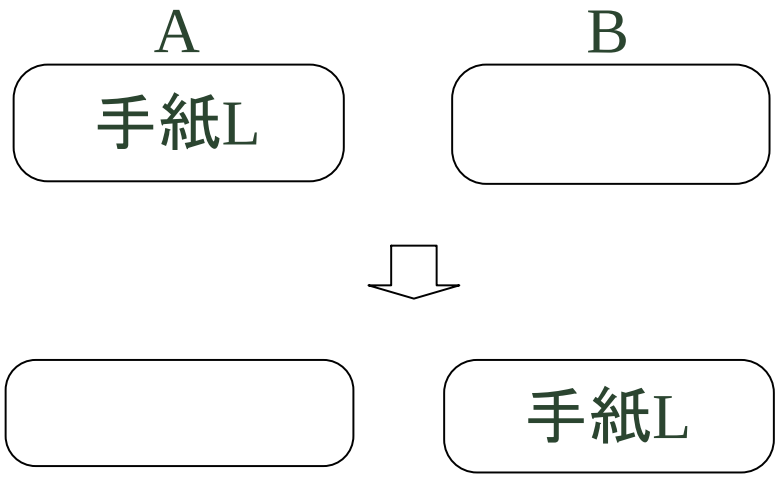
Introduction to the New Information World

第3章 情報の伝達と通信

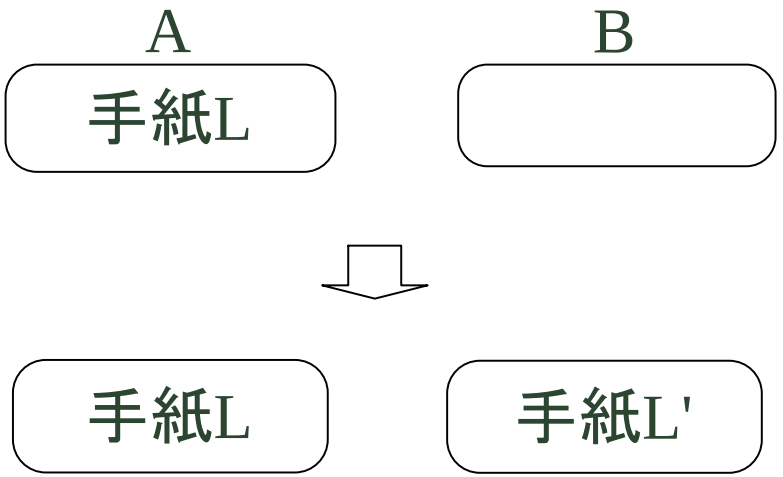
3.1 情報の伝達と情報量

- **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

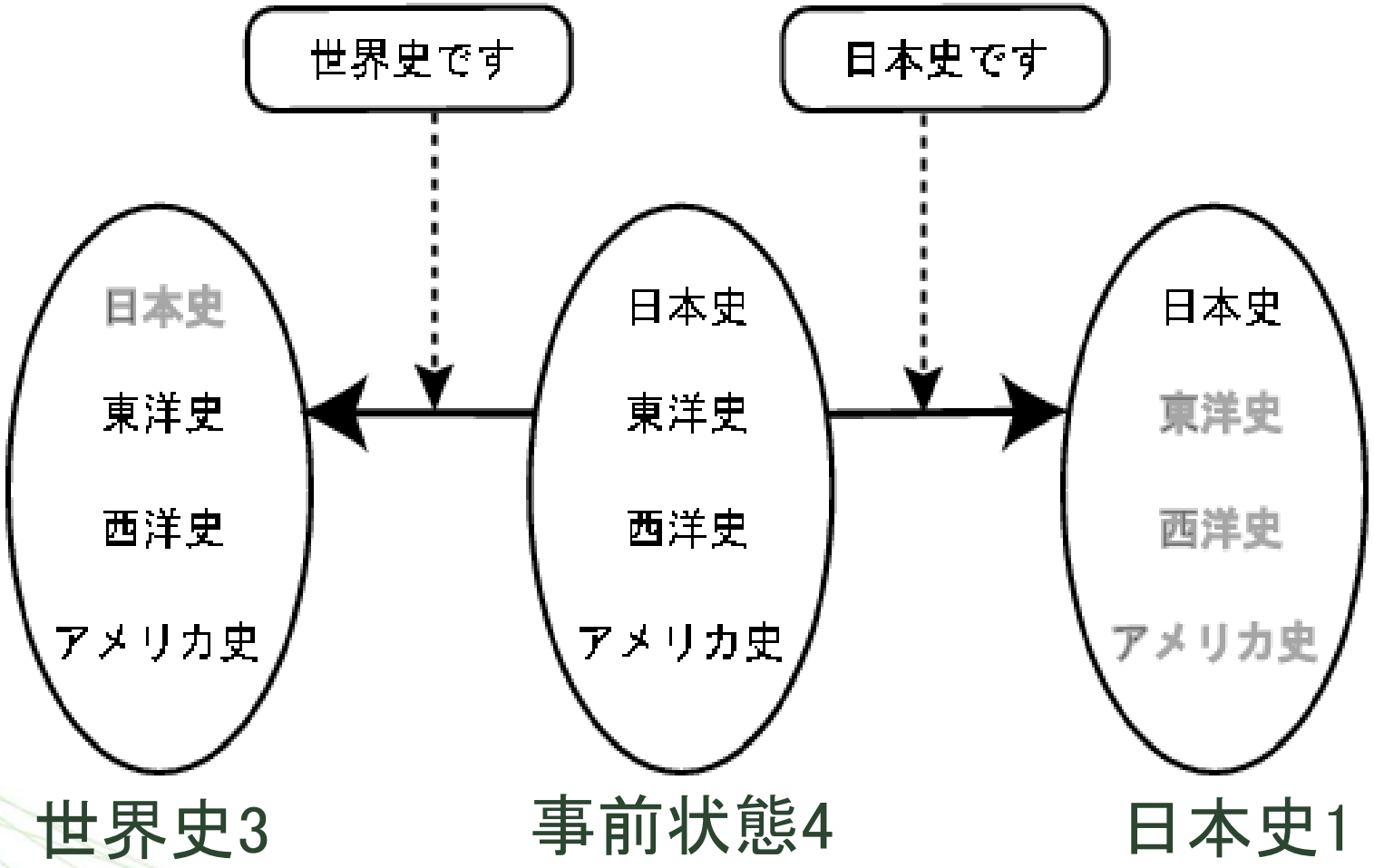
情報を受け取った効果とは?

