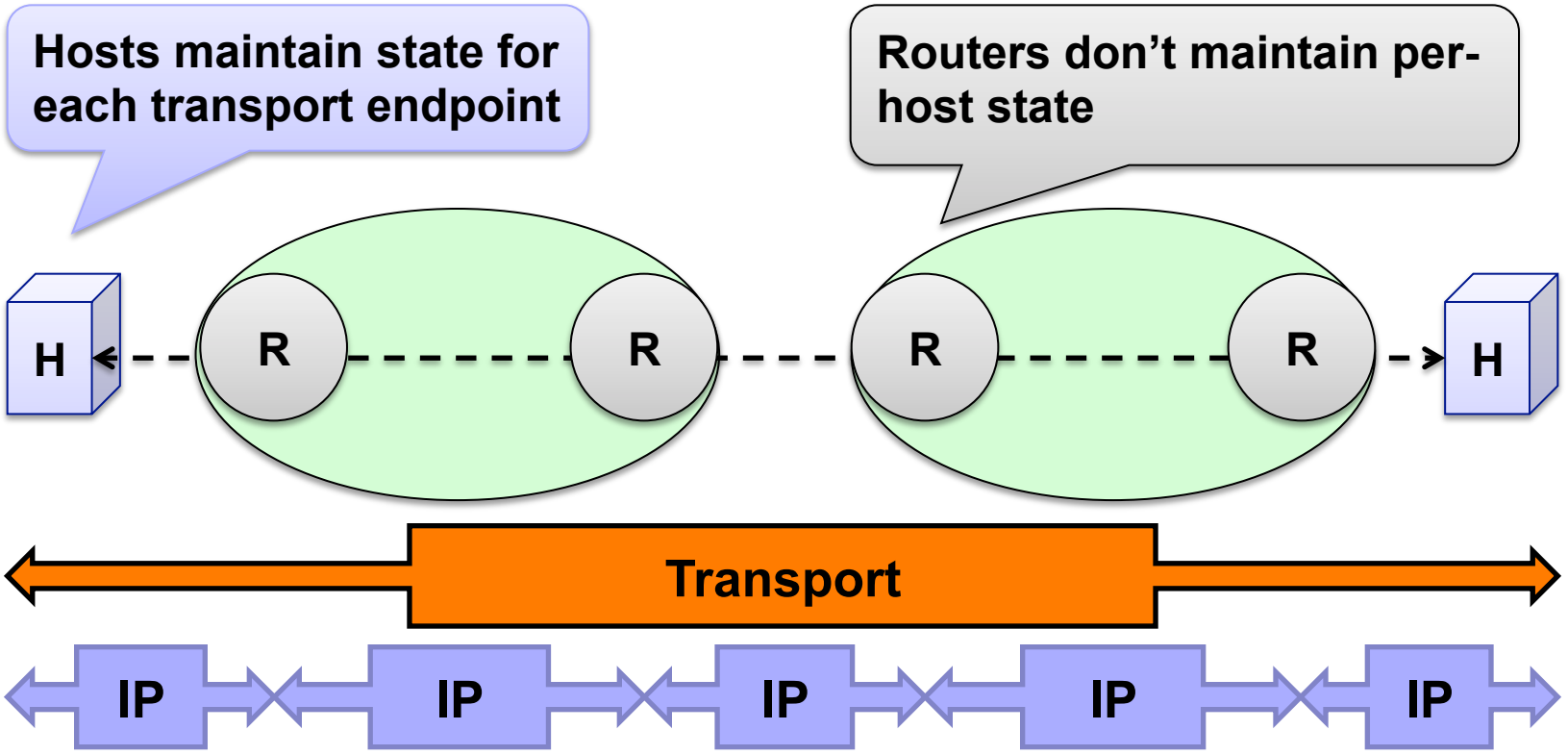
情報ネットワーク論I
TCP 1/2

門林　雄基
奈良先端科学技術大学院大学

Transport layer: a birds-eye view



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
  - コネクション指向 （仮想回線）
  - コネクションを伴わない （データグラム）
- 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)

次回講義

Internetプロトコル群でのトランスポート

- TCP　(RFC793)
  - “Transmission Control Protocol”
  - コネクション指向
  - 多機能
- SCTP (RFC4960)
- DCCP (RFC4340)

