

Sieci komputerowe

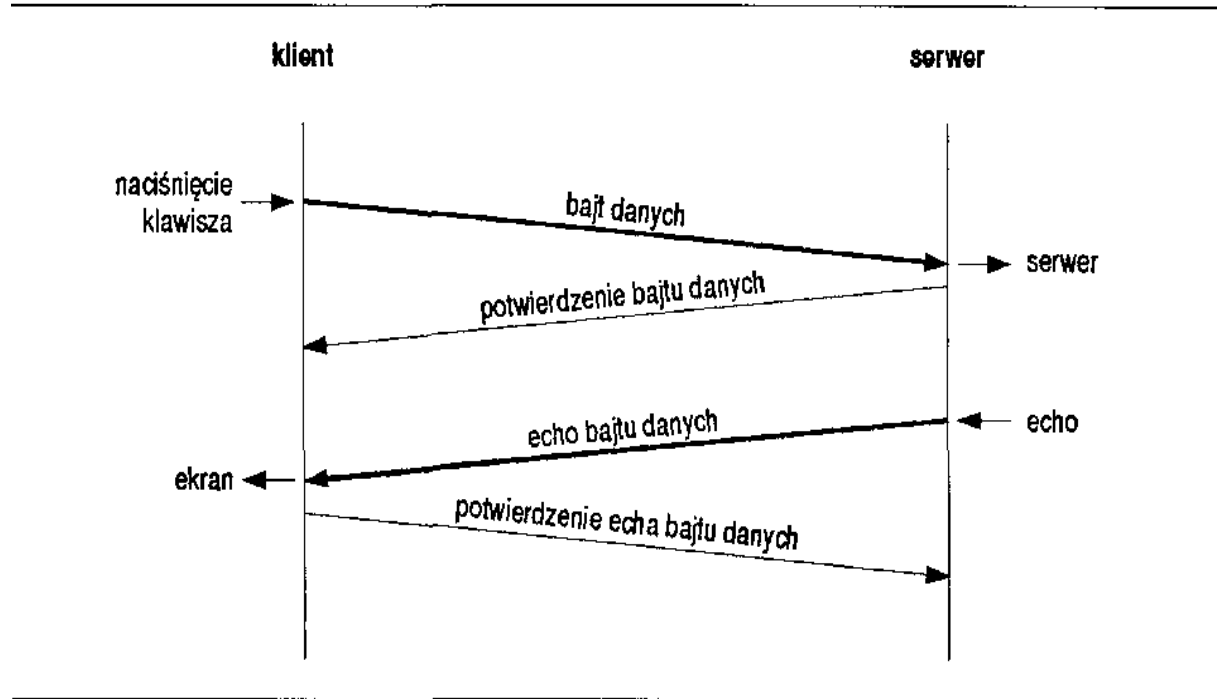
Wykład 8
Protokół TCP c.d.

Slajdy

- Slajdy i kilka przykładowych zadań egzaminacyjnych:

