第3章
情報の伝達と通信

3.1 情報の伝達と情報量

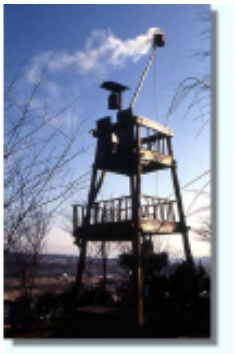
- 2章「情報」をどう表わすか
- 3章「情報」をどう伝えるか?
  - 「情報が伝わる」とはどういうことだろうか?
  - 「物質の移動」と同じこと?
  - 情報の「正確さ」「量」を数量化する方法

情報の表現と伝達は裏表の関係

- 情報の「定義」
  - 人から人に伝達される意味を持った記号系列
- 伝えるための表現方法
- 表現に応じた伝達方法

通信

- 遠隔の伝達
  - マラソン，飛脚，早馬
    →郵便制度



  - のろし，手旗信号，光の点滅信号
    電信，電話
    →通信制度

情報の伝達とは何か?

- **情報の伝達は物質の移動と同じだろうか?**
  - 同じならば、色々な物理法則が適用できる
    - 手紙で情報を伝える場合も、結局は「紙」が移動している
  - しかし、物質の移動は元の場所からなくなる効果もある

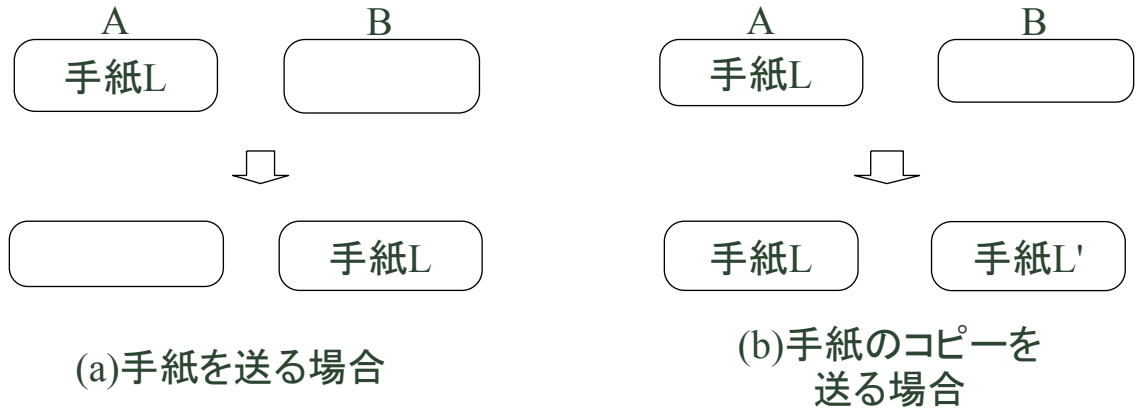
- **情報の伝達は、物質の移動とは違う**
  - 伝えた先の「情報」が増える / 伝える元にも残る
  - 適切な理論が必要!
    - もちろん物理法則の支配下にある (e.g., 光の速度より速くは伝わらない)

配布不可

3.1 情報の伝達と情報量

- **情報の伝達 (3.1.1) とは:**
  受取側の状態の変化が本質
  - 様々な伝え方で同じ「情報」(メッセージ)が伝わる
    - 「手紙」を送る / 「手紙のコピー」を送る
    - 電子メールを送る
  - 手紙の物理的な移動は本質でない
- **情報量 (3.1.2): 情報を受け取った効果を測る**
  - メッセージ: 「今回は日本史から出題する」
  - このメッセージの効果を量で表現すると?

情報の伝達と伝達手段



(a)手紙を送る場合

(b)手紙のコピーを送る場合

- **Bさんが受け取る情報はどちらの場合も同じ**
- **物理的な手紙の移動は無関係**
  - Aさんの手元から手紙が消えることは本質でない

情報を受け取った効果とは?

- **直感的な説明**
  - 情報を受け取った場合
    - 自分に影響がある、これまで知らなかった事実を知った
    - なんらかの判断の材料にできる事実を知った
  - 情報を受け取ったと言い難い場合
    - 関心のない手紙を受け取った (e.g. 迷惑メイル)
- **情報を受け取る効果は、受け取る人の「状態」と関係がある**
- **メッセージの効果を「情報量」として表現したい**

情報の量は何で測るべきか

- **文字数では直感に反する**
  - 1000文字のダイレクトメール vs. 100文字の「情報」試験問題
  - 明日の東京の天気 vs. 明日のブエノスアイレスの天気
    → 受け手にとっての役立ち度が違う
  - 前回のルーレットの目 vs.次回のルーレットの目
    → 受け手がすでに知っているかどうかが違う
  - サイコロの次の目 vs. ルーレットの次の目
    → ルーレットの方が価値が高い?
- **直感に合う情報の量のあるべき性質**
  - 受け手が知っているかどうかを反映したもの
  - 受け手の役立ち度/価値を反映したもの