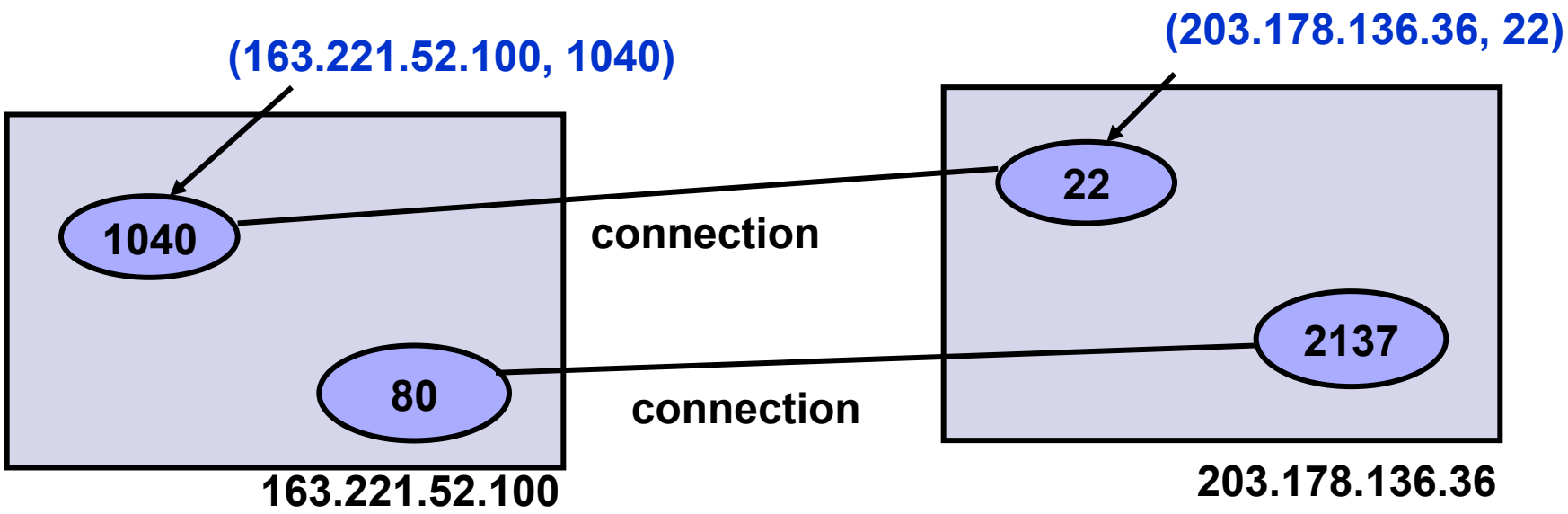
Advanced topic; out of scope

- UDP　(RFC768)
  - “User Datagram Protocol”
  - コネクションを伴わない
  - IP + プロセスの識別機能

独習

TCPにおけるプロセス、接続の識別

- プロセスの指定
  - (IP, port)
- TCP コネクションの識別
  - (source IP, source port, destination IP, destination port)



TCPのサービスモデル(1)

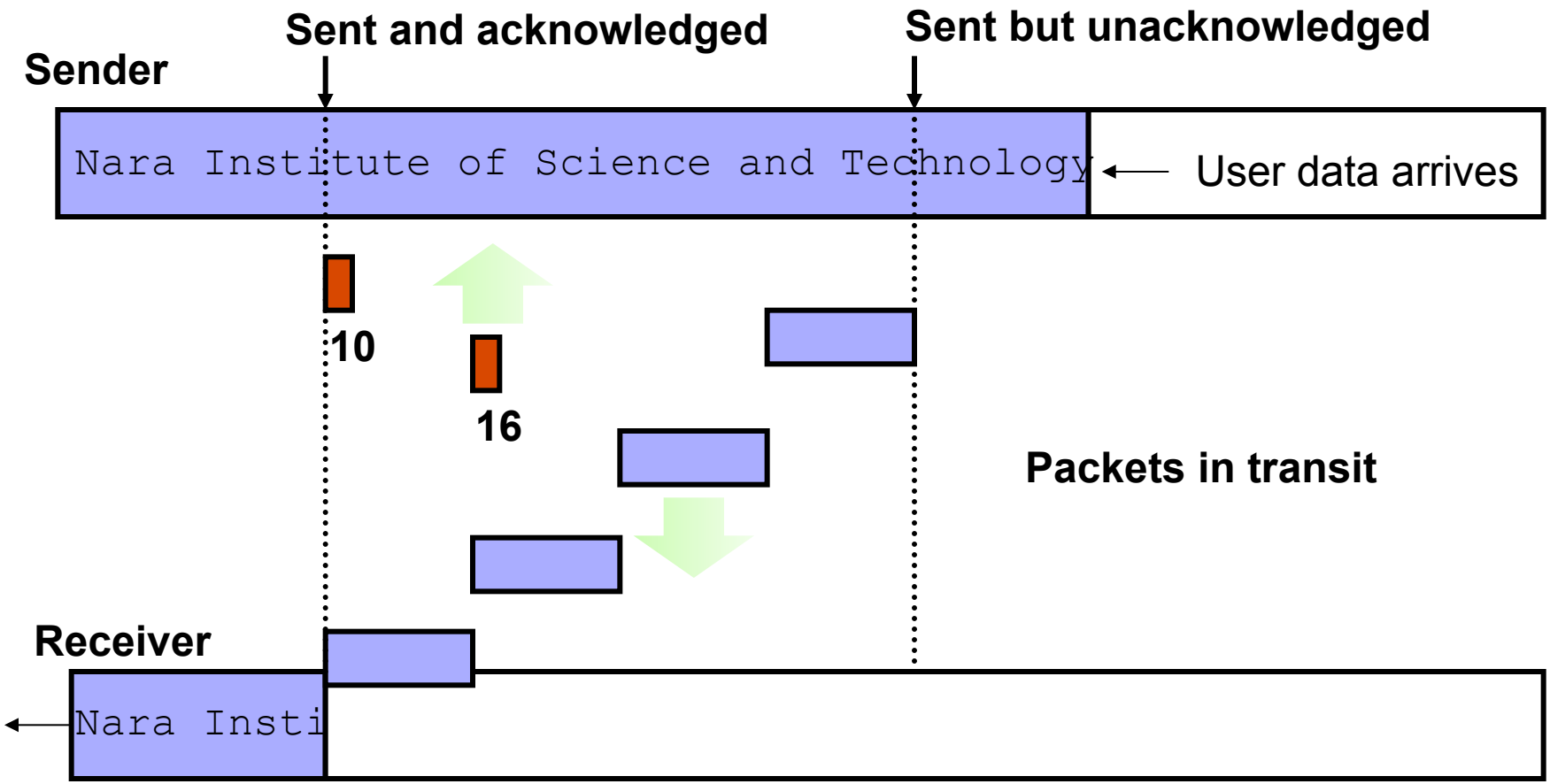
- コネクション指向型
- バイト列
  - 上位層からはパケットの境界は見えない
  - 境界がない　ー　上位層での構造化が必要
- 全二重 (full duplex)
  - 単一コネクション内に独立した2つのストリーム
- 信頼性
  - 配送順序の入れ替わり、重複、パケット廃棄、ビット誤り等を隠蔽