- 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)
    - 既知の情報を受け取った
- 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

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
- 場合の数の変化～情報量？

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)
  - 利点:上記の問題は，対数の性質を利用すると，解消できる

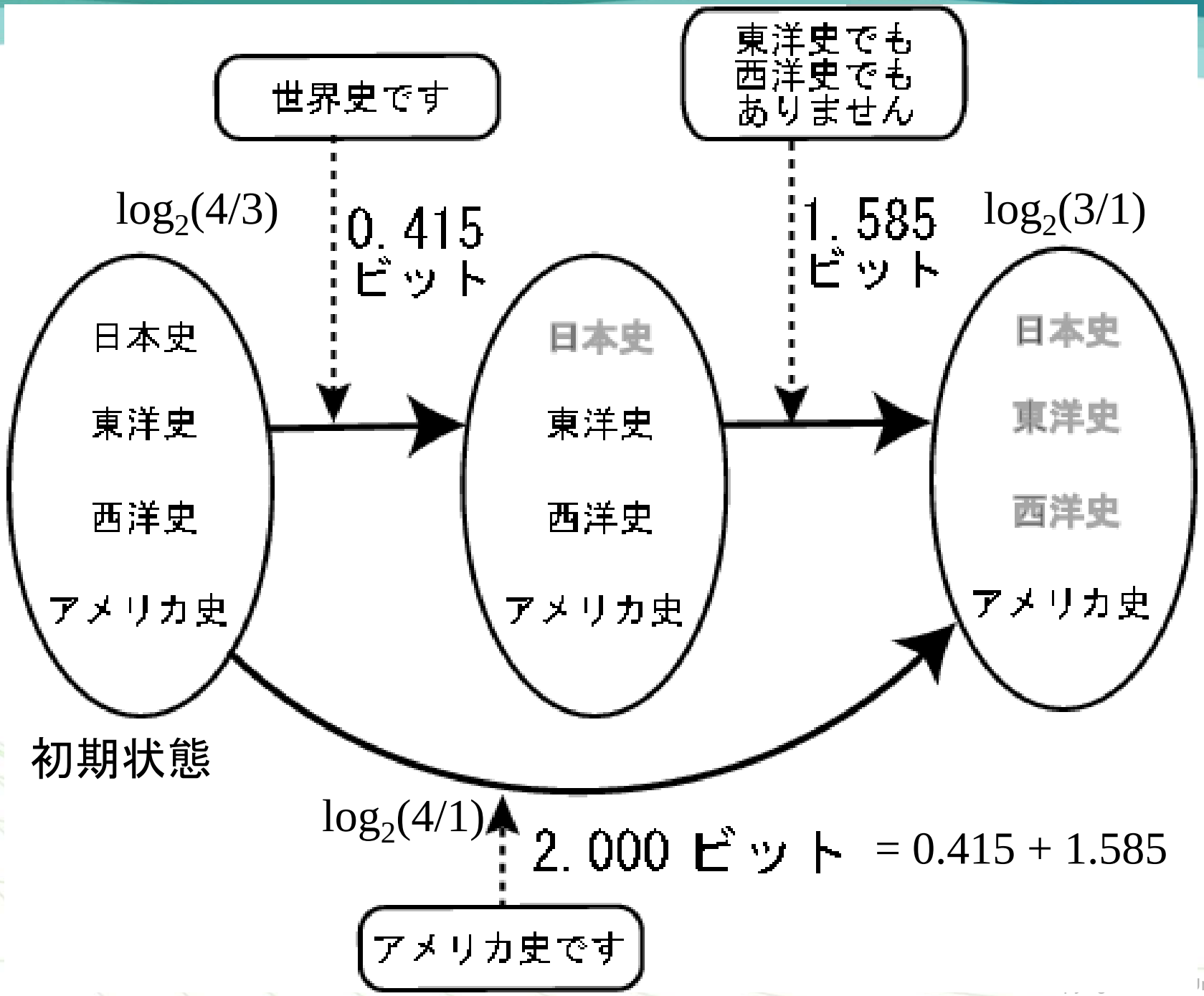
情報量の加法性

- **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
    - A x B = Cのとき, logA + logB = logC

情報量の加法性の確認


世界史です
東洋史でも
西洋史でも
ありません
log2(4/3)
0.415
ビット
1.585
ビット
log2(3/1)
日本史
東洋史
西洋史
アメリカ史
日本史
東洋史
西洋史
アメリカ史
日本史
東洋史
西洋史
アメリカ史
初期状態
log2(4/1)
2.000 ビット = 0.415 + 1.585
アメリカ史です

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，確率の低い事象が発生すると知らされた方が嬉しい！**
  - ＝価値が高い！

確率に基づく情報量の定義

これはニュースになる.
なぜならば,
確率が低い.
＝情報量が多い.

- 定義: -log2(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　　$-\log_2(1.0)=0$　　以下同様
    - 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

試験の例：確率から情報量を求める

- 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」＝−log2(3/4)＝0.4150
  - 「日本史が出題される」＝−log2(1/4)＝2.0

本日の出席表

- 身近な話題について，以下のような例を見つけなさい．
  - 情報を与えられる前は，いくつかの事象が発生することが予想されているが，各々の発生確率は異なることは知っている．．．と言う条件を満たす事象の集合．（例：日本史，または世界史が出題されることは分かっている．）
- その例について，与えられる情報の情報量を計算して示しなさい．
- その例の平均情報量(3.1.3)を求めなさい．
  - ただし，平均情報量＝Σ（確率i X 情報量i）

情報量の差異の応用：符号化と情報量 (3.1.4)

- 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 X 確率i))／記号の数

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

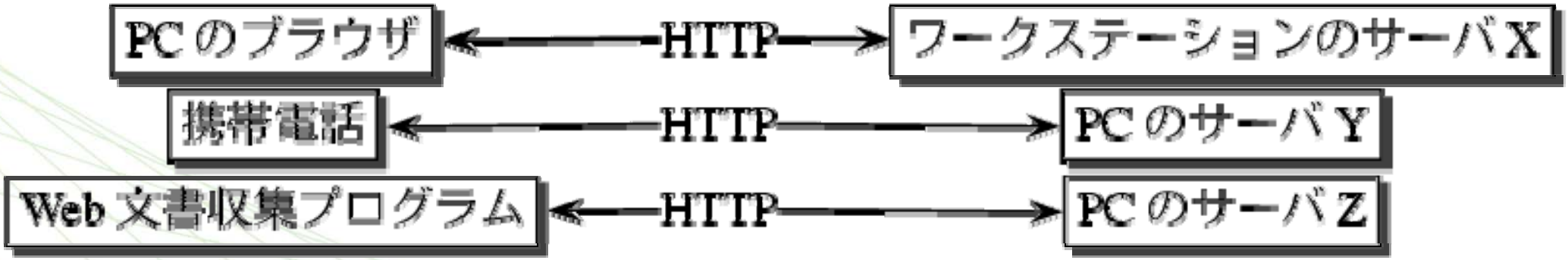
- **平均符号長＝0.844**
  - 1年分の試験の情報を表す符号長(=1)より平均符号長が短くなっている！

3.2 情報通信：：確実に，安全に情報を伝える

- **プロトコル（3.2.2）(a)：確実に**
  - 通信の際の決めごと
- **通信の秘密と相手の認証（3.2.3）(a)：安全に**
  - 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)

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）：Hyper Text Transfer Protocol
  - 電子メイル（SMTP）：Simple Mail Transfer Protocol
- プロトコルを正しく使える機器ならば，通信が可能