<http://cia.mimuw.edu.pl/sieci/>

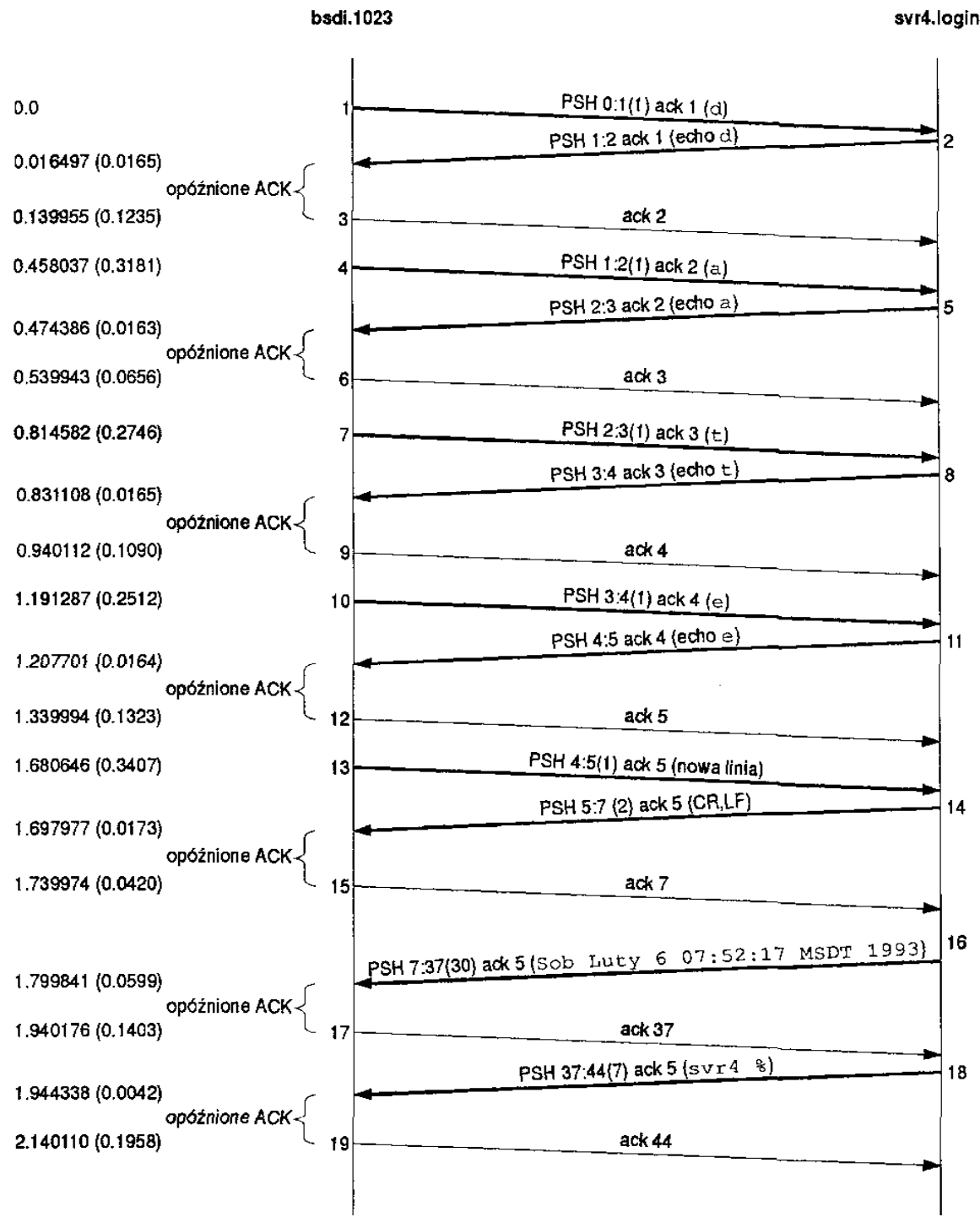
Przepływ danych interaktywnych



rysunek 19.1 Jeden z możliwych sposobów wymiany echa wpisywanych znaków

- Przesyłane są pojedyncze znaki, pakiet IP dla przesłania 1 znaku ma długość 41 bajtów!
- Dla jednego przesłanego znaku tworzone są zwykle 4 segmenty