信頼性を有するストリームの実現

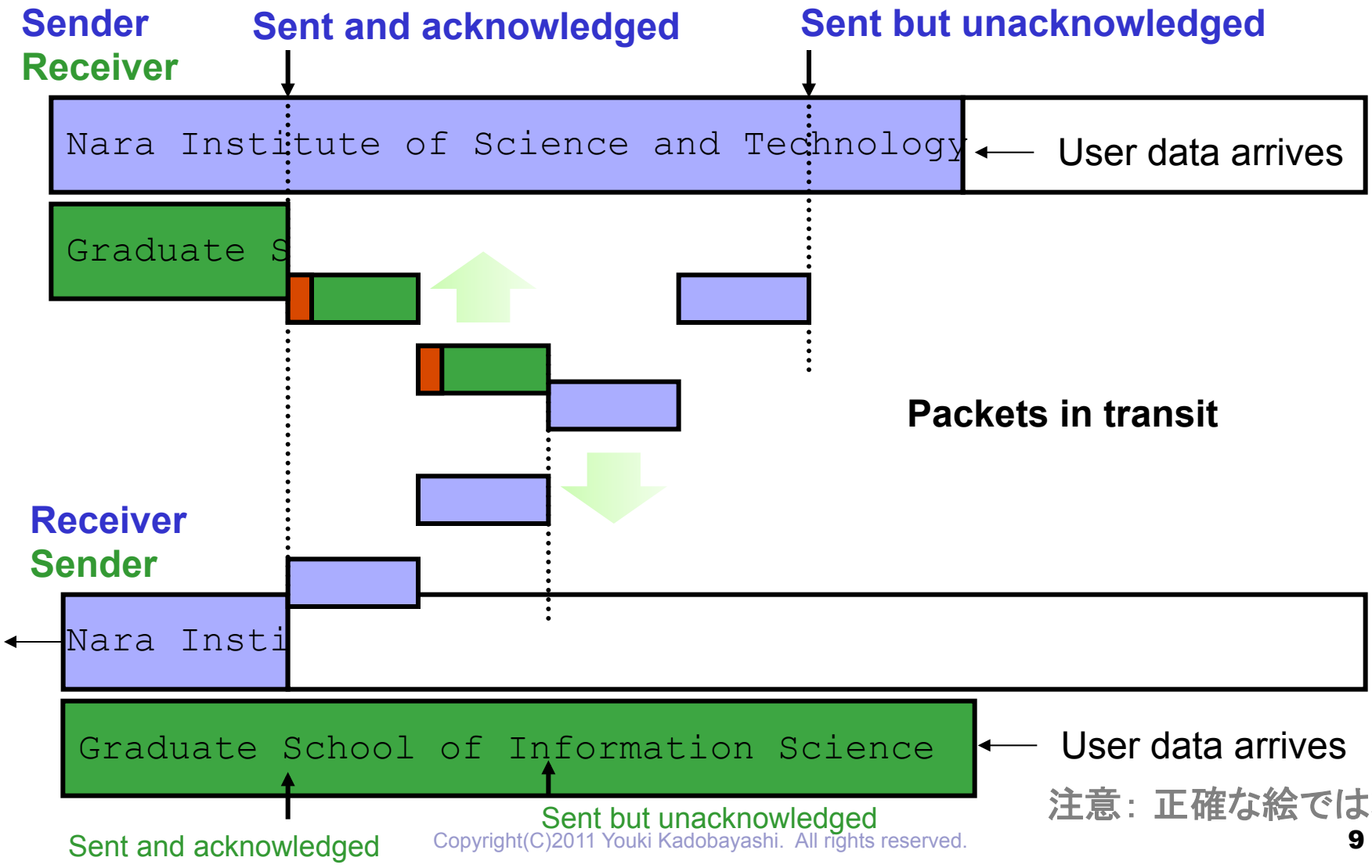
- 確認応答 (ACK: Acknowledgment)
  - Active acknowledgment
    - 明示的に受信の確認応答を返す
  - Duplicate ACK
    - パケットが落ちたことを伝える