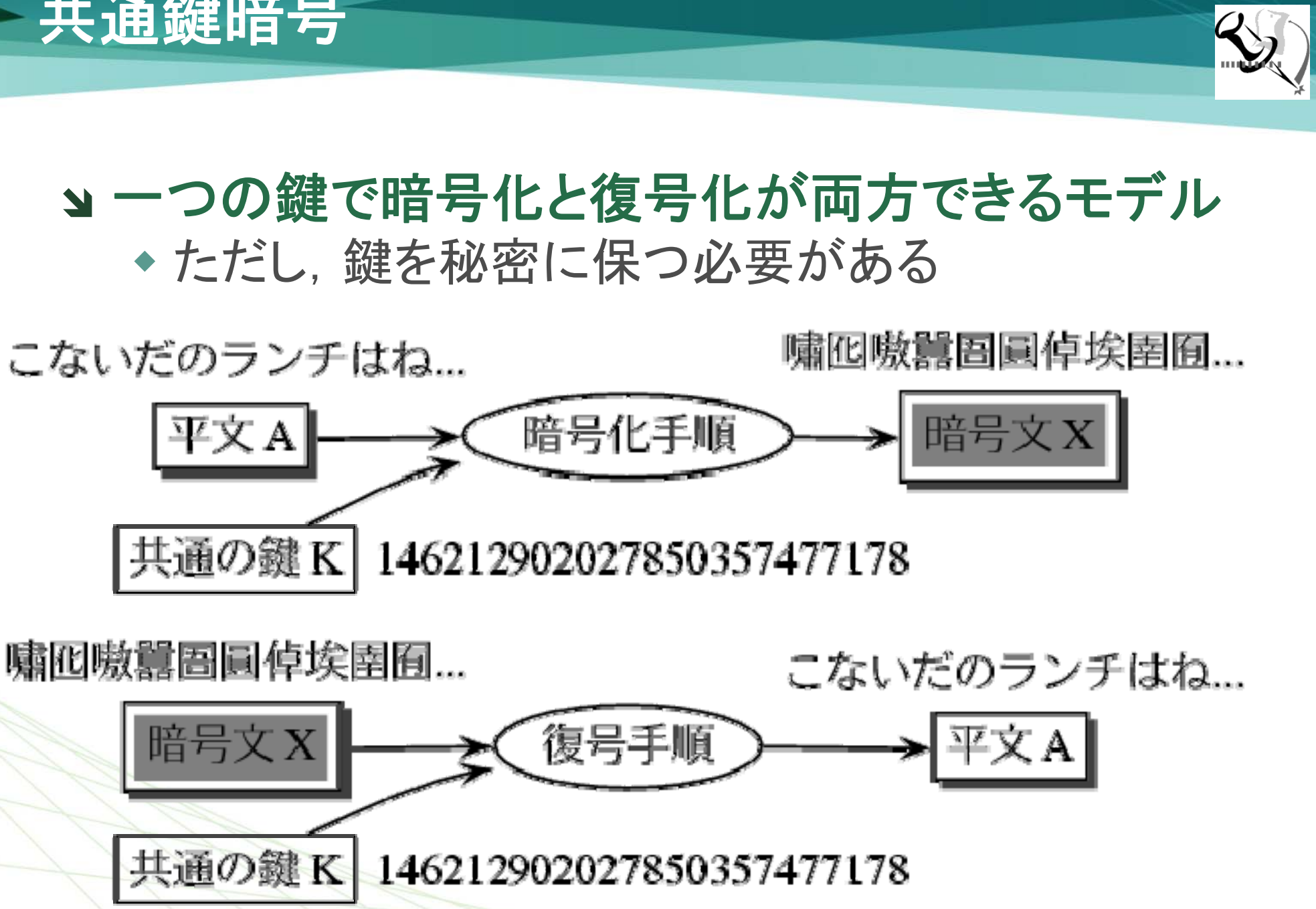


Webの相互運用性

3.2.3 通信の秘密と相手の認証

- 目的：正しい相手だけに，正しい内容を．正しい相手からのみ，正しい内容だけを．
- 平文（ひらぶん）：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。
  - 「嘯囮噉囂圄圓倬埃圉囿…」
- 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

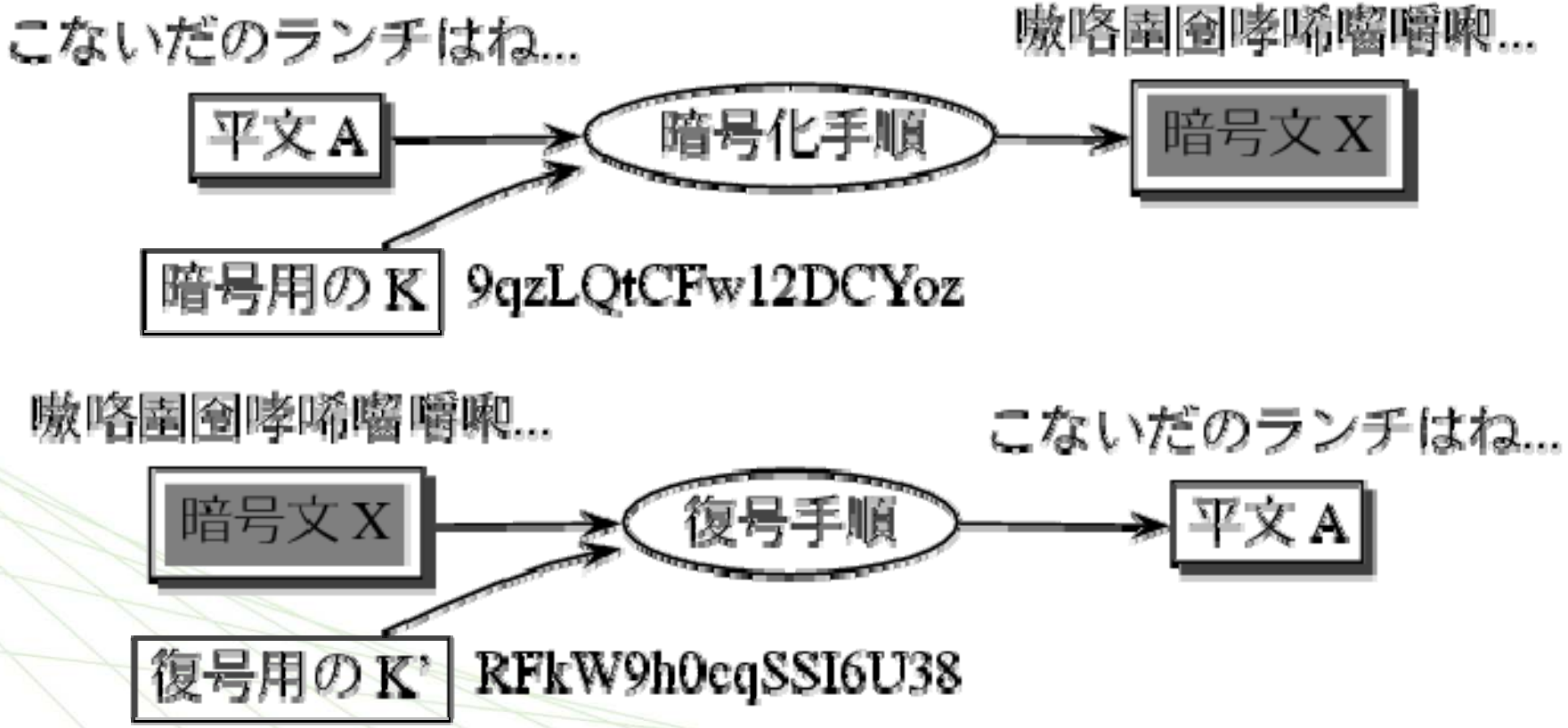


- 一つの鍵で暗号化と復号化が両方できるモデル
  - ただし，鍵を秘密に保つ必要がある

公開鍵暗号

暗号化と復号化の鍵が異なるモデル

- 鍵の組みを準備し，一方を公開し（公開鍵），他方を秘密にしておく（秘密鍵）

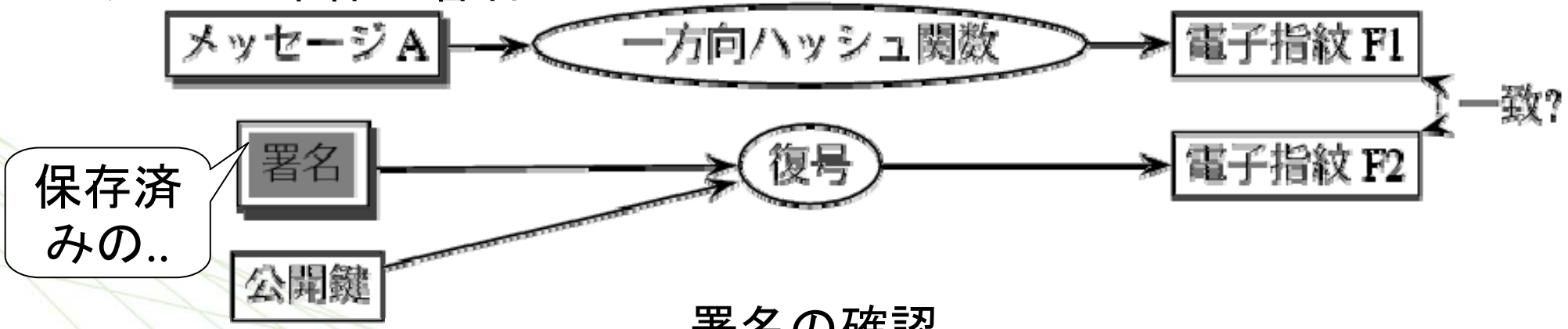


デジタル署名と検証

メッセージ本体＋署名

メッセージに対する署名の獲得

メッセージ本体＋署名



署名の確認

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

- 仮想的な通信路を確保，占有．通信速度一定保証．
  通信路が確保できなければ通信×

パケット交換

- データを細かく分け順番に，指定した宛先と通信．