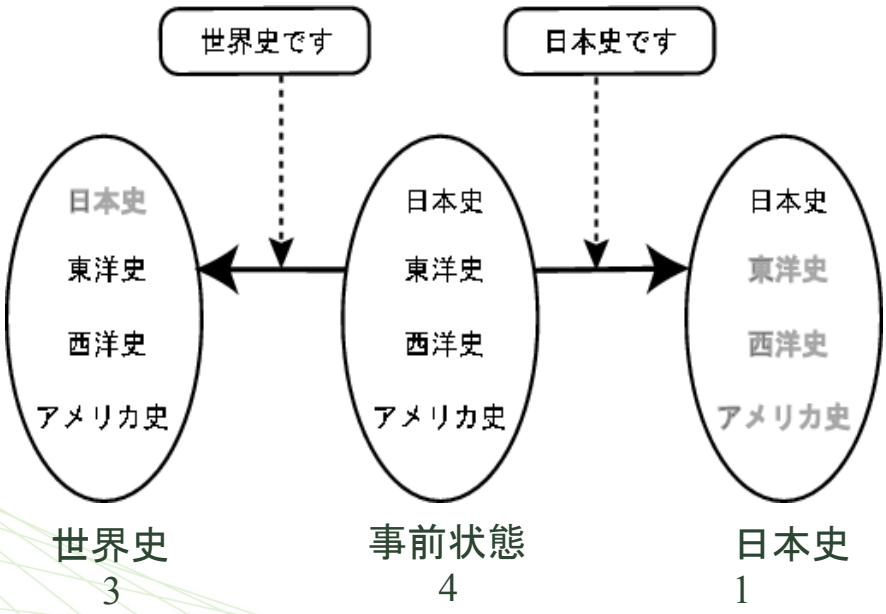
情報の価値 = 場合の数の減少量

- **おおまかな定義: 情報の価値は、受け手の選択肢をどれだけ減らすかで測る**
  - 1000文字のダイレクトメール vs. 100文字の「情報」試験問題
    → 選択肢: 教科書のどのページを勉強するか
  - 明日の東京の天気 vs. 明日のブエノスアイレスの天気
    → 選択肢: どの服を着るか
  - 前回のルーレットの目 vs.次回のルーレットの目
    → すでに知っている情報は選択肢を減らさない
  - サイコロの次の目 vs. ルーレットの次の目
    → 選択肢の数: 6→1 vs. 100→1
- **以降、「選択肢の数」で単純化**

試験に関する情報の価値

- 科目「歴史」の試験
  - 日本史、東洋史、西洋史、アメリカ史のどれか一つが出題
  - 事前にはどれが出題されるかは分からない
- メッセージ:「今回は日本史から出題する」
- 状況の変化
  - 事前: 日本史からアメリカ史の4種類全部の試験勉強が必要である
  - 事後: 日本史の勉強だけですむ

メッセージによる場合の数の変化



場合の数に基づく情報量(の候補)

- **案1: 差**
  - 定義: 事前の場合の数 − 事後の場合の数
  - 問題点: 100 → 97 の場合と 4 → 1 が同じ価値?
- **案2: 商**
  - 定義: 事前の場合の数 / 事後の場合の数
  - 問題点: 情報量の加法性(後述)を満たさない
- **案3: 商の対数**
  - 定義: log(事前の場合の数 / 事後の場合の数)