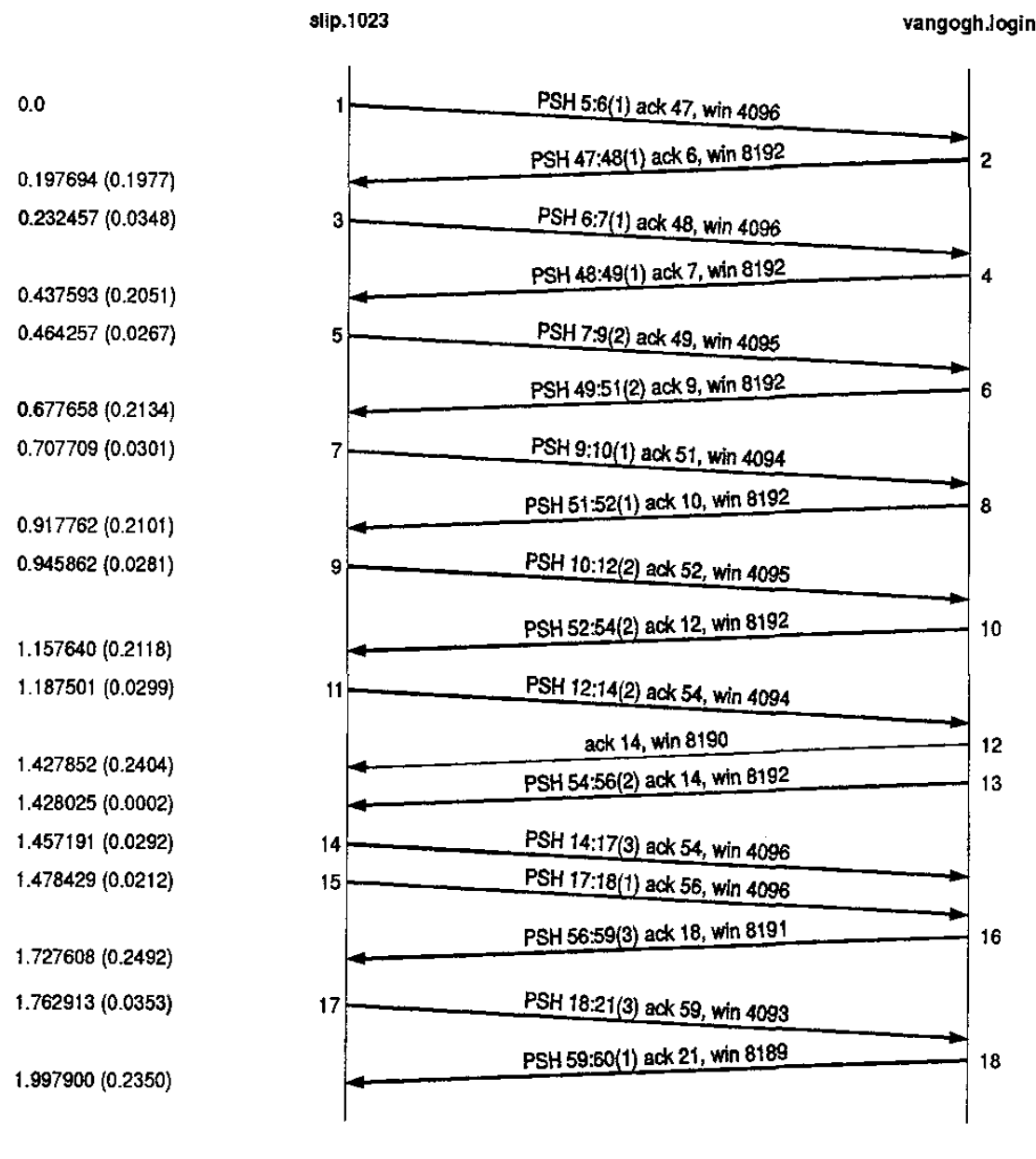
Opóźnione potwierdzenia w przesyłaniu interaktywnym



Algorytm Nagle'a

- Służy do minimalizacji liczby wysyłanych segmentów
 - Algorytm Nagle'a mówi, że TCP może mieć tylko jeden mały niepotwierdzony segment
 - Może więc występować wstrzymanie wysyłania danych, tak, aby później wysłać więcej w jednym segmencie
 - Istnieje możliwość wyłączenia algorytmu Nagle'a, za pomocą odpowiedniej opcji API gniazd

Algorytm Nagle'a c.d.