- タイムアウトと再送
  - 一定時間を経過しても確認応答を受け取らなかった場合
    → タイムアウト
    送信が正しく完了しなかったと仮定し、パケットを再送
    - 指数的バックオフ (Exponential back-off)

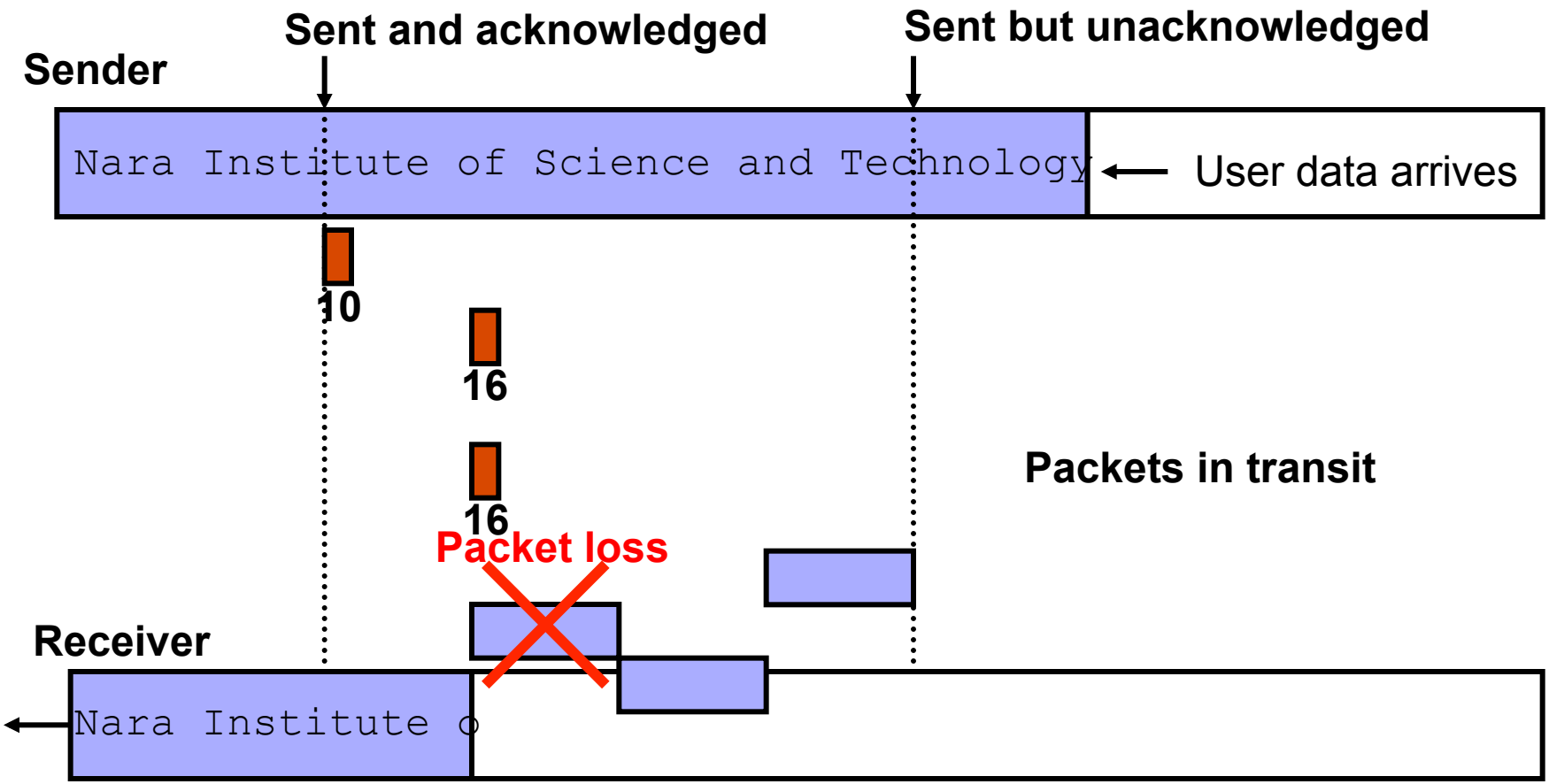
ACK



Piggybacking: 全二重通信時の効率化



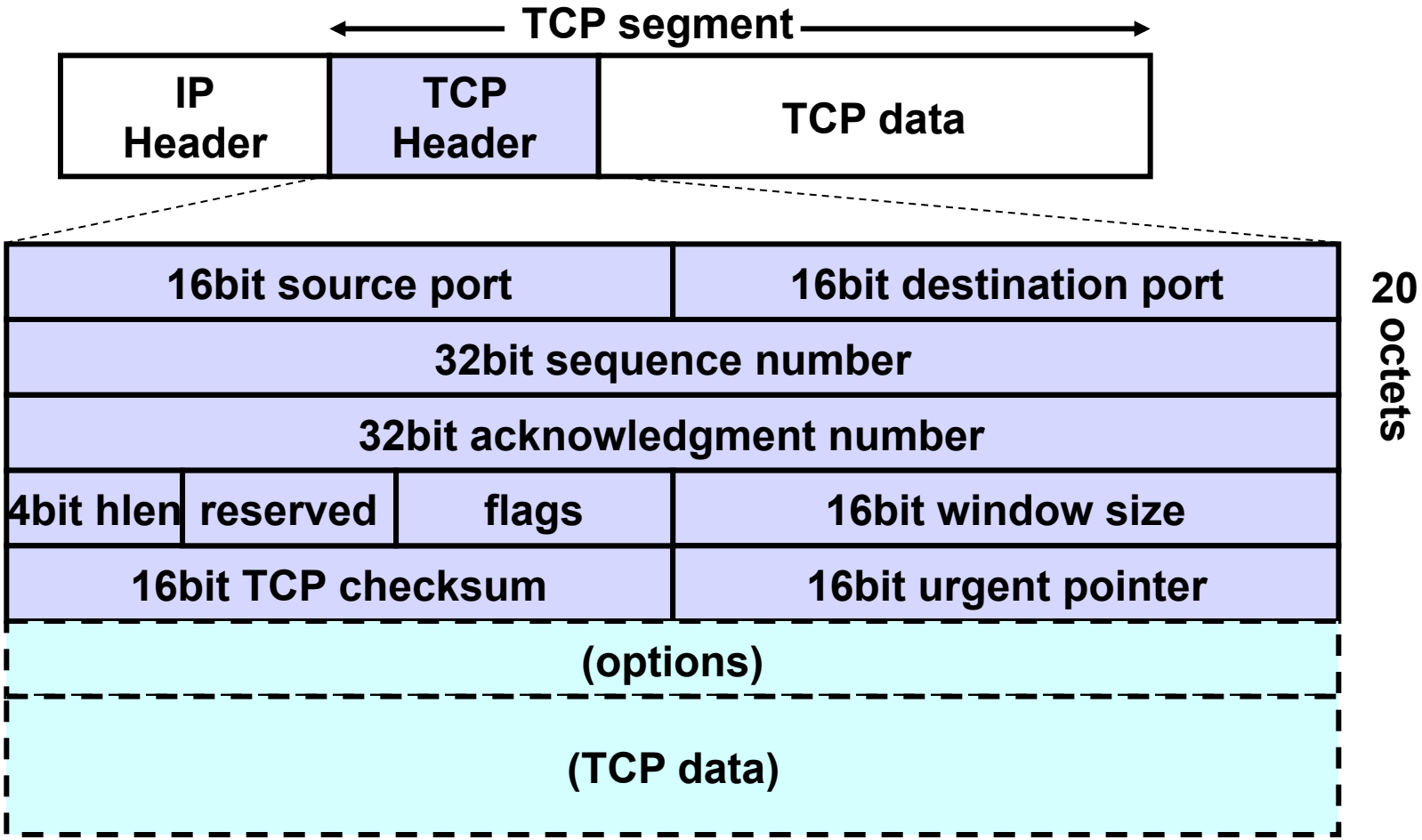
Duplicate ACK



Questions?

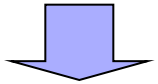
バイト列からパケットへ: TCPのヘッダ

バイト列のバッファから適当な長さを取り出し、TCPヘッダを付与



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 (RFC896)
  - 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

