без данни), <u>не</u> консумират поредни номера и <u>не</u> се потвърждават.
- Ако сегментът е <u>повреден</u>, го отхвърля <u>без</u> изпращане на отрицателна квитанция (NAK).
- Ако сегментът пристигне <u>непоред</u>, отлага АСК, докато не пристигнат всички липсващи сегменти, запълващи
- празнината. Подател
 - Ако таймерът се нулира преди пристигането на АСК, предава повторно сегмента.

 Тъй като или сегмента.
 - Тъй като или сегментът е повреден или изгубен, или потвърждението му се изгубило

b. Иван



ТСР: Piggybacking

ACK=53; Seq. No.=127

• АСК=53 специфицира номера на следващия байт данни, който А очаква да получи от В.

• Seq. No.=127 специфицира номера на първия байт данни, съдържащи се в сегмента, изпращан от А (т.е. номера на сегмента).

• В отоговор В изпраща своя байт 53 и последващите го байтове данни, капсулирани в сегмент.

• В същото време В потвърждава успешното приемане на байтове, изпратени от A (т.е. байт 130 е следващият байт, очакван от В).

41

ТСР: Нормално функциониране (piggybacking Rule 1 ACK-delaying Rule 1 Seq: 4001–5000 Ack: 1401 (delayed 500 m Rule 1 Ack: 1401 Rule 1 6001-7000 < 500 n (ACK any other segment)Time Forouzan, B.A., TCP/IP Protocol Suit, 4th ed., McGraw-Hill, 2010 re 15.29

TCP: Hobpeden cerment

Segment 1

Seq: 1201, 200 bytes
Segment 2

Segment 3

Seq: 1601, 200 bytes

Segment 3

Seq: 1601, 200 bytes

Segment 3

Seq: 1601, 200 bytes

Segment 3

Segment 3

Segment 3

Seq: 1601, 200 bytes

Segment 3

Segment 4

Segment 3

Segment 3

Segment 3

Segment 4

Segment 3

Segment 3

Segment 3

Segment 3

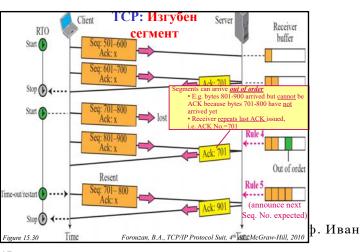
Segment 3

Segment 3

Segment 4

Segment 3

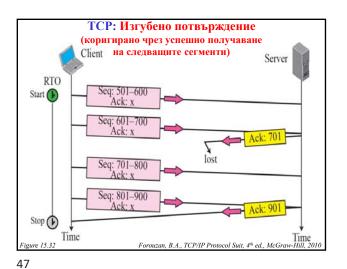
43

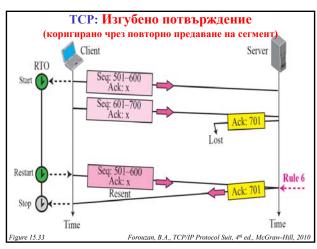


ТСР изгубен сегмент: Бързо повторно Client предаване RTO time tart (1) 3 дублирани повторни потвържtop (P) дения Seq: 301-400 Ack: x (ACK=300) се възприемат като сигнал, че q: 501–600 Ack: x получатепят все още очаква Seq: 601-700 съответния сегмент гаймерът за Fast retransa сегмент все O още не се е нулирал. Сегментът се предава повторно веднага! Figure 15.31 Forouzan, B.A., TCP/IP Protocol Suit, 4th ed., McGraw-Hill, 2010

44

42





ТСР контрол на грешките: Проблеми

- Сегментите могат да се забавят при транзита
 - Може да се наложи повторното им предаване
 - Но с различен диапазон от байтове
 в сравнение с първоначалното им предаване
- Изисква се внимателно администриране
 - За да се следи кои байтове са били правилно доставени досега

ТСР контрол на задръстванията

49 61

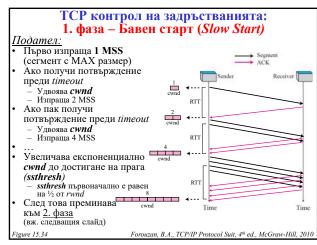
Динамично манипулиране на **ТСР:** Контрол на размера на прозореца Подателят поддържа 2 прозореца Прозорец, зададен от получателя (receiver window, rwnd) сегver window, rwnd)
Ако подателят се придържа
към него, няма да възникнат
проблеми поради препълване
на буфер, но могат да
възникнат проблеми поради
задръствания в мрежата. Прозорец, предизвикан от задръствания (congestion window, cwnd) Във всеки един момент подателят използва помалкия от тези 2 прозореца T.e. swnd = MIN (rwnd, cwnd)Основно действие Когато таймерът се занули, подателят започва да предава по-малко (по обем) данни. Намалява натоварването на мрежата (a) Fast network feeding low-capacity receiver (b) Slow network feeding high-capacity receiver b. Иван С пел облекчаване на 6-22 задръстванията*Tanenbaum* A.S. & Wetherall D.J., Computer Networks, 5th ed., Pearson, 201

62

TCP: Контрол на задръстванията (прод.)