  個々の通信専用の通信路を占有することはない．

回線交換　　パケット交換

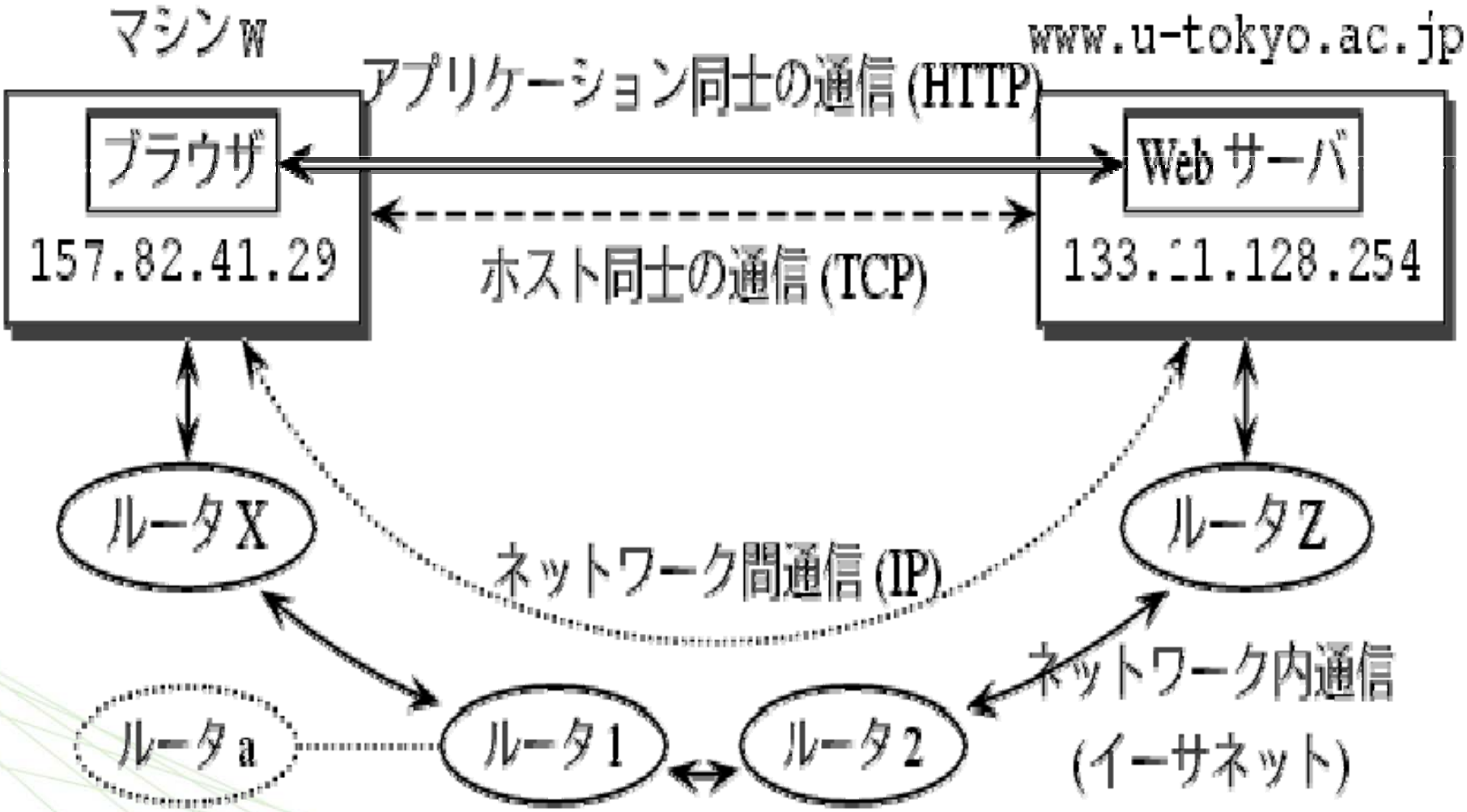
3.4 インターネット

- 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）

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
  - jp, ac, u-tokyo
  - 157, 82, 41....133, 11, 128, ...など
- 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(メール, ファイル転送, ウェブ情報通信など)毎に定義されている

Webブラウザにおける通信の行われ方(1)



Webブラウザにおける通信の行われ方(2)