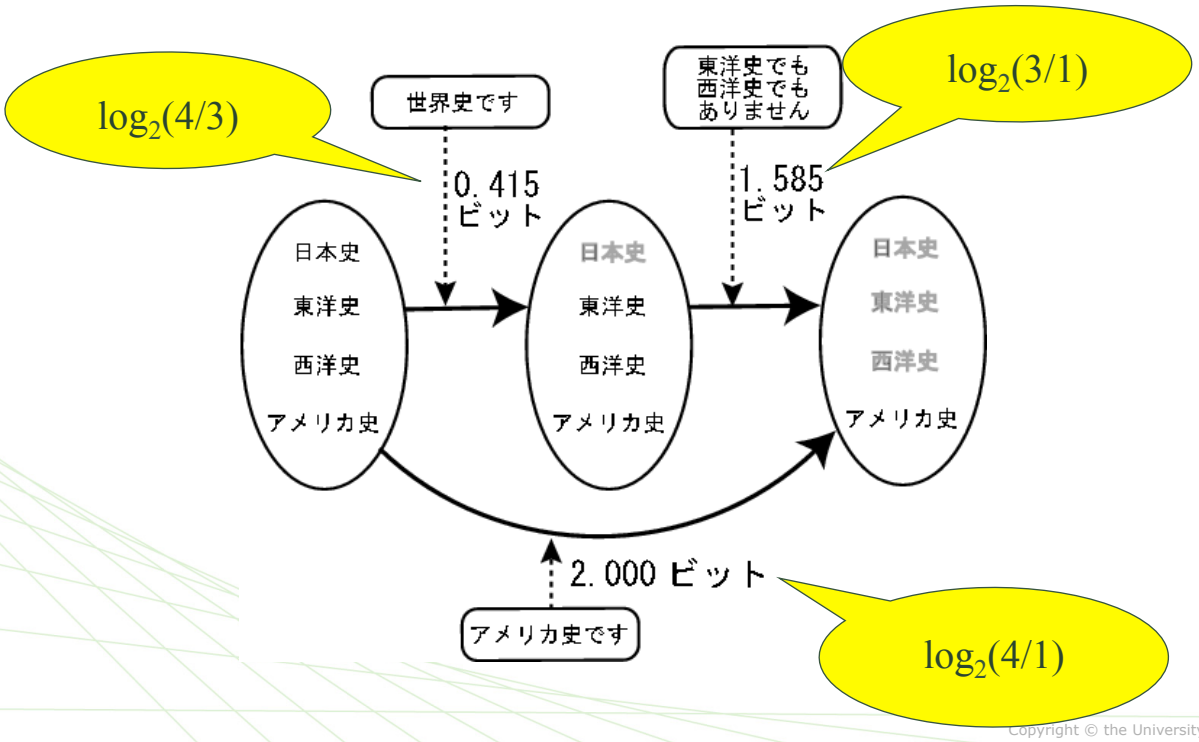
情報量の加法性

- **情報を一度に受け取った場合 (A)**
  - メッセージA:「アメリカ史を出題する」
    場合の数 4 → 1
- **分割して受け取った場合 (B+C)**
  - メッセージB:「世界史を出題する」
    場合の数 4 → 3
  - メッセージC:「東洋史と西洋史は出題しない」
    場合の数 3 → 1
- **情報量(A) = 情報量(B)+情報量(C) としたい**

場合の数に基づいた情報量の定義

- 定義:
  $\log_2$(事前の場合の数 / 事後の場合の数)
- 単位: ビット (bit)
- 性質
  - 場合の数が大きく減る程数が大きい
  - 底が2なので
    二者択一(場合の数が2から1になる場合)に 1.0
  - 情報量の加法性を満たす

情報量の加法性の確認



情報量の定義

↘ 定義: m通りの選択肢をn通りに減らす情報の量

$\log_2(m/n)$ (ビット)

- サイコロの次の目 — $\log_2(6/1)$ ≒ 2.59 ビット
- ルーレットの次の目 — $\log_2(100/1)$ ≒ 6.64 ビット
- 3桁の宝籤の下2桁 — $\log_2(1000/10)$ ≒ 6.64 ビット
- 前回のルーレットの目 — $\log_2(1/1) = 0$ ビット
- コイン投げの裏表 — $\log_2(2/1) = 1$ ビット

情報量の加法性

↘ 2つの情報をまとめた情報の情報量 ＝ 個々の情報量の和


20ビット


30ビット


50ビット

- 例: 3桁の宝籤

?34 — $\log_2(1000/10)$ ≒ 6.64 ビット

＋

8?? — $\log_2(1000/100)$ ≒ 3.32 ビット

＝

834 — $\log_2(1000/1)$ ≒ 9.97 ビット

配布不可

場合の数から確率へ

- 例題: 科目「歴史」の試験 (改)
  - 日本史、世界史のどれか一つが出題させる
  - 事前にはどれが出題されるかは分からない
  - 世界史が75%、日本史 は25%で出題される
- どちらが価値が高いメッセージか?
  - 「世界史が出題される」
  - 「日本史が出題される」
- 場合の数の変化はどちらも同じだが...

確率に基づく情報量の定義

- 定義: $-\log_2$(確率)
- 単位: ビット (bit)
- 性質
  - 確率が低いことを伝えるメッセージほど大きい
    確率1.0 → 情報量 0, 確率0.5 → 情報量 1.0
    確率 0.25 → 情報量 2.0, 確率0 → 情報量無限大
  - c.f. 犬が人間を噛んだ v.s. 人間が犬を噛んだ
  - 場合の数に基づく定義の一般化:
    全てが等確率で起こる時は、場合の数の定義と同じ

平均情報量

➘ 平均情報量

- **p1*log2(p1) - p2*log2(p2) - …**
- **日本史: 25%,世界史: 75%の場合**
  - −0.25 log2(0.25) − 0.75 log2(0.75)

  = 0.25*2 +0.75*0.415

  = 0.811 < 1
  - 1−0.811=0.189はヤマの分

クイズ

天秤を使って9枚のコインから偽物を見つけるための最小回数は?

- 見掛けは全く同じ / 1枚が偽物 /
  本物の重さは全て同じ / 偽物は本物より軽い

- 方法1) A vs B, A vs C, A vs D, …, A vs I (最悪8回)
  - 実は8回目は不要(A〜Hが同じ→Iが偽物)
- 方法2) A vs B, C vs D, E vs F, G vs H (最悪4回)
- 方法3) ABC vs DEF
  → ABCが軽かったら A vs B (2回)

クイズ

天秤を使って9枚のコインから偽物を見つけるための最小回数は?

- 見掛けは全く同じ / 1枚が偽物 / 本物の重さは全て同じ / 偽物は本物より軽いか重いかのどちらか
- 3回でできる

コインが13枚の場合は?

偽物が2枚の場合は?