- Основно допускане от страна на ТСР
 - Изтичане на времето за изчакване (timeout) се причинява от задръствания, а не поради грешки в IP пакетите.
- Вярно в кабелните мрежи
 - Комуникационните линии в днешно време са много надеждни (напр. при използване на влакнесто-оптични кабели)
 - Много рядко пакетите пристигат с грешки
- Погрешно в безжичните мрежи
 - Изключително ненадеждни
 - Пакети се губят или пристигат с грешки винаги, по всяко време.
 - Друг подход е необходим за контрол на задръстванията





65

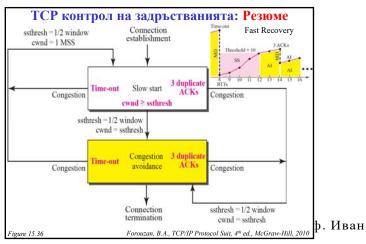


TCP контрол на задръстванията:
3. фаза – Рязко свиване на прозореца
(Multiplicative Decrease)

- Агресивно
 - Защото на мрежата ѝ е трудно да се възстанови от претоварването
- Подател:
 - Ако възникне *timeout*
 - Намалява ssthresh до ½ от cwnd
 - Връща се към 1. фаза (*Slow Start*)
 - Т.е. възвръща първоначалната стойност на *cwnd* (=1 MSS)

66

67



TCP: Oпции

End of option list

No operation

Maximum segment size

Window scale factor

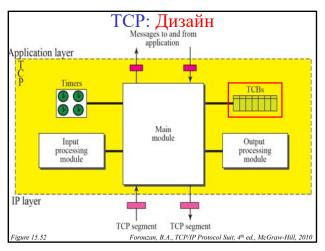
Multiple-byte

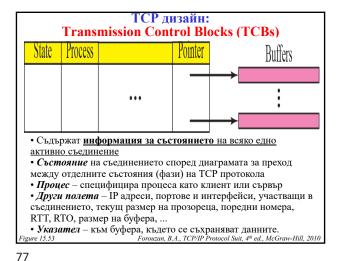
Timestamp

SACK-permitted

SACK

Forouzan, B.A., TCP/IP Protocol Suit, 4th ed., McGraw-Hill, 2010





ТСР: Реализации RFC 1122 TCP Tahoe Measure TCP Reno NewReno RTT Variance Estimation Exponential RTO Backoff Karn's Algorithm Slow Start Dynamic Window Sizing 1 1 1 on Congestion Fast Retransmit Fast Recovery Modified Fast Recovery

- Бързо възстановяване (fast recovery)
- В случай на бързо повторно предаване, прозорецът се намалява наполовина и директно се преминава към 2. фаза (плавно нарастване на прозореца)
 Модифицирано бързо възстановяване
 Подобрена реакция при загуба на 2 сегмента от един прозорец ble 22.5 Stallings, W., Data and Computer Communications, 9th ed., Prentice Hall, 2011

79

Stream Control Transmission Protocol (SCTP) Application layer SMTP FTP H.248 H.323 DHCP Transport layer SCTP TCP UDP IGMP | ICMP IP ARP Data link laver Underlying LAN or WAN technology Forouzan, B.A., TCP/IP Protocol Suit, 4th ed., McGraw-Hill, 2010 igure 16.1

84

SCTP

- Дефиниран в RFC 2960
- Първоначално стандартизиран от IETF SIGTRAN WG за транспортиране на SS7 сигнализация през IP мрежи
- По-късно еволюира до транспортен протокол с общо предназначение, осигуряващ надеждно, пълно-дуплексно съединение с подобрени опции за доставка.
 - Необходим за нови приложения като IP телефония (VoIP), ISDN over IP, media gateway control, ..

