東京大学工学部 電気系学科
学生実験テキスト
IP ネットワークアーキテクチャ
第三版

江崎研究室

2015 年 10 月 1 日

1　目的

ネットワークアーキテクチャの一つである IP ネットワークは最も大規模に普及している実装であると言っても過言ではない。本実験では、その実装に対し理論を基礎とした実習を行うことで、実用的なネットワークアーキテクチャとしての知見を深め、(IP に限らず) 通信に必要な要素と、その要素の依存関係を身につける。

2　はじめに

”インターネット”や”IP アドレス”という単語について、その名称は知られていても、どのような手順で通信が成立し、ネットワークが分散的に運用され、相互接続の結果、全体として世界規模の実用的な大規模ネットワークサービスが提供されているかが認識されていない事が多い。

本実験では、以下の項目を中心に、IP ネットワークアーキテクチャの基本要素を確認し、最終的には、自らが利用しているインターネットの構造を知り、説明が出来る水準に達する事を目標とする。

2.1　アドレッシングと経路表

IP ネットワークアーキテクチャの基礎であるアドレッシング、ルーティングについて理解する。(IP に限らず) ネットワークアーキテクチャは、エンドホストとその間に存在するルータによって構成される。

特に IP ネットワークアーキテクチャでは、ネットワーク、ルータおよびホストに付与された識別子と、ルータおよびホストが持つ経路表により、逐次的にデータ転送時の経路選択が行なわれ、通信が成立することを確認する。

2.2 経路制御アーキテクチャ

経路制御 (IP ルーティング) の方法には、”静的な制御方法”と”動的な制御方法”が存在する事を知り、動的経路制御法により、IP ネットワークが、”単一故障点が回避できる事”と”スケーラビリティが与えられる事” を理解する。

2.3 DNS

IP ネットワークアーキテクチャには、IP アドレスとは異なる構造を持った DNS と呼ばれる識別子抽象化システムが存在し、ユーザの利便性を高めていることを知る。また、DNS は、事実上、世界規模で実現している唯一の大規模分散データベースであることも理解する。

3 実験環境

本実験では複数人でチームをつくり、身近な計算機 (PC) を使って、基本的なネットワークトポロジを組み上げる。普段利用している PC が、インターネットでのエンドホストとなるだけでなく、設定によっては、インターネットの内部で利用されているルータの機能を持つことを確認する。このため、学生が所持している学科貸与の PC を利用することを前提とする。事前準備として、所持している PC において、Linux(ubuntu) が起動できるようにさせ、なおかつアップデートを行い、最新の状態にしておくこと。 PC の接続に利用するネットワークケーブルは実験 TA が部材を用意するので、自身で作成すること。実験最終日には、それまでに確認した要素技術を組み合わせ、ネットワーク相互接続実験を行う。そのため、 最終日前までに、本テキストの内容を理解した上で、実験計画(特にトポロジ、検証事項およびその方法)を明確にし、最終実験に備えること。その他、時間に余裕があれば、実際に大規模ネットワークで利用されている製品ベースのルータおよびスイッチも、学生自身で設定する機会を設ける。

4　アドレッシングと経路表

4.1　通信の基本要素

通信の基盤となるネットワークは、ホストとルータで構成されている。ホストは通信データの発信源または終着点であり、ルータはそのデータを転送させる機能をもつ機器である。通信を行う際は、これらの機器が一定のルールの元、データの送信、受信、転送を行っている。通信では、このルールを「プロトコル (Protocol)」と呼んでいる。インターネットでは、その名の通り、「インターネットプロトコル（IP:Internet Protocol)」が利用されている。

4.2　internet と The Internet

あるネットワークと別のネットワークが相互接続されたネットワークをインターネット (internet) と呼ぶ。internet の inter とは、相互接続 (interconnect) しているという意味を示すためである。一方で、我々が日常的にメール送信をしたり、Web 閲覧をするネットワークもインターネットと呼ばれている。このインターネットは、IP により、世界規模で相互接続がされた唯一のネットワークであるため、上記の internet” と区別して、The Internet” と呼ばれることもある。

インターネットは IP パケットと呼ばれるデータ単位に、細切れになってデータが送信されている。このようなデータ転送方式をパケット交換方式と呼ぶ。

- 課題
  - パケット交換と比較される通信方式には回線交換がある。それぞれどのような長所、短所があるかを考察せよ。
  - 以降、各自の PC のネットワーク接続を、有線接続で行う際には、UTP ケーブルを利用する。TA の指示に従い、自身で利用するケーブルを作成し、その品質を測定せよ。一定基準を満たさないケーブルