

情報ネットワーク論I

TCP 1/2

門林 雄基
(インターネット工学講座)
TA: Omar, 衛藤、鈴木、横山

トランスポート層の機能

- プロセス間の通信
 - プロセスの指定
 - プロセス間の通信路の識別
- 上位層へのインターフェース
 - Connection-oriented (virtual circuit)
 - Connectionless (datagram)
- ネットワーク資源の競合と調停
 - フロー制御 (flow control)
 - 輻輳制御 (congestion control)

Copyright(C)2004 Youki Kadobayashi. All rights reserved.

Internetプロトコル群でのトランスポート

- TCP
 - Connection-oriented
- UDP
 - Connectionless
 - IP + プロセスの識別機能

Copyright(C)2004 Youki Kadobayashi. All rights reserved.

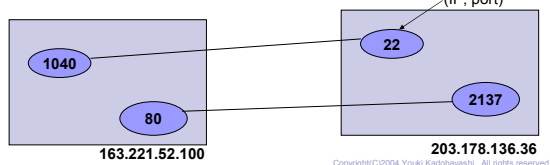
TCPにおけるプロセス、通信路の識別

■ プロセスの指定

- (IP, port)

■ TCP connection の識別

- (source IP, source port, destination IP, destination port)



TCPのサービス

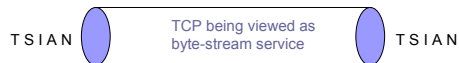
■ コネクション指向型 (connection-oriented)

■ 信頼性

- 配送順序の入れ替わり、重複、パケット廃棄、ビット誤り等を隠蔽

■ バイト列 (byte-stream service)

- パケット、パケットの境界は見えない



TCPの特徴(1)

■ ストリーム通信 (stream oriented)

- バイト列
- 境界がない
- アプリケーション側での構造化が必要

■ 効率化のための工夫

- Piggybacking
- Nagle algorithm
- ...

■ 全二重 (full duplex)

- 独立した2つのストリーム

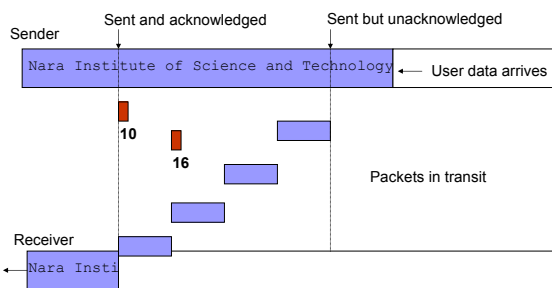
Copyright(C)2004 Youki Kadoyayashi. All rights reserved.

信頼性を有するストリームの実現

- 確認応答 (ACK: Acknowledgment)
 - Active acknowledgment
 - 明示的に受信の確認応答を返す
 - Duplicate ACK
 - パケットが落ちたことを伝える
- タイムアウトと再送
 - 一定時間を経過したらタイムアウトし、送信が正しく完了しなかったと仮定してパケットを再送する
 - 指数的バックオフ (Exponential back-off)

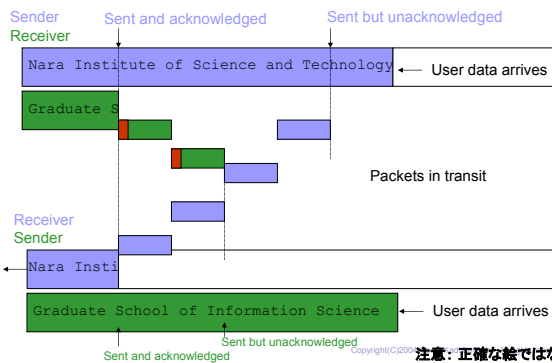
Copyright(C)2004 Youki Kadoyayashi. All rights reserved.

ACK



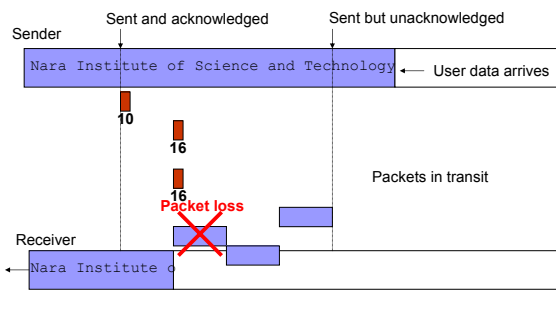
Copyright(C)2004 Youki Kadoyayashi. All rights reserved.

Piggybacking: 全二重通信時の効率化



Copyright(C)2004 Youki Kadoyayashi. All rights reserved. 注意: 正確な絵ではない

Duplicate ACK

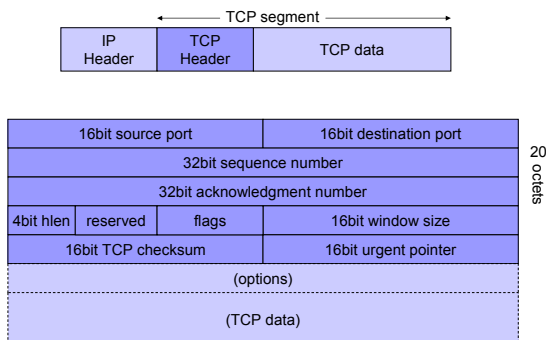


Copyright(C)2004 Youki Kadoyayashi. All rights reserved.

Questions?

Copyright(C)2004 Youki Kadoyayashi. All rights reserved.

TCPのヘッダ



Copyright(C)2004 Youki Kadoyayashi. All rights reserved.

Nagleのアルゴリズム

- Q. 1byte のデータに20byte+20byteのヘッダを付けるとオーバーヘッドが大きいのではないか？



- Nagle algorithm
 - ACKされていない小さいセグメントは1つしかネットワーク中に存在させない
 - RTTが小さい場合
 - LANなのでオーバーヘッドは許容できるはず
 - 少ないバッファリングでパケットを送出
 - RTTが大きい場合
 - WANなのでオーバーヘッドを小さく

Copyright(C)2004 Youki Kadobayashi. All rights reserved.

Q.

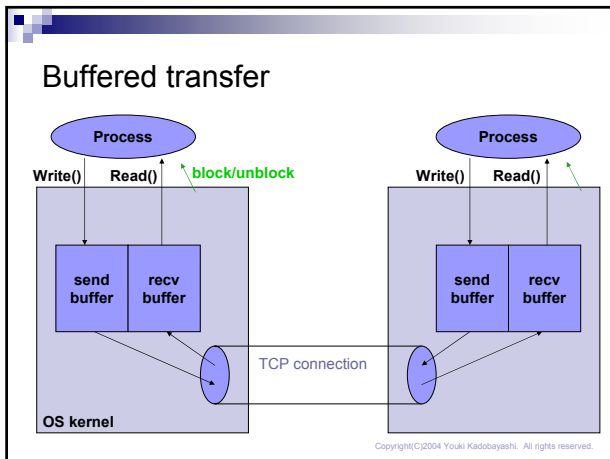
- Nagle algorithm をオフにする必要があるのはどういうときか？

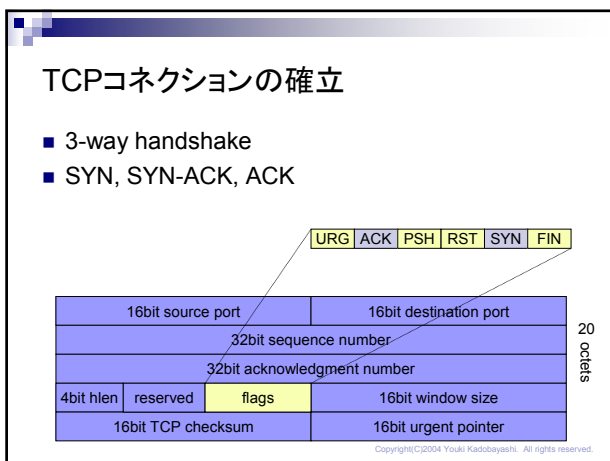
Copyright(C)2004 Youki Kadobayashi. All rights reserved.

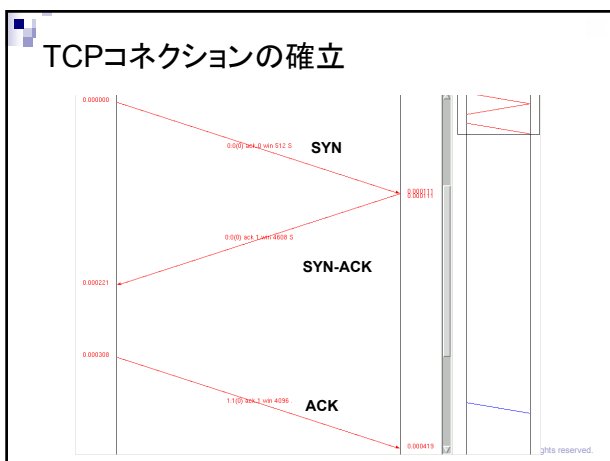
TCPの特徴(2)

- バッファつき転送 (buffered transfer)
 - いくらでも書き込める
 - アプリケーション側での同期は不要
 - オペレーティングシステムでプロセス状態を変更
- 仮想回線 (virtual circuit)
 - コネクションの確立と解放 (connection setup and release)
 - 切断を検出可能

Copyright(C)2004 Youki Kadobayashi. All rights reserved.