Protocol	Port Number	Description
IUA	9990	ISDN over IP
M2UA	2904	SS7 telephony signaling
M3UA	2905	SS7 telephony signaling
H.248	2945	Media gateway control
H.323	1718, 1719, 1720, 11720	IP telephony
SIP	5060	IP telephony
Table 16 1	Forouzan, B.A., TCF	P/IP Protocol Suit. 4th ed., McGraw-Hill, 2010

b. Иван

SCTP: Прилики с TCP

- Логическо съединение
 - <u>Асоциация</u>
- Пълен дуплекс
- Надежден транспорт
- Поредни номера за контрол на потока и на грешките
 - Същия диапазон като TCP, т.е. (0, 2³²-1).
- · Piggybacking
- Контролна сума
 - **32** бита (SCTP) / **16** бита(TCP)
- Бързо повторно предаване в случай на:
 - 4 SACKs (SCTP) / 3 ACKs (TCP)
- Бързо възстановяване
- Контрол на задръстванията (същите фази като ТСР)
 - Но прилагани поотделно за всеки поток (stream)

SCTP: Разлики с TCP

- Ориентиран към съобщения, а не към байтове. (запазва границите на съобщенията на приложния слой)
- Номерира парчета от данни (chunks)
 - ТСР номерира байтове
- Последователни номера се използват само за потвърждение на данни (data chunks)
 - Control chunks се потвърждават с други control chunks
- Потвърждава последния получен номер, не следващия очакван
- Т.е. ACK_X означава 'Изпрати ми X+1'
- Няма опции в заглавната част
 - Опциите се задават чрез дефиниране на нови *chunk* видове
- Поддръжка на множество потоци от данни (multi-streaming) ако един поток е блокиран, останалите продължават да доставят данни, аналогично на магистралните платна.
 - ТСР използва само 1 поток от данни

Multi-homing (хостовете могат да дефинират <u>няколко IP адреса / мрежови интерфейси</u> във всеки край на комуникацията за постигане на отказоустойчивост; ако един от пътищата за доставка се провали, започва използването на друг)

TCP използва само по 1 IP адрес във всеки край на комуникацията 87

 До голяма степен е независим от другите потоци Може да е: Строго подреден и 	SCTP: Multi-streaming ending Receiver
пабежден - Надежден — данните се доставят към приложението, веднага след пристигането им. - Частично падежден — приложението определя време на живот за данните; данните не се предават повторно. - Неподреден — подобно на UDP; данните са независими по отношение на другите данни, изпратени по-рано; затова се запазват в буфер веднага при	Stream of data chunks STP Stream of data chunks SCTP Stream of data chunks SCTP SCTP Stream of data chunks

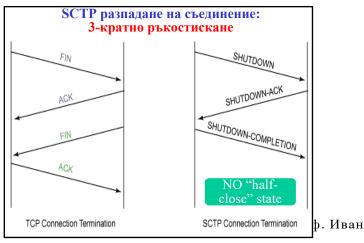
88

92

Feature	UDP	TCP	SCTP
Connection oriented	No	Yes	Yes
Reliable transport	No	Yes	Yes
Unreliable transport	Yes	No	Yes
Preserve message boundaries	Yes	No	Yes
Ordered delivery	No	Yes	Yes
Unordered delivery	Yes	No	Yes
Data checksum	Yes	Yes	Yes
Checksum size (bits)	16	16	32
Path MTU	No	Yes	Yes
Flow-, error-, congestion control	No	Yes	Yes
Multiple streams	No	No	Yes
Multi-homing support	No	No	Yes

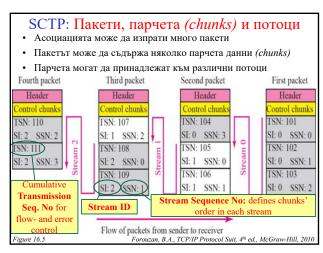


89



91

SCTP: Пакет Source port address Destination port address Source port address Destination port address Sequence number Verification tag each association) Acknowledgment number Checksum Window size Control flag Control chunks ACK No. and window size Urgent pointer Checksum Options • carry data chunks Data bytes • each chunk may belong to different stream A segment in TCP A packet in SCTP Forouzan, B.A., TCP/IP Protocol Suit, 4th ed., McGraw-Hill, 2010



SCTP: Multi-homing Хостове Могат да използват <u>няколко IP адреса</u> Обвързани едновременно с една асоциация SCTP асоциация Може да използва множество налични ІР адреси Setup (INIT chunk) – всяка крайна точка посочва списък с налични IP Setup (INIT-ACK chunk) – крайните точки избират двойка валидни IP адреси като <u>основни (primary)</u>, между които се осъществява <u>предаването</u> на данните по подразбиране (всички други адреси се маркират като вторични / secondary) Път за предаване Наблюдаван чрез изпращане на <u>HEARTBEAT</u> chunks, за изпробване достижимостта на даден адрес. Счита се за неактивен, ако предаването по него се провали многократно (данните се предават повторно към избран вторичен адрес). Приложението се информира за състоянието на пътя (в случай на промени)

Приложението може да инструктира SCTP да използва друг път за основен