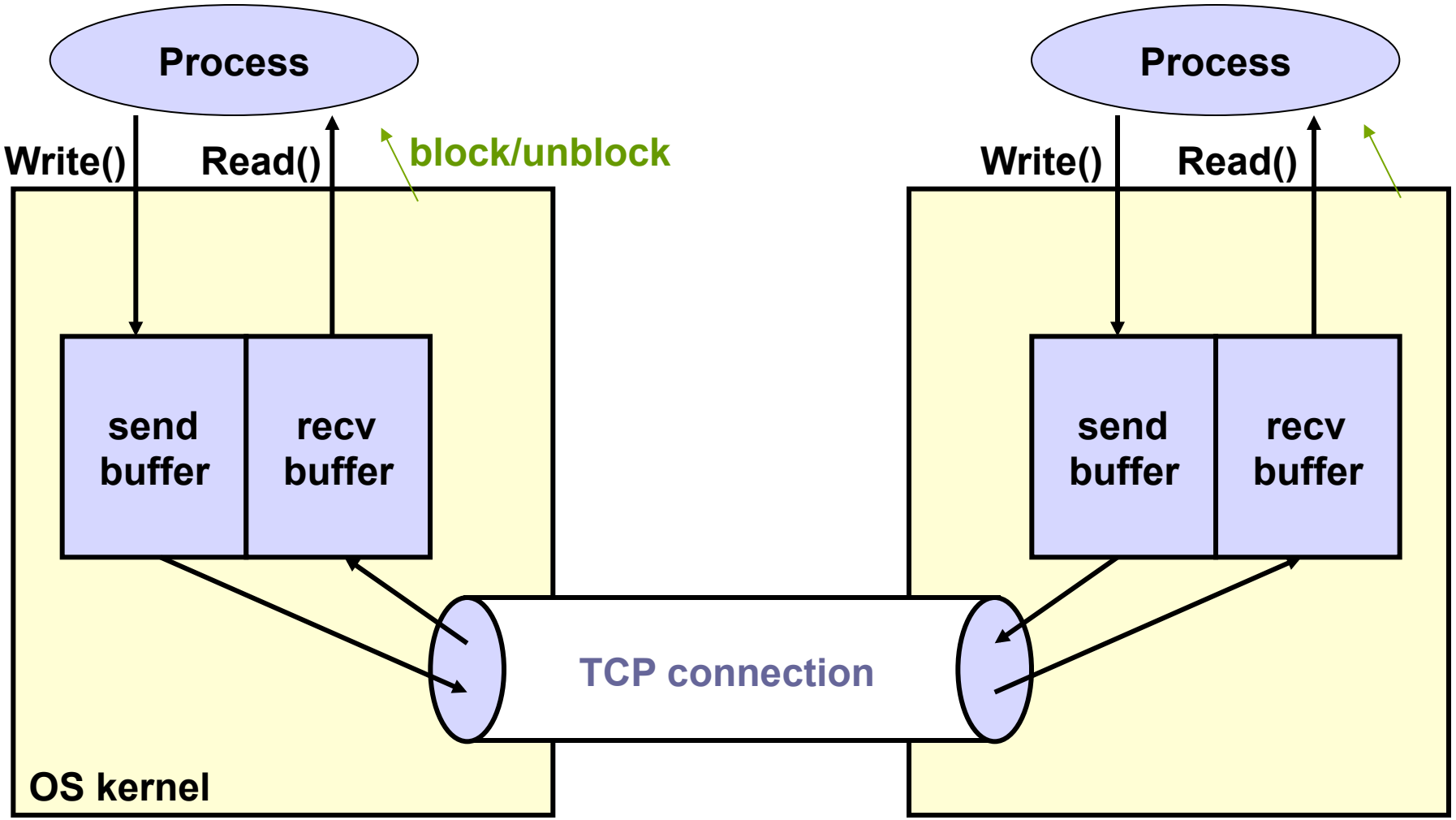
Q.

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

TCPのサービスモデル(2)

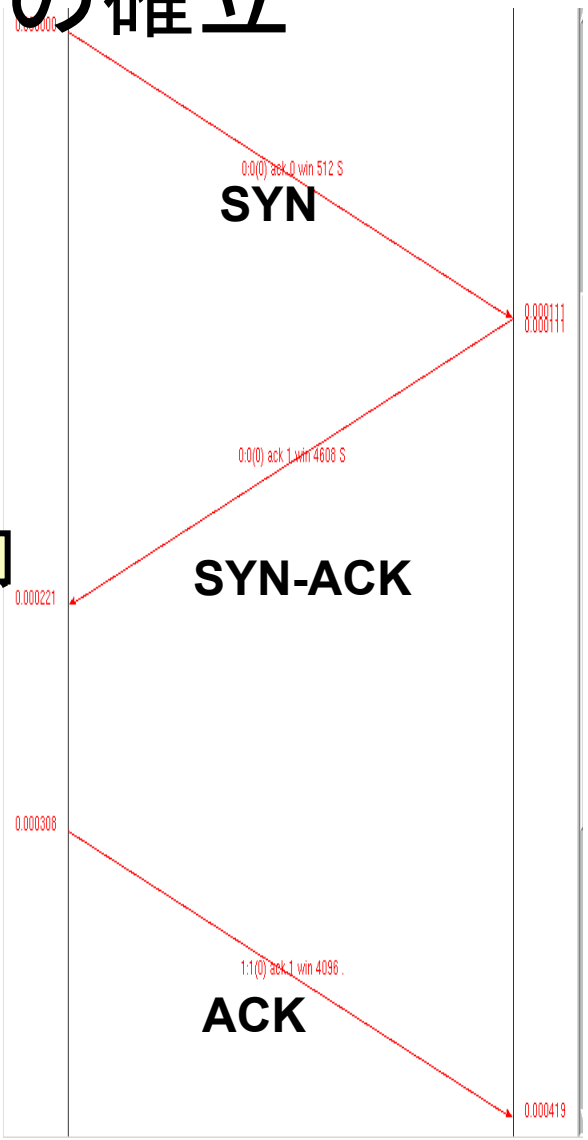
- バッファつき転送
  - いくらでも書き込める（ように見える）
  - アプリケーション側での同期は不要
  - オペレーティングシステムでプロセス状態を変更
- 仮想回線
  - コネクションの確立と解放
  - 切断を検出可能