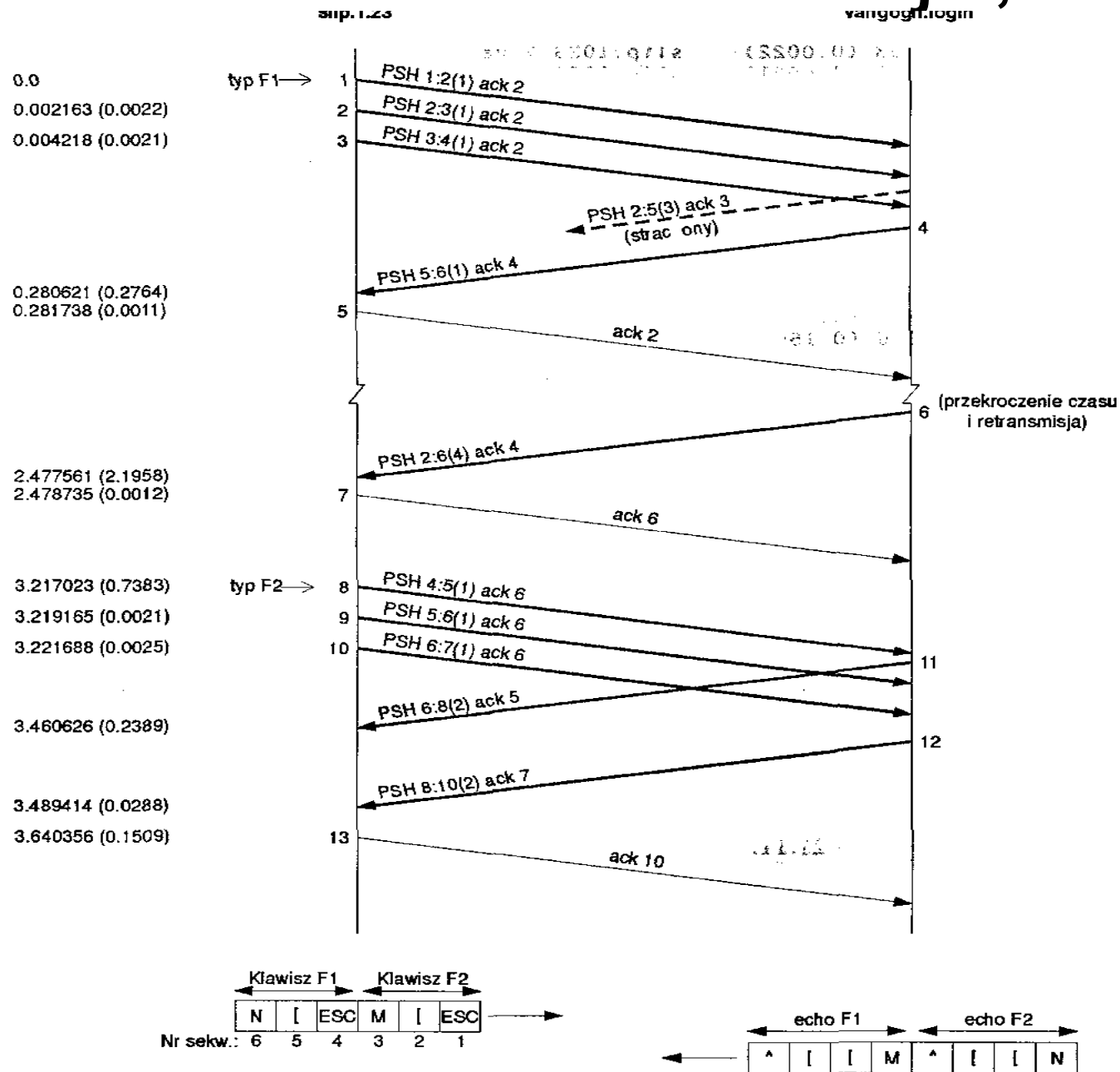


Rysunek 19.4 Przepływ danych pomiędzy *slip* a *vangogh.cs.berkeley.edu* przy połączeniu *rlogin*

Retransmisja

- Retransmisja następuje w momencie, gdy TCP nie otrzymał potwierdzenia dla któregoś z segmentów
- Nie ma mechanizmu potwierdzeń selektywnych!
- Konieczne jest wyznaczenie czasu oczekiwania, po którym ma nastąpić retransmisja
- Aby wyznaczyć czas oczekiwania, TCP musi zmierzyć czas podróży segmentów (tzw. RTT-round trip time)
 - Czas podróży: $R = \alpha R_p + (1 - \alpha)M$, $\alpha = 0.9$
 - Czas oczekiwania: $RTO = R\beta$, $\beta = 2$

Retransmisja, c.d.



Retransmisja, c.d.