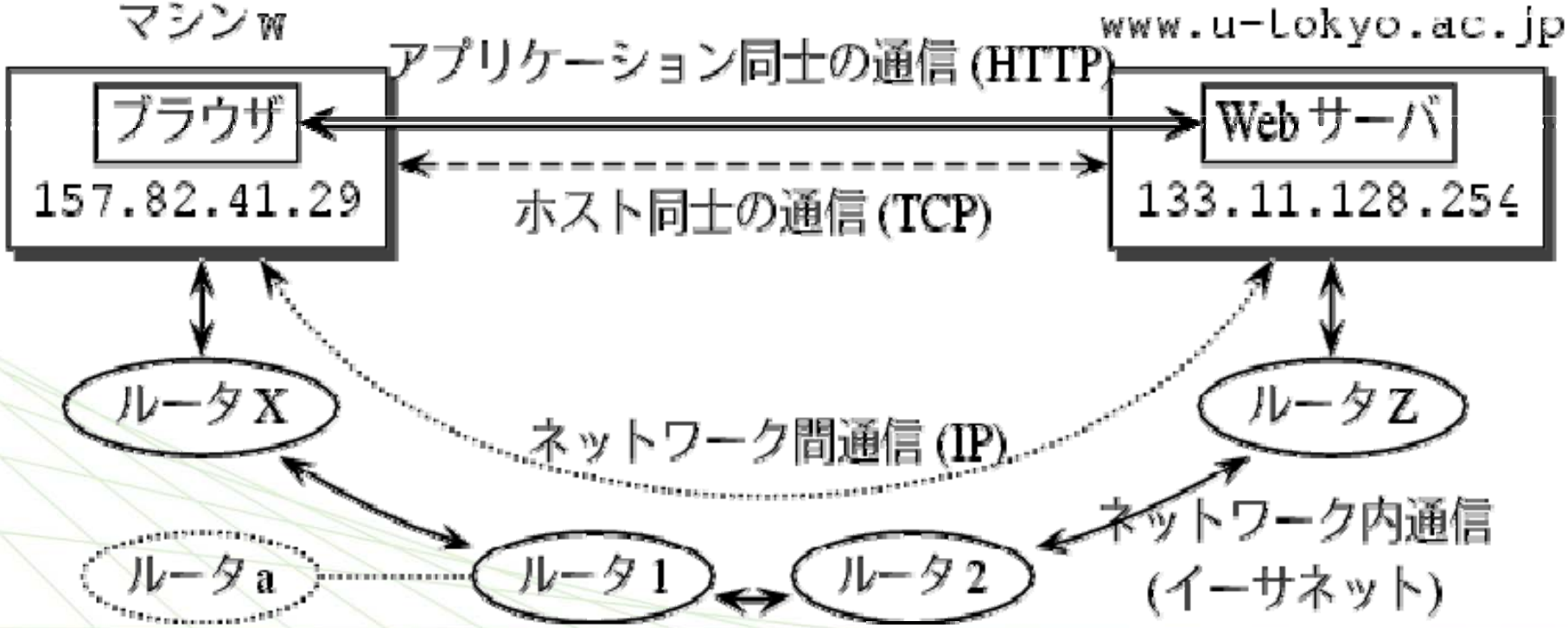
- ブラウザがURL(http://www.u-tokyo.ac.jp)から相手先のIPアドレス(133.11.128.254)を調べる.
- ブラウザが宛先のIPアドレスのウェブサーバに対して,HTTPのメッセージを送る.
  1. 送信元マシンWがメッセージをパケットに分割する.
  2. マシンWがパケット毎に宛先のIPアドレス(133.11.128.254)に送信する.
  3. 各パケットは,マシンWと同一ネットワーク内にあるルータXに届けられる.
  4. ルータXは宛先に至るまでのルータを次々と選びながら,最終的にルータZを介してWebサーバまで届けられる.
  5. パケットを受信したマシンは,パケットを元通りの順番に並べて,HTTPメッセージを取り出し,ウェブサーバに渡す.

階層プロトコル（例）

- 全てのデータ種別，利用目的に備えたプロトコルを準備するのは困難
  - 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：通信そのものに必要なプロトコルと，ゲームなどの利用目的に対応するプロトコルを分離
  - 場合に応じて組み合わせ可能に