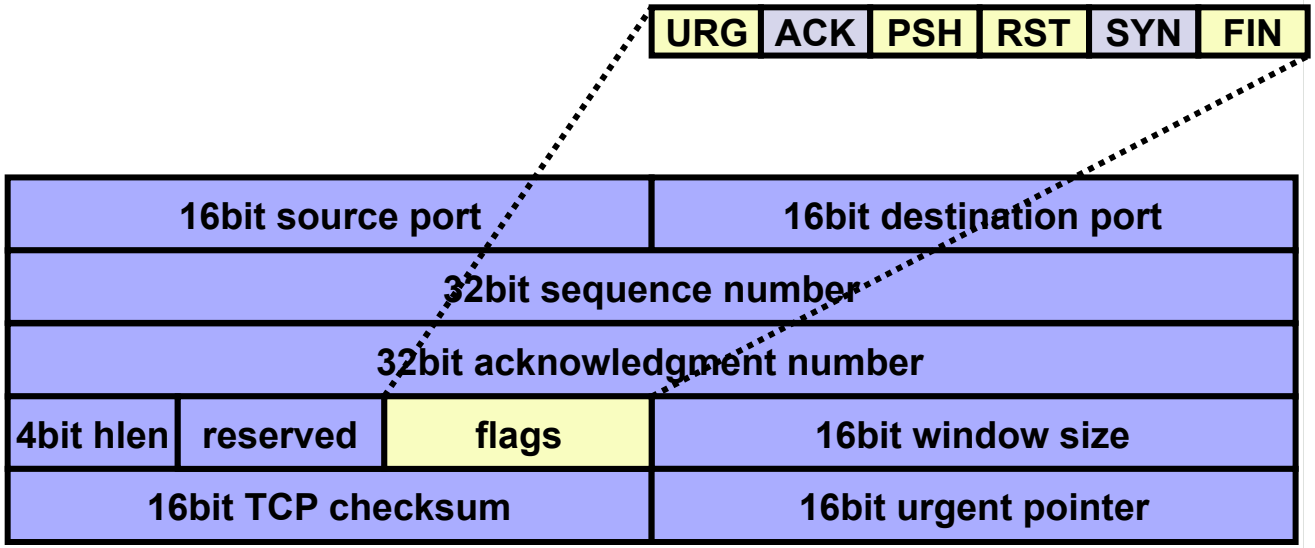
Buffered transfer



仮想回線(1): TCPコネクションの確立

- 3-way handshake
- SYN, SYN-ACK, ACK
- 全二重通信路を検証





TCPコネクションの確立: 実例

- 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

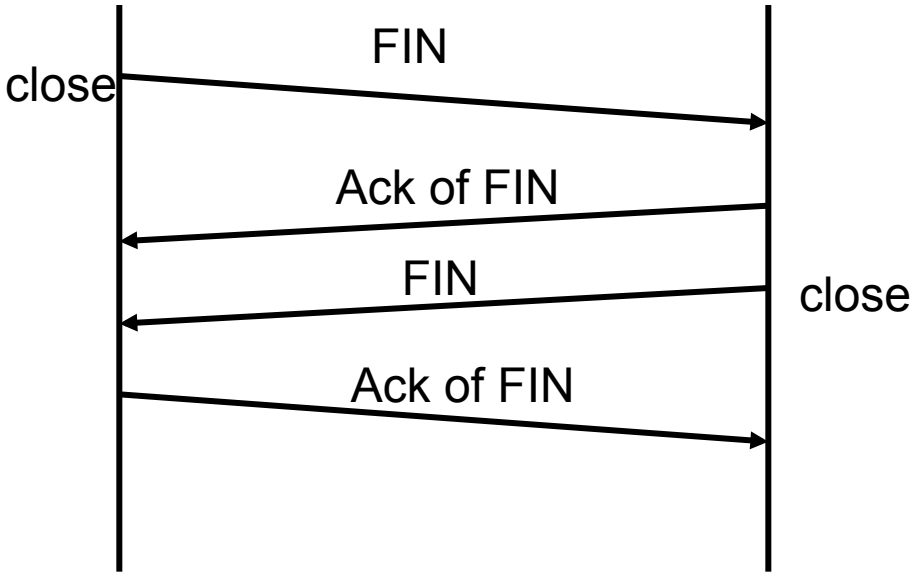
tcpdump出力の意味

- 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)

Questions?