- W momencie wykrycia segmentu innego niż oczekiwany, jest wysyłane tzw. zduplikowane ACK natychmiast (bez opóźnienia)
 - Może to pomóc ustalić, że segment zaginął i dokonać retransmisji jeszcze przed upłynięciem czasu oczekiwania (algorytm szybkiej retransmisji)

Repakietyzacja

- Po zakończeniu odliczania czasu oczekiwania następuje retransmisja, ale TCP nie musi retransmitować identycznych segmentów
- Może wykonać repakietyzację, czyli wysłać segment większy

Algorytm powolnego startu

- Służy do kontroli przepływu związanego z obciążeniem sieci (inaczej niż w przypadku ogłaszania wielkości okna)
 - Problem jest wykrywany, gdy pojawiają się straty segmentów
- Nie następuje transmisja mająca na celu wypełnienie okna ogłaszanego przez odbiorcę
- Liczba wysyłanych segmentów bez potwierdzenia zaczyna się od 1, i jest stopniowo zwiększana wraz z otrzymywaniem kolejnych ACK

Powolny start c.d.

- Algorytm powolnego startu wymaga wprowadzenia i obliczania dla każdego połączenia zmiennej *cwnd* (ang. congestion window) – okna przeciążenia
 - wartość *cwnd* jest ustawiona na początku na jeden segment i zwiększana o wielkość jednego segmentu za każdym otrzymanym ACK
 - wysyłana liczba segmentów nie może przekroczyć wartości *cwnd* i wartości okna ogłaszanego przez odbiorcę
 - za każdym otrzymanym ACK segmentu, wielkość *cwnd* wzrasta o 1 segment, powoduje to wykładniczy wzrost liczby wysyłanych segmentów (1,2,4,8...)

Algorytm zapobiegania zatorom

- Wprowadza zmienną $ssthresh$ (próg powolnego stratu)
- Gdy wystąpi zator (wykryty poprzez przekroczenie czasu) $ssthresh := 1/2 * cwnd$; $cwnd := 1$ (powolny start)
- Gdy $cwnd \geq ssthresh$, zaczyna działać algorytm zapobiegania zatorom:
 - $cwnd += segsize * segsize / cwnd$
 - powoduje to wolniejszy (linowy) wzrost liczby wysyłanych segmentów

Algorytm zapobiegania zatorom

- Jeśli zator zostanie wykryty poprzez otrzymywanie zduplikowanych ACK, nie jest wykonywany powolny start (nie zmniejszamy rozmiaru cwnd do jednego segmentu), od razu zaczyna działać algorytm zapobiegania zatorom
 - cwnd jest nadal zwiększane, ale wg algorytmu zapobiegania zatorom: $cwnd += \text{segsize} * \text{segsize} / cwnd$