情報量からの考察

- **計測1回の情報量: 一方が重い/軽い/同じの3通り→ $\log_2(3)$ ≒ 1.58 ビット**
- **偽物の存在する場合の数**
  - 9枚中1枚だけ軽い: $\log_2(9)$ ≒ 3.17 ビット

    → 3.17/1.58 = 2回で分かる
  - 9枚中1枚だけ軽いか重い: $\log_2(9x2)$ ≒ 4.17 ビット

    → 3 > 4.17/1.58 > 2 → 3回以上
  - 13枚中: $\log_2(13x2)$ ≒ 4.7 ビット

    → 3 > 4.7/1.58 > 2 → 3回以上

情報量の差異の応用：符号化と情報量

- 情報は0,1の符号で表され，伝達される
- 伝送速度が一定ならば，小さいデータほど早く伝送できる
  - データは復元可能なように圧縮して伝送する
- 例：二年分の試験出題情報を符号化する場合，珍しい情報には長い符号を，珍しくない符号には短い符号を割り当てる.
  - 平均符号長＝Σ（符号長i×確率p）

符号化と情報量（例）

出題	確率	符号	符号長
日本史＋日本史	1／16	111	3
日本史＋世界史	3／16	110	3
世界史＋日本史	3／16	10	2
世界史＋世界史	9／16	0	1

- 平均符号長＝0．844
  - 1年分の試験の情報を表す符号長（＝1）より平均符号長が短くなっている

3.2 情報通信

- プロトコル (3.2.2) (a)
  - 通信の際の決めごと
- 通信の秘密と相手の認証 (3.2.3) (a)
  - 暗号 盗聴を防ぐ
  - (認証 通信参加者の身元の保証)
  - (署名 通信内容の改竄の防止、否認の防止)

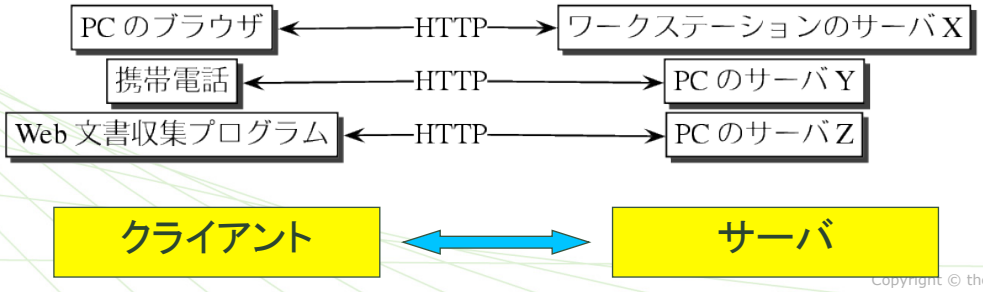
通信プロトコル

- 決められた種類の情報を伝える場合には、あらかじめ伝え方に約束事がある
  - 例)手紙: 宛名を書く場所、「気付」、差出人を書く場所、日付を書く場所、等々
    → 間違えると届かなかったり、誤解されたり
- インターネットの通信にも、種類によって約束事がある
  - 例) 電子メール: 宛先のメールアドレス、差出人のメールアドレス、日付を書く場所、等々
  - 例) WWW: URL, データの大きさ, データの種類, 更新された日時, 言語、等々

→ プロトコルと呼ぶ

プロトコル（protocol）

- 通信の意図を理解するための決めごと
  - 電話「もしもし」, トランシーバ「どうぞ」
- コンピュータ同士の通信: 人間の場合より厳密
  - WWW（HTTP, HyperText Transfer Control P.）
  - 電子メール（SMTP, Simple Mail Transfer P.）
- プロトコルを正しく使えば機器によらず通信可能



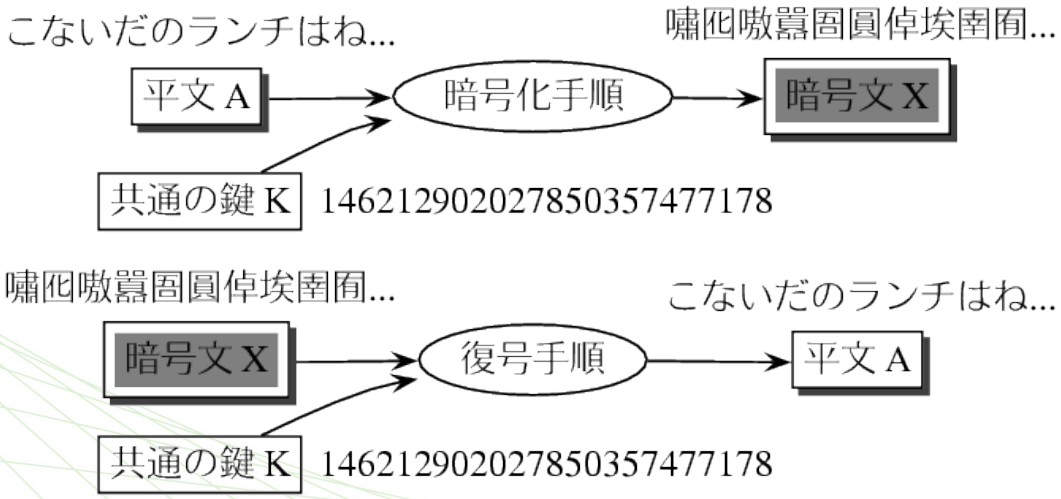
暗号

- 平文: 元のデータ。第三者に読まれたくないもの
  - 「明日のランチはね…」
- 暗号文: 変換後のデータ。盗聴されても平文を(簡単には)取り出せない。
  - 「嘯囮噉囂圀圓倬埃圉囿…」
- 暗号化: 平文から暗号文を作成すること
- 復号: 暗号文から平文を取り出すこと
- 鍵: 暗号化や復号の際に用いられるデータ