TCPコネクションの確立

- `dv# tcpdump tcp and host mint100.aist-nara.ac.jp`
- `tcpdump: listening on de0`
- 12:16:00.146101 `dv.aist-nara.ac.jp.49626 > mint100.aist-nara.ac.jp.telnet: S 686685713:686685713(0) win 65535 <mss 1460,nop,wscale 1,nop,nop,timestamp 110942140 0> (DF) [tos 0x10]`
- 12:16:00.146350 `mint100.aist-nara.ac.jp.telnet > dv.aist-nara.ac.jp.49626: S 2312441307:2312441307(0) ack 686685714 win 17520 <mss 1460> (DF)`
- 12:16:00.146405 `dv.aist-nara.ac.jp.49626 > mint100.aist-nara.ac.jp.telnet: . ack 1 win 65535 (DF) [tos 0x10]`

Sequence number + 1 を Ack として返すと、確認応答

Copyright(C)2004 Youki Kadobayashi. All rights reserved.

tcpdump出力の意味

- `time src.port > dst.port flag [ from:to(nbytes) | ack # ] win # opt`
- 12:16:00.146101 `dv.aist-nara.ac.jp.49626 > mint100.aist-nara.ac.jp.telnet: S 686685713:686685713(0) win 65535 <mss 1460,nop,wscale 1,nop,nop,timestamp 110942140 0> (DF) [tos 0x10]`
- 12:16:00.146350 `mint100.aist-nara.ac.jp.telnet > dv.aist-nara.ac.jp.49626: S 2312441307:2312441307(0) ack 686685714 win 17520 <mss 1460> (DF)`

32bit sequence number
32bit acknowledgment number
flags

Copyright(C)2004 Youki Kadobayashi. All rights reserved.

Questions?

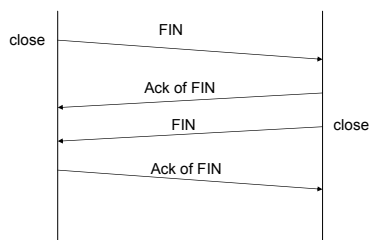
Copyright(C)2004 Youki Kadobayashi. All rights reserved.

TCPコネクションタイムアウト

- `dv# tcpdump tcp and host 163.221.1.252`
- `tcpdump: listening on de0`
- `13:50:58.896853 dv.aist-nara.ac.jp.49641 > 163.221.1.252.telnet: S 3631940833:3631940833(0) win 65535 <mss 1460,nop,wscale 1,nop,nop,timestamp 111511961 0> (DF) [tos 0x10]`
- `13:51:01.889346 dv.aist-nara.ac.jp.49641 > 163.221.1.252.telnet: S 3631940833:3631940833(0) win 65535 <mss 1460,nop,wscale 1,nop,nop,timestamp 111512261 0> (DF) [tos 0x10]`
- `13:51:04.889543 dv.aist-nara.ac.jp.49641 > 163.221.1.252.telnet: S 3631940833:3631940833(0) win 65535 <mss 1460,nop,wscale 1,nop,nop,timestamp 111512561 0> (DF) [tos 0x10]`
- `13:51:07.889836 dv.aist-nara.ac.jp.49641 > 163.221.1.252.telnet: S 3631940833:3631940833(0) win 65535 <mss 1460> (DF) [tos 0x10]`
- `13:51:10.890215 dv.aist-nara.ac.jp.49641 > 163.221.1.252.telnet: S 3631940833:3631940833(0) win 65535 <mss 1460> (DF) [tos 0x10]`
- `13:51:13.890411 dv.aist-nara.ac.jp.49641 > 163.221.1.252.telnet: S 3631940833:3631940833(0) win 65535 <mss 1460> (DF) [tos 0x10]`
- `^C`
- 39 packets received by filter
- 0 packets dropped by kernel

Copyright(C)2004 Youki Kadobayashi. All rights reserved.

TCPコネクションの解放



Copyright(C)2004 Youki Kadobayashi. All rights reserved.