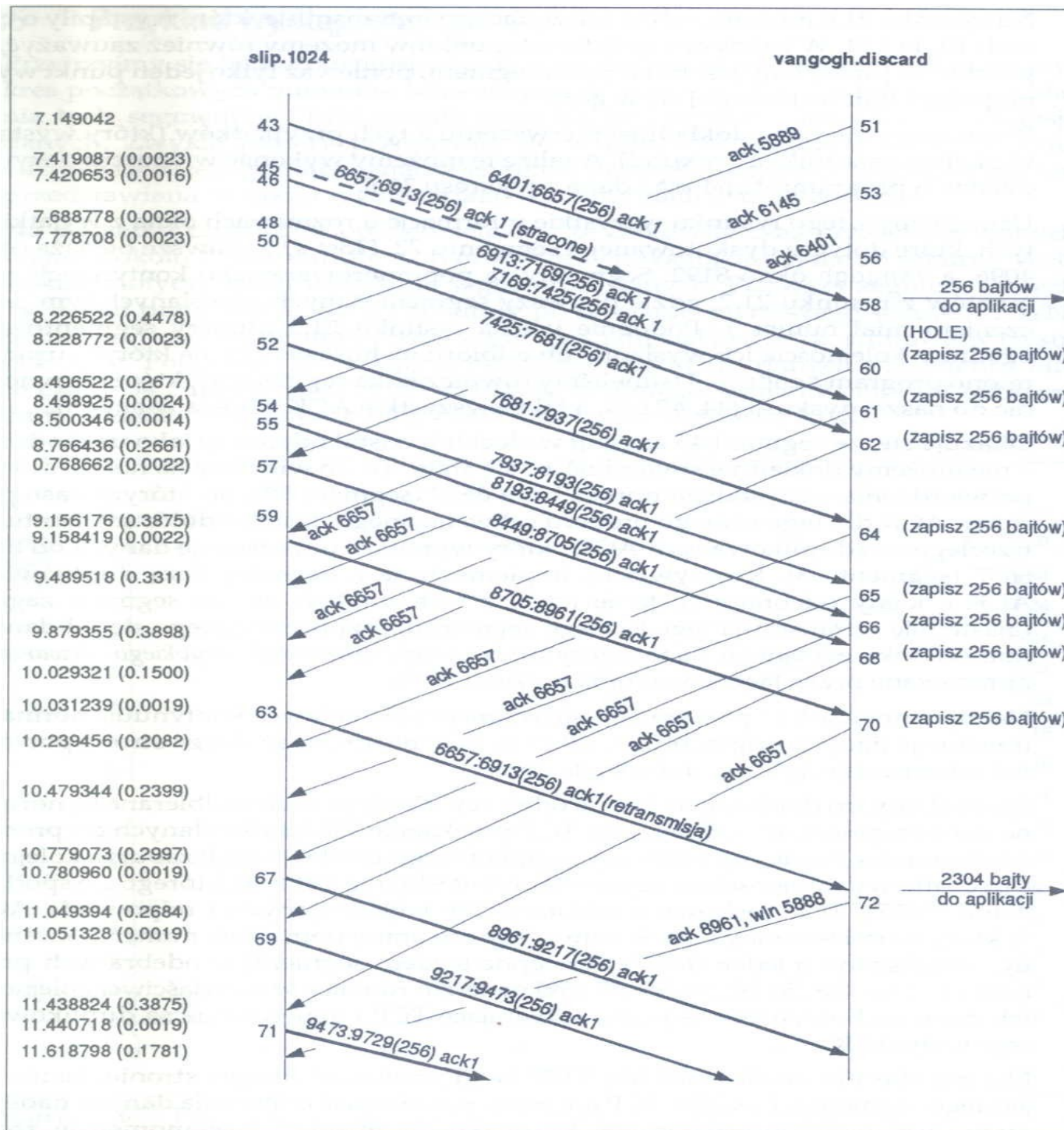
Algorytm zapobiegania zatorom c.d.

- Algorytm powolnego startu i zapobiegania zatorom służą do kontroli przepływu ze strony nadawcy

Algorytm szybkiej retransmisji

- Jeśli wystąpi 3 lub więcej zduplikowanych ACK można wysłać segment ponownie, przed upłynięciem czasu oczekiwania
- Nie trzeba stosować powolnego startu, gdyż pojawianie się zduplikowanych ACK oznacza, iż segmenty są odbierane, więc nie wystąpił zator
- Jeśli otrzymamy mniej niż 3 zduplikowane ACK, to znaczy że wystąpiło tylko zakłócenie porządku segmentów i nie należy stosować szybkiej retransmisji