共通鍵暗号

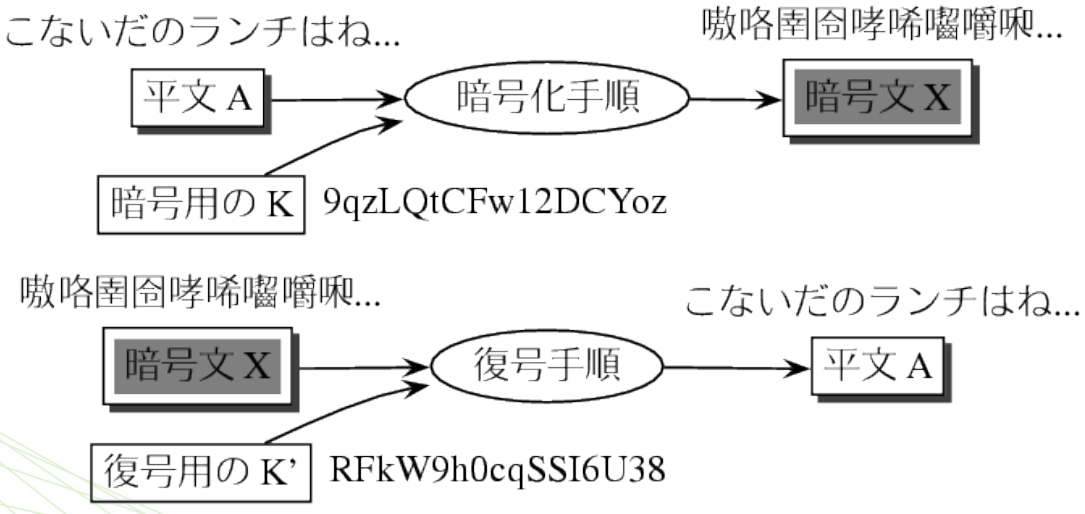
一つの鍵で暗号化と復号化の両方を行うモデル

- 鍵を秘密に保つ必要がある

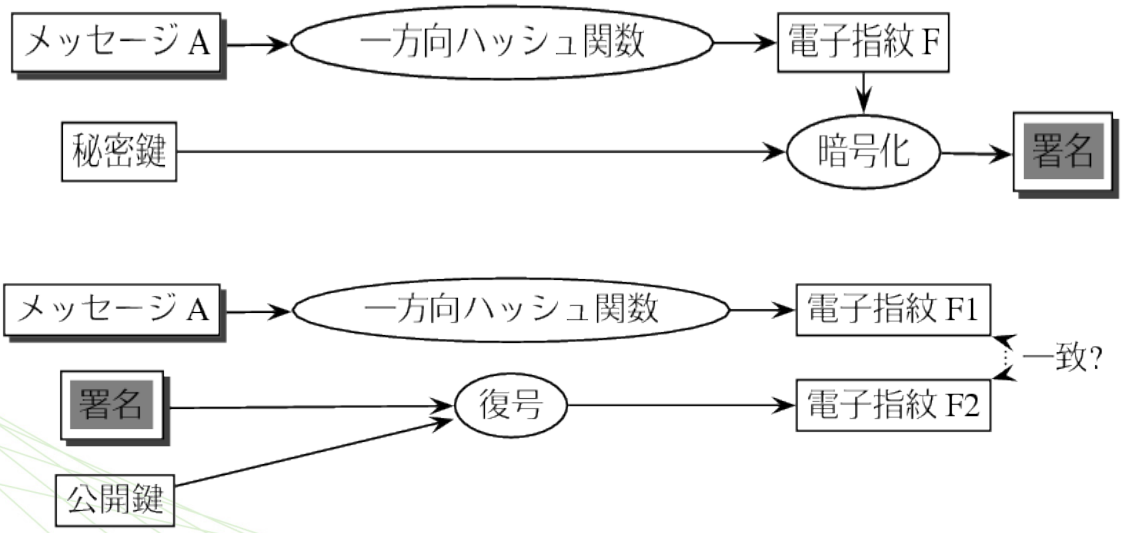


公開鍵暗号

暗号化の鍵を公開し，復号化の鍵を秘密にする方法



デジタル署名と検証



発展

鍵を渡さずに鍵を持っていることを証明する

チャレンジ・レスポンス

確認する人がランダムに選ぶ

q843fjf8e374343

証明する人

q843fjf8e374343

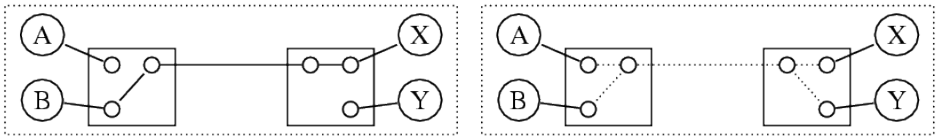
確認する人

ICカード学生証

3.3 情報ネットワークの枠組

交換の方式 (3.3.1)

- 交換機: 通信される情報を経路に振り分ける
- 回線交換: 通信路を確保
- パケット交換: データを細かく分けて順番に通信
  - ちぎっては投げ，ちぎっては投げ



回線交換　　パケット交換

遅延（latency）とスループット（throughput）

broadcastとunicast

交換方式の特性

交換方式	回線（電話）	パケット（インターネット）
流れる情報の種類	音声のように途切れては困るもの	WWWのように時間がかかっても構わないもの
料金体系	回線を占有している時間に対して課す	全体の設備を使う権利に対して課す
端末の能力	単純でよい	データをためる・送り直す等の能力が必要
交換機の能力	高くないといけない	比較的低い

3.4 インターネット

- ネットワークの集合体と通信 (3.4.1)
- 階層プロトコル (3.4.2)
- IPアドレスとポート番号 (3.4.3)

3.4 インターネット

- インターネットを使って2つのコンピュータが通信をしているとき、実際にはどこで何が起きているのか?