TCPコネクションの解放

- `12:16:07.086568 mint100.aist-nara.ac.jp.telnet > dv.aist-nara.ac.jp.49626: FP 713:721(8) ack 245 win 17520 (DF) [tos 0x10]`
- `12:16:07.086738 dv.aist-nara.ac.jp.49626 > mint100.aist-nara.ac.jp.telnet: . ack 722 win 65535 (DF) [tos 0x10]`
- `12:16:07.086998 dv.aist-nara.ac.jp.49626 > mint100.aist-nara.ac.jp.telnet: F 245:245(0) ack 722 win 65535 (DF) [tos 0x10]`
- `12:16:07.087180 mint100.aist-nara.ac.jp.telnet > dv.aist-nara.ac.jp.49626: . ack 246 win 17519 (DF) [tos 0x10]`
- `^C`
- 355 packets received by filter
- 0 packets dropped by kernel

Copyright(C)2004 Youki Kadobayashi. All rights reserved.

TCPコネクションのリセット

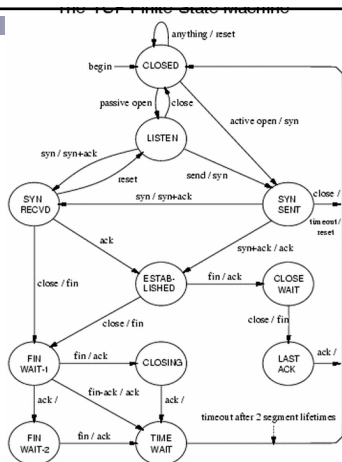
■ RST

- ☐ Abortive release (中断)
- ☐ Nonexistent port (該当ポートなし)

- 13:48:35.948096 dv.aist-nara.ac.jp.49635 > mint100.aist-nara.ac.jp.8080: S 1342403683:1342403683(0) win 65535 <mss 1460,nop,wscale 1,nop,nop,timestamp 111497668 0> (DF) [tos 0x10]
- 13:48:35.948265 mint100.aist-nara.ac.jp.8080 > dv.aist-nara.ac.jp.49635: R 0:0(0) ack 1342403684 win 0

Copyright(C)2004 Youki Kadobayashi. All rights reserved.

TCP FSM



TCPオプション

■ 3-way handshake

- ☐ 利用オプションについて通信相手と合意

- 12:16:00.146101 dv.aist-nara.ac.jp.49626 > mint100.aist-nara.ac.jp.telnet: S 686685713:686685713(0) win 65535 <mss 1460,nop,wscale 1,nop,nop,timestamp 110942140 0> (DF) [tos 0x10]
- ☐ MSS option
- ☐ Window scale option
- ☐ Timestamp option, etc.

Copyright(C)2004 Youki Kadobayashi. All rights reserved.

Questions?

Copyright(C)2004 Youki Kadobayashi. All rights reserved.

TCPを用いた対話型データ転送

- 12:16:00.149276 dv.aist-nara.ac.jp.49626 > mint100.aist-nara.ac.jp.telnet: P 1:37(36) ack 1 win 65535 (DF) [tos 0x10]
- 12:16:00.189184 mint100.aist-nara.ac.jp.telnet > dv.aist-nara.ac.jp.49626: P 1:16(15) ack 37 win 17484 (DF) [tos 0x10]
- 12:16:00.189491 dv.aist-nara.ac.jp.49626 > mint100.aist-nara.ac.jp.telnet: P 37:40(3) ack 16 win 65535 (DF) [tos 0x10]
- 12:16:00.189656 mint100.aist-nara.ac.jp.telnet > dv.aist-nara.ac.jp.49626: P 16:40(24) ack 40 win 17517 (DF) [tos 0x10]
- 12:16:00.189957 dv.aist-nara.ac.jp.49626 > mint100.aist-nara.ac.jp.telnet: P 40:108(68) ack 40 win 65535 (DF) [tos 0x10]
- 12:16:00.190124 mint100.aist-nara.ac.jp.telnet > dv.aist-nara.ac.jp.49626: P 40:64(24) ack 108 win 17452 (DF) [tos 0x10]

Copyright(C)2004 Youki Kadobayashi. All rights reserved.

まとめ

- トランスポート層
- Internetにおけるトランスポート – TCP
- TCPのサービス、特徴
- ACK, piggybacking
- Nagle algorithm
- コネクションの確立、解放
- TCP FSM
- 簡単なデータ転送 (telnet を例として)

Copyright(C)2004 Youki Kadobayashi. All rights reserved.