Szybka retransmisja - przykład

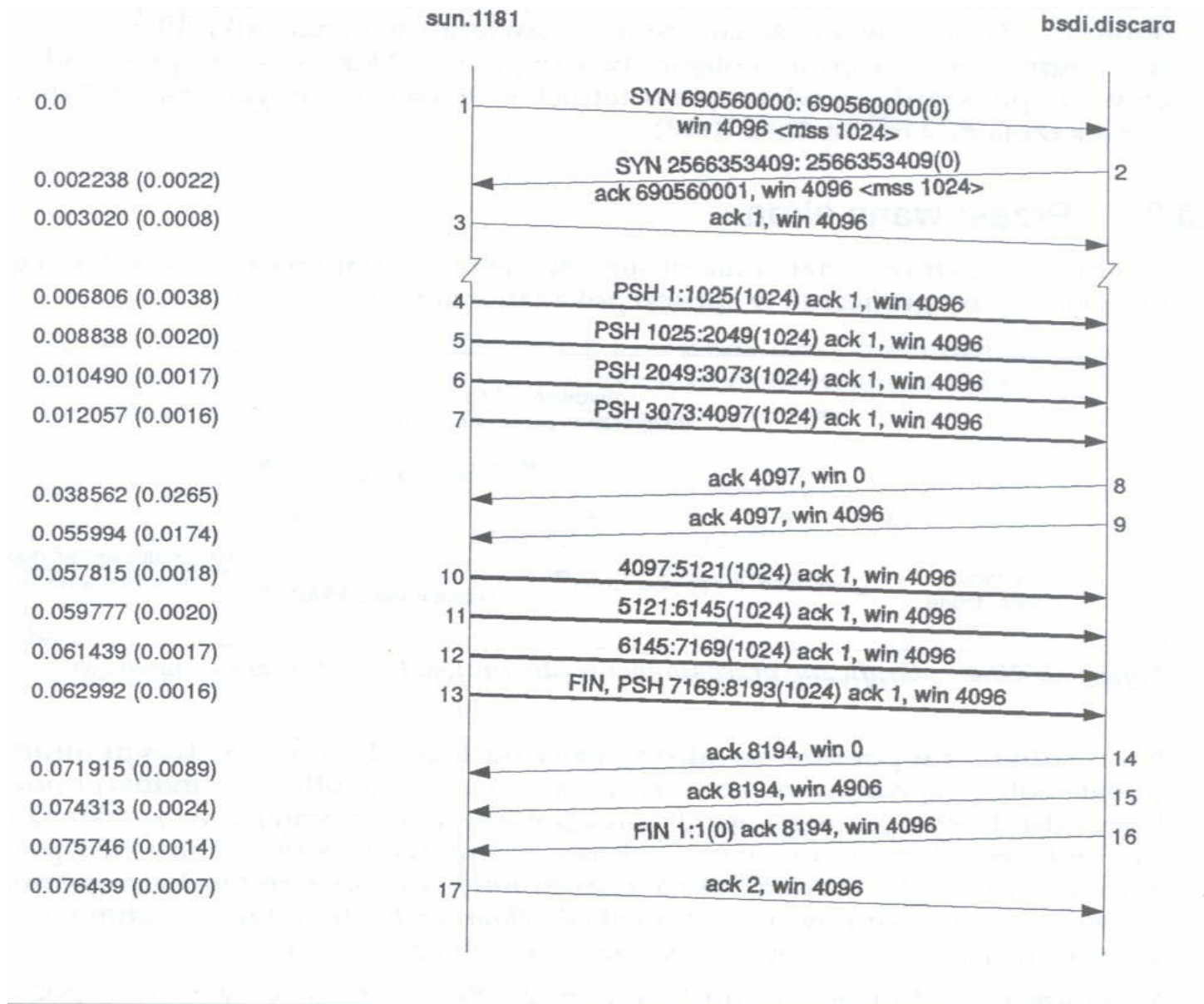


Rysunek 21.7 Wymiana pakietów retransmisji odbywającej się w okolicy 10 sekundy

Mechanizm persist TCP

- Może wystąpić problem, gdy zaginie segment ogłaszający, iż okno powiększyło się (przy poprzednio ogłoszonej zerowej wielkości okna)
 - W tym przypadku transmisja nie mogłaby być kontynuowana, bo nadawca czekałby w nieskończoność
 - Aby temu zapobiec, nadawca wysyła segmenty sondujące o długości jednego bajta (TCP może to robić mimo ogłoszonej zerowej długości okna)

Persist TCP c.d.



ysunek 20.3 Przesyłanie 8192 bajtów z szybkiego nadawcy do wolnego odbiorcy

Łączy o dużych przepustowościach i opóźnieniach

- Pojemność potoku można zdefiniować jako:
 - $\text{pojemność}(\text{bity}) = \text{szerokość pasma (b/s)} * \text{czas podróży(s)}$
- Potok powinien być wypełniony, aby uzyskać oczekiwaną przepustowość, okno określone liczbą 16 bitową nie wystarcza
- Stosuje się opcję skalowania okna

Opcja skalowania okna

- Dla sieci o dużych przepustowościach i dużych opóźnieniach okno opisane liczbą 16 bitową może być zbyt małe
- Stosowana jest opcja TCP skalowania okna
 - stosowany jest licznik jedno bajtowy, może przyjmować wartości od 0 do 14.
 - Największy rozmiar okna to $65535 * 2^{14}$
- Opcja skalowania może się pojawiać jedynie w segmentach SYN, więc jest stała dla danego połączenia
 - Jeśli strona, która wykonuje aktywne otwarcie umieści tę opcję w segmencie SYN, ale nie dostanie jest w segmencie SYN przesłanym przez drugą stronę, to opcja ta nie może być używana