注意:

- 2つのコンピュータは直接つながっていない
- 各コンピュータは、同時に色々な相手と通信している
- 同時に色々な種類の通信もしている (例: メールとWWW)
- 世界中のコンピュータが通信できる

インターネットの通信

- インターネットとは広義には「ネットワークどうしをつないだネットワーク」のこと
- どうやって通信しているか?
  - インターネットで通信をする場合は、情報は色々なネットワークを渡り歩いてゆく
  - 個々のネットワークの中では、コンピュータどうしが直接通信できる
  - ネットワークどうしをつなぐ機器を「ルータ」とよぶ

インターネット通信の実際

例: 東大からテルアビブ大学へのメッセージ

```
ux104$ traceroute post.tau.ac.il
traceroute to post.tau.ac.il (132.66.16.11), 30 hops max, 40 byte pack
 1  133.11.50.158 (133.11.50.158)  0.78 ms  0.226 ms  0.224 ms
 2  192.168.254.65 (192.168.254.65)  0.612 ms  0.49 ms  0.466 ms
 3  133.11.249.242 (133.11.249.242)  0.775 ms  0.67 ms  0.647 ms
 4  ra36-vlan2.nc.u-tokyo.ac.jp (133.11.127.43)  0.757 ms  0.76 ms  0.
 5  ra37-vlan3.nc.u-tokyo.ac.jp (133.11.127.78)  0.803 ms  0.74 ms  0.
 6  tokyo-s1-g2-0.sinet.ad.jp (150.99.197.169)  0.939 ms  1.229 ms  0.
 7  jt-tokyo-s1-p3-0.sinet.ad.jp (150.99.197.37)  2.186 ms  2.924 ms
 8  nii-s1-p4-0.sinet.ad.jp (150.99.197.22)  2.779 ms  2.615 ms  2.598
 9  nii-gate2-p2-0.sinet.ad.jp (150.99.199.174)  2.641 ms  2.473 ms  2
10  nii-gate3-p3-0.sinet.ad.jp (150.99.198.246)  189.415 ms  189.298 m
11  sinet.ny1.ny.geant.net (62.40.103.233)  199.775 ms  199.893 ms  19
12  ny.uk1.uk.geant.net (62.40.96.170)  263.855 ms  263.747 ms  263.85
13  uk.nl1.nl.geant.net (62.40.96.181)  280.256 ms  280.372 ms  280.35
14  nl.il1.il.geant.net (62.40.96.118)  345.219 ms  344.887 ms  345.01
15  iucc-gw.il1.il.geant.net (62.40.103.70)  343.861 ms  343.667 ms  3
16  tau-gp1-fe.ilan.net.il (128.139.191.69)  343.89 ms  343.845 ms  34
17  * * *
18  * * *
```

インターネットの多様性と階層プロトコル

- インターネットの形態は様々
  - 通信内容は色々: ビデオ・電話・WWW・メール・・・
  - 機器の種類も様々: 無線LAN, 有線LAN, CATV, FTTH, 電話回線, 電力線, ...
- 1つのネットワークで色々できるのは何故か?

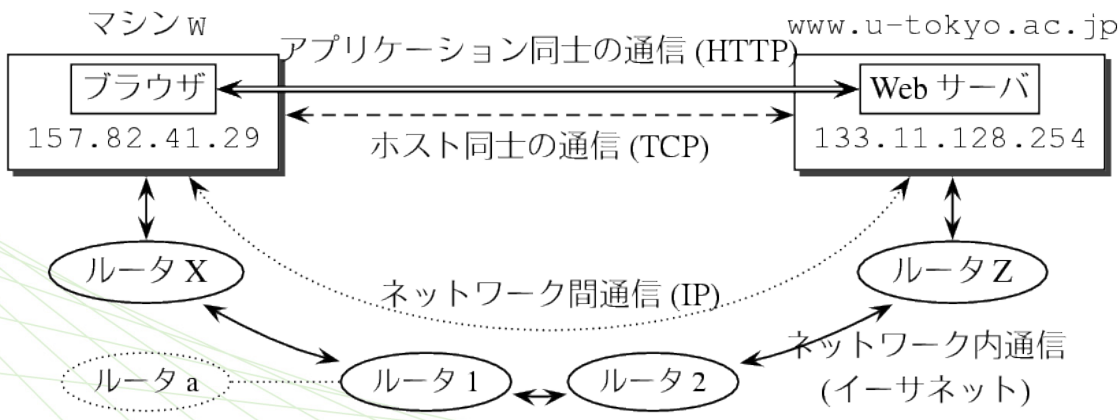
  cf. 電話とケーブルテレビは別々のネットワーク

→ 階層プロトコル

  - 色々なレベルでのプロトコルを決めておく
  - プロトコルが同じであれば取り替え可能になる

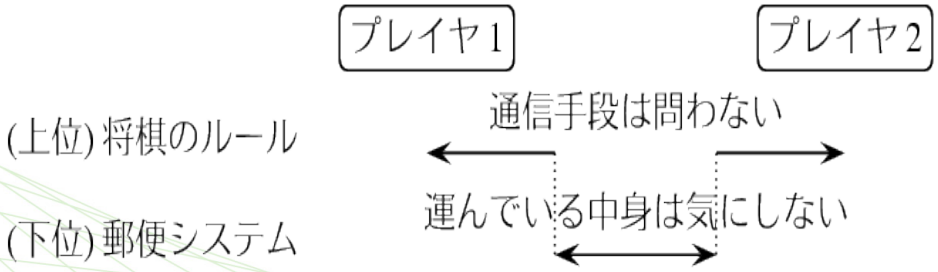
インターネットの通信

- ネットワークの集合体: グループごとに管理
- ルータ: ネットワーク間の通信を中継
- 様々なプロトコル: 役割毎に分割



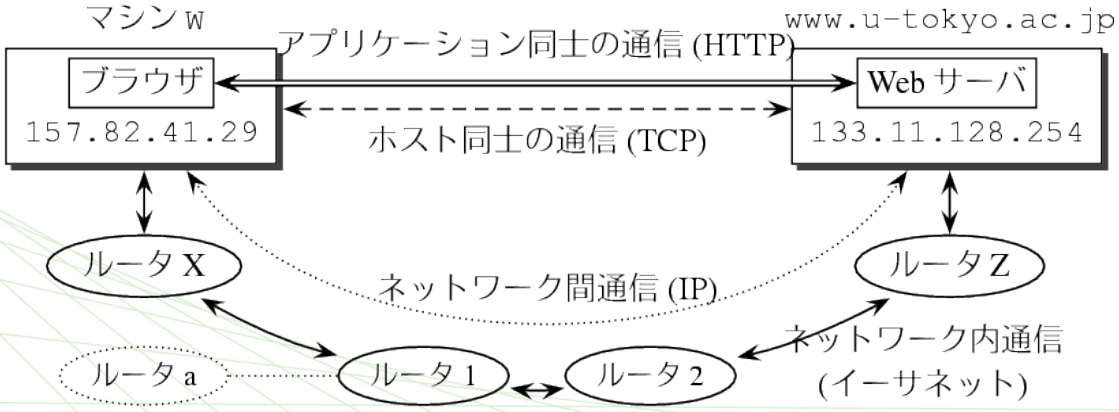
階層プロトコル（例）

- 全ての場合に備えたプロトコルを準備する無理
  - インターネットでオセロをするプロトコル
  - 郵便で将棋を指すプロトコル，携帯電話で囲碁…
- 解：通信とゲームのプロトコルを分離
  場合に応じて組み合わせ可能に



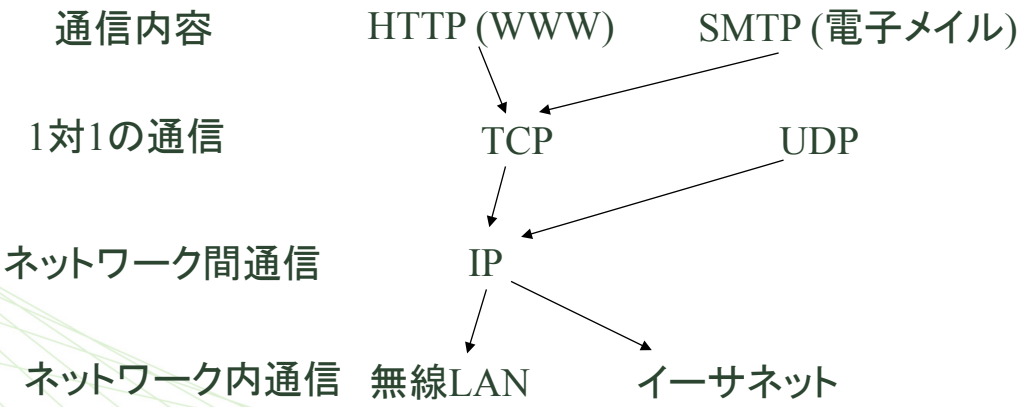
インターネットの場合

- アプリケーション(WWW,電子メール…):
  1対1の通信の部分が共通 → TCP
- ネットワーク内通信:
  媒体(無線LAN, イーサネット…)毎に異なる



TCP/IP 階層プロトコル

- 共通の通信手順は同じプロトコル
- 異なる部分だけ取り換え可能



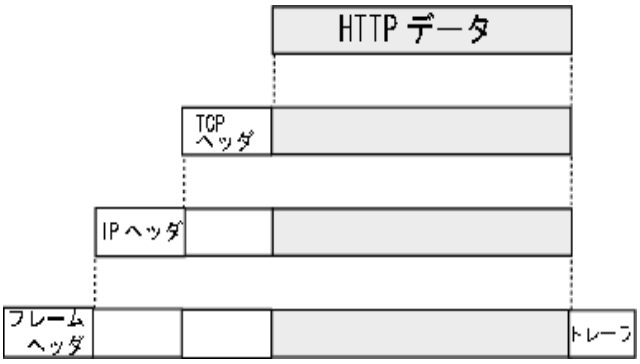
OSI（Open Systems Interconnect）の7階層モデル

発展

アプリケーション層	いまどきはーまとめ
プレゼンテーション層	HTTP, SMTP, SCP, …
セッション層	
トランスポート層	TCP, UDP
ネットワーク層	IP
データリンク層	Ethernet, PPP, X.25,…
物理層	光ファイバ, 銅線, 伝書鳩,…