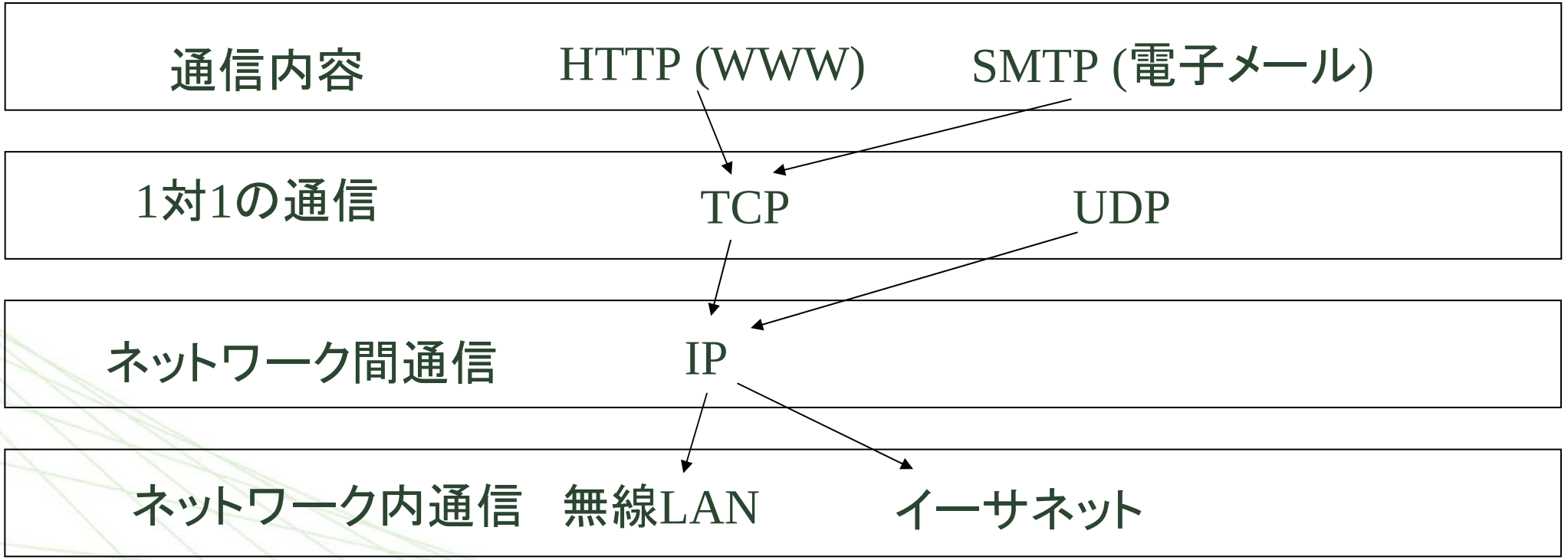
インターネットの場合

- 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



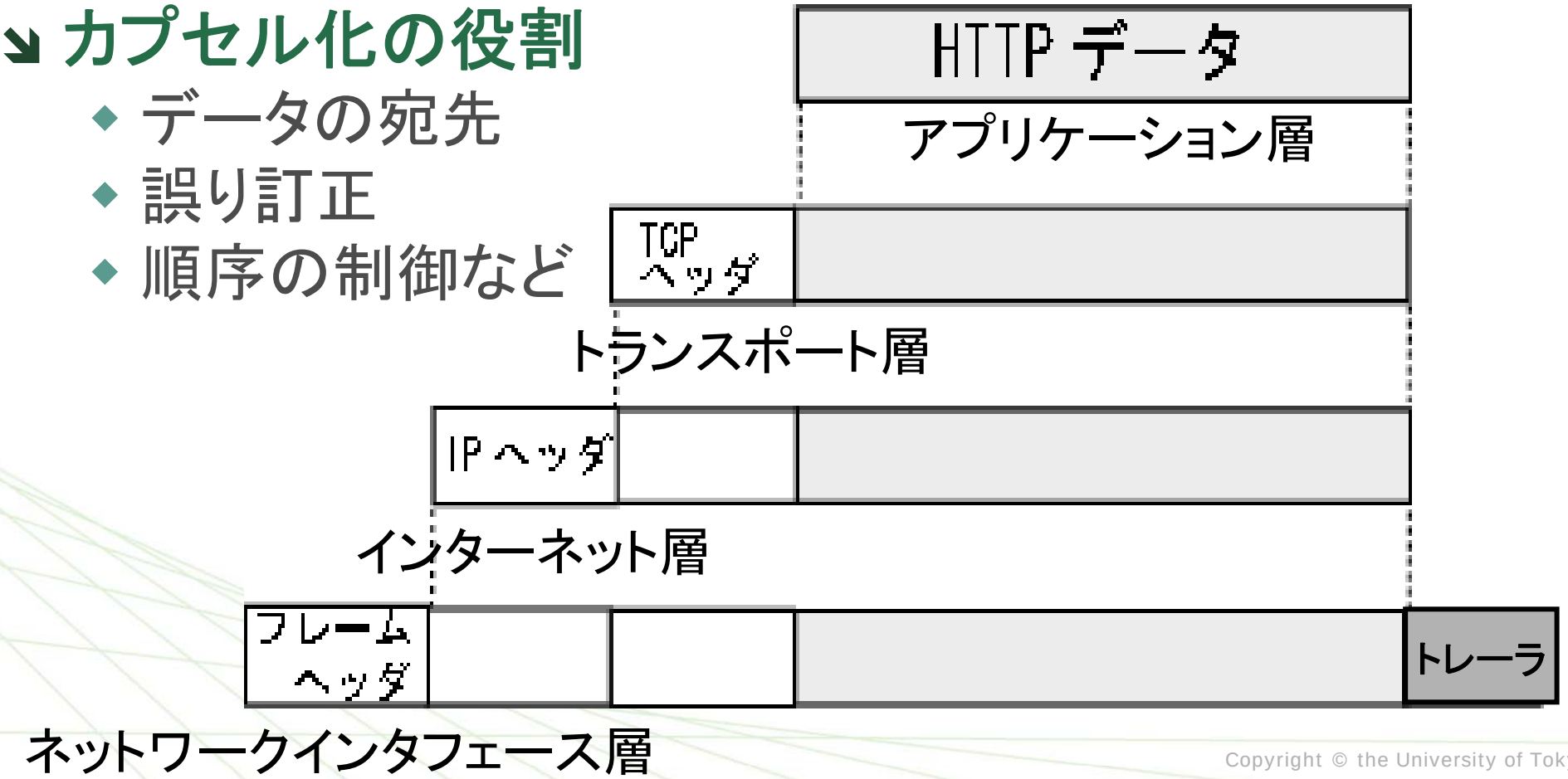
TCP / IP 階層プロトコル

- 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



カプセル化による階層プロトコルの実現

- 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
- カプセル化の役割
  - データの宛先
  - 誤り訂正
  - 順序の制御など



IP(Internet Protocol)アドレスとポート番号

- **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
    → 全世界で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
  - 192.168.. . はローカル(内部ネットワークでのみ有効で，内部ネットワークでのみ一意な)IPアドレス.
- **ポート番号: 同じホストの複数の通信を区別**
  - ポート:8080. . . httpのポート

例

- ネットワーク 172.16.0.0/16の中のホスト

172	.	16	.	2	.	10
10101100		00010000		00000010		00001010
ネットワーク番号				ホスト番号		

本日の出席票

- TCP(アプリケーション間通信)
  - Transmission Control Protocolの略
  - 接続を確立し，データを分割／組立て，確認応答(ACK)を行う.
- 以下の用語を上記の例に則って説明せよ.
- DNS
  - Domain Name Server の略
- DHCP
  - Dynamic Host Configuration Protocolの略
- MACアドレス
  - Media Access Controlの略

５／１６（来週）は１３１１