Opcja znaczników czasowych

- Umożliwia umieszczanie 32 bitowych znaczników czasowych w wysyłanych segmentach
- Odbiorca wysyła echo znacznika
- Umożliwia to dokładny pomiar RTT
- Umożliwia również ochronę przed powtarzającymi się numerami sekwencyjnymi
 - Zjawisko to może wystąpić w sieciach o przepustowościach $\geq 1\text{Gb/s}$
- Opcja ta może być używana jedynie w segmentach SYN

Opcje używane we współczesnych implementacjach

- 12:15:18.513011 IP cia.mimuw.edu.pl.59542 > ftp-osl.osuosl.org.www: S
2063872163:2063872163(0) win 5840 <mss 1460,sackOK,timestamp 18560100930,nop,wscale 6>
- 12:15:18.513233 IP ftp-osl.osuosl.org.www > cia.mimuw.edu.pl.59542: S
58996676:58996676(0) ack 2063872164 win 5792 <mss 1460,sackOK,timestamp 880023258 1856010093,nop,wscale 7>

Przesyłanie pliku (tcpdump)

```
12:15:19.964869 IP cia.mimuw.edu.pl.59542 > ftp-osl.osuosl.org.www: . ack 98584 win 1002
<nop,nop,timestamp 1856010456 880023693>
12:15:19.964929 IP ftp-osl.osuosl.org.www > cia.mimuw.edu.pl.59542: P 98584:99784(1200) ack
167 win 54 <nop,nop,timestamp 880023693 1856010405>
12:15:19.964935 IP ftp-osl.osuosl.org.www > cia.mimuw.edu.pl.59542: P 99784:100032(248) ack
167 win 54 <nop,nop,timestamp 880023693 1856010405>
12:15:19.964962 IP cia.mimuw.edu.pl.59542 > ftp-osl.osuosl.org.www: . ack 100032 win 1002
<nop,nop,timestamp 1856010456 880023693>
12:15:19.965194 IP ftp-osl.osuosl.org.www > cia.mimuw.edu.pl.59542: . 100032:101480(1448)
ack 167 win 54 <nop,nop,timestamp 880023693 1856010456>
12:15:19.965201 IP ftp-osl.osuosl.org.www > cia.mimuw.edu.pl.59542: P 101480:102928(1448)
ack 167 win 54 <nop,nop,timestamp 880023693 1856010456>
12:15:19.965205 IP ftp-osl.osuosl.org.www > cia.mimuw.edu.pl.59542: . 102928:104376(1448)
ack 167 win 54 <nop,nop,timestamp 880023693 1856010456>
12:15:19.965208 IP ftp-osl.osuosl.org.www > cia.mimuw.edu.pl.59542: . 104376:105824(1448)
ack 167 win 54 <nop,nop,timestamp 880023693 1856010456>
12:15:19.965210 IP ftp-osl.osuosl.org.www > cia.mimuw.edu.pl.59542: P 105824:107024(1200)
ack 167 win 54 <nop,nop,timestamp 880023693 1856010456>
12:15:19.965263 IP ftp-osl.osuosl.org.www > cia.mimuw.edu.pl.59542: . 107024:108472(1448)
ack 167 win 54 <nop,nop,timestamp 880023693 1856010456>
12:15:19.965266 IP ftp-osl.osuosl.org.www > cia.mimuw.edu.pl.59542: . 108472:109920(1448)
ack 167 win 54 <nop,nop,timestamp 880023693 1856010456>
12:15:19.965268 IP ftp-osl.osuosl.org.www > cia.mimuw.edu.pl.59542: P 109920:111120(1200)
ack 167 win 54 <nop,nop,timestamp 880023693 1856010456>
12:15:19.965298 IP ftp-osl.osuosl.org.www > cia.mimuw.edu.pl.59542: . 111120:112568(1448)
ack 167 win 54 <nop,nop,timestamp 880023693 1856010456>
12:15:19.965329 IP cia.mimuw.edu.pl.59542 > ftp-osl.osuosl.org.www: . ack 112568 win 1002
<nop,nop,timestamp 1856010456 880023693>
```