カプセル化

- 階層毎に制御用のデータを付加する
  - ヘッダ: 先頭に付加されたもの
  - トレーラ: 末尾に付加されたもの
- 役割
  - データの宛先
  - 誤り訂正
  - 順序の制御など



IPアドレス

- IPアドレス: インターネット内の住所
  - 32bit の数値、8bit毎に表記: 192.168.1.3
  - インターネットに接続するホスト
    → 一意のアドレスを必ず持つ
  - 連続する番号が意味を持つ
    - 組織毎にIPアドレスのまとまりで使用を許可される
    - ネットワークの住所を表す
- (ポート番号: 同じホストの複数の通信を区別)

IPアドレス，もう少々

ホストのアドレスとネットワークのアドレス

- IPアドレスの2進表示の末尾に0が並んでいるものがネットワークアドレス

  192.168.12.240

  $= 11000000\ 10101000\ 00001100\ 11110000_2$

  192.168.12.241～192.168.12.254までがそのネットワークで使えるホストのアドレス （192.168.12.255は同報のために使う）

  このようなネットワークを 192.168.12.240/28 と書く

  28は先頭から28ビットまでがネットワークアドレスの範囲

- 2進数のよい訓練になる

IPアドレス，もう少々

プライベートアドレス

- 内線番号のようなもの

  192.168.0.0/16

  172.16.0.0/12

  10.0.0.0/8

特殊アドレス

- ブロードキャストアドレス
  - ホストの部分が2進数ですべて1のもの
- 自分自身をあらわす

  127.0.0.1

ドメイン名

- **IPアドレス:　32ビットの整数・パケットの宛先として実際に使用**
- **ドメイン名:　人間が使うためのもの・階層化されて整理**
- **ドメイン名システム**
  - ドメイン名からIPアドレスを調べる仕組み(cf.電話帳)

mail. ecc. u-tokyo. ac. jp
情報基盤センタ　東京大学　学術機関　日本

どうやって4億個の名前を管理して調べるか?

- 各ドメインには、ドメイン名サーバ(DNS)というコンピュータが用意されている
- DNSとインターネットを使って通信して調べる

ネットワーク間通信-IP

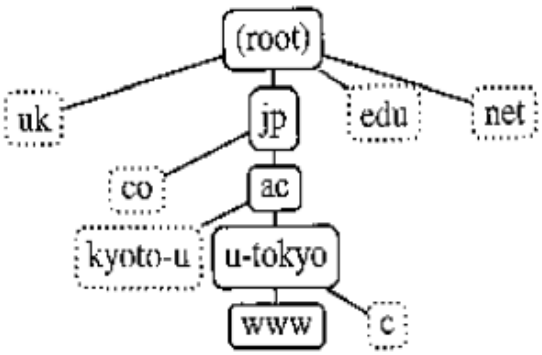
- **ネットワーク間伝達の仕組み**
  - 同一ネットワーク内に宛先があれば直接転送
  - そうでない場合は，宛先に送るのに適した同一ネットワーク内のルータに転送
- **IPアドレスネットワークアドレス**
- **経路制御**
  - 経路表の利用
    - 静的経路制御
    - 動的経路制御

ネットワーク内の通信

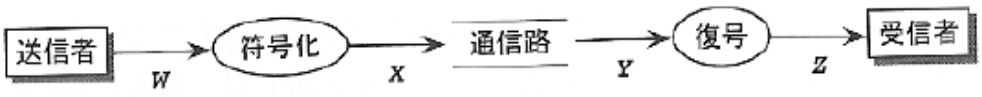
- 通信の媒体によって異なる
- 代表的な通信媒体であるイーサネットの例
  - コンピュータを識別する48ビットのMACアドレス
  - 製造会社の番号と会社内の一意の番号の組み合わせ
  - IPアドレスとの変換にはARP（Address Resolution Protocol)が利用される

IPアドレスとホスト名の対応付け-DNS

- 人間に使いやすい名前（ホスト名）の利用
- 分散管理
  - 階層ごとに限られた情報を管理
  - 反復問い合わせによる解決
  - 問い合わせの結果の再利用



通信路のモデル



二元対称通信路

- 使われる文字は0と1の二つ
- 確率qで0と1が反転
- q=0.5の時は役に立たない通信路

結合エントロピー

↘ 結合エントロピーH(X,Y)

$$H(X,Y) = -\sum_i \sum_j p(x_i, y_j) \log p(x_i, y_j)$$

$$H(X|y=0) = -\sum_i p(x_i|y=0) \log p(x_i|y=0)$$

$$\begin{aligned} H(X|Y) &= -\sum_j p(y_j) H(X|y=y_j) \\ &= -\sum_i \sum_j p(x_i, y_j) \log p(x_i|y_j) \\ &= H(X,Y) - H(Y) \end{aligned}$$

相互情報量

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) = I(Y;X)$$

通信路容量

- Xの確率分布を変更したときの相互情報量I(X;Y)の最大値
- 符号化の方法によって相互情報量を通信路容量に近づけることが出来る