仮想回線(2): 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]

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）

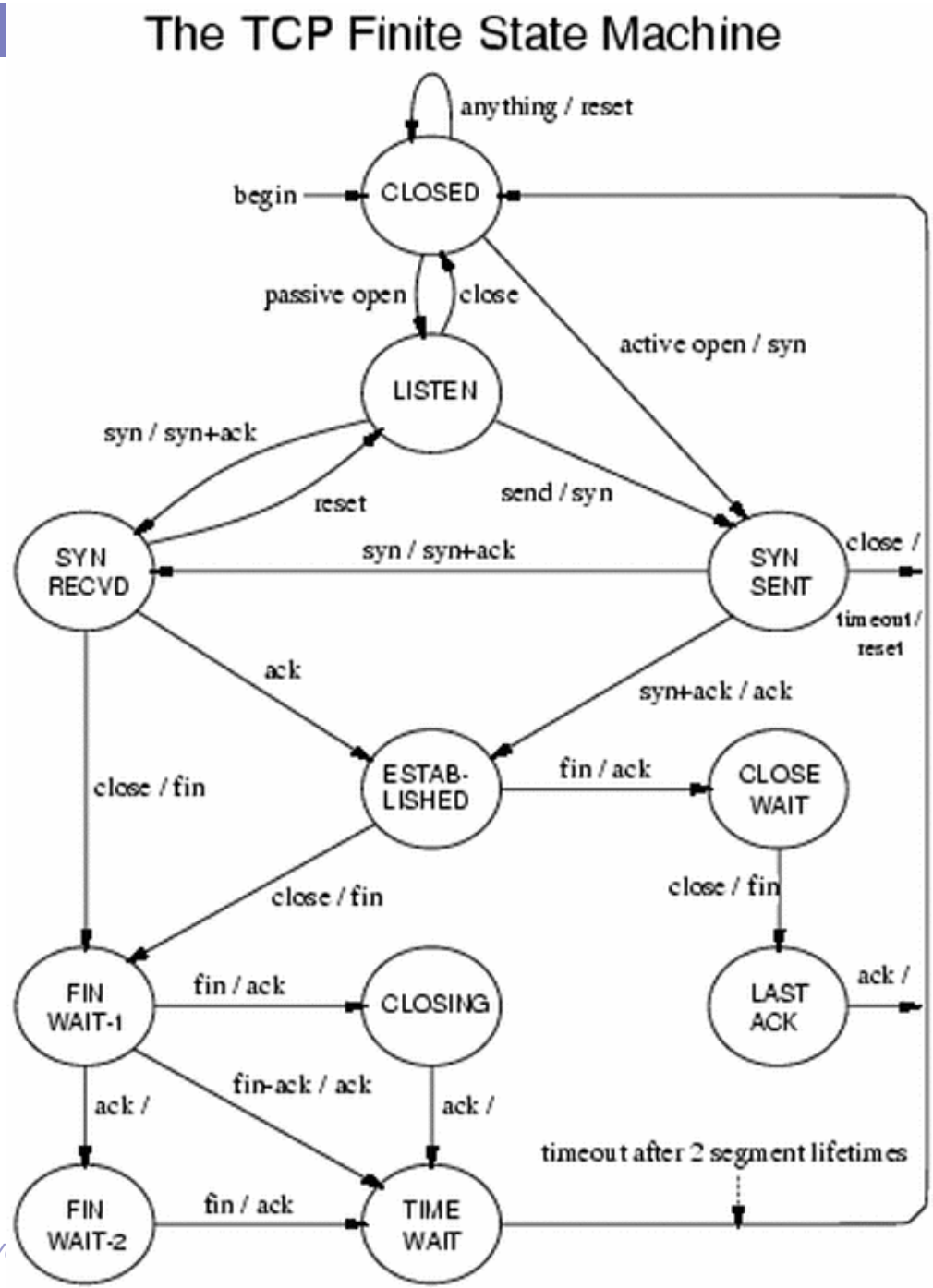
- 13:48:35.948096 dv.aist-nara.ac.jp.49635 > mint100.aist-nara.ac.jp.8080: S 1342403683:1342403683(0) win 65535 <mss 1460,nop,wscale 1,nop,nop,timestamp 111497668 0> (DF) [tos 0x10]
- 13:48:35.948265 mint100.aist-nara.ac.jp.8080 > dv.aist-nara.ac.jp.49635: R 0:0(0) ack 1342403684 win 0

仮想回線(3): 拡張機能の利用

- TCP options in 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 (RFC793, Sep 1981)
  - Window scale option (RFC1323, May 1992)
  - Timestamp option (RFC1323)
  - Selective ACK option (RFC2018, Oct 1996)
  - etc.

仮想回線(まとめ):TCPの状態遷移



Questions?

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のサービスモデル、特徴
- 効率化: ACK, piggybacking, Nagle algorithm
- コネクションの確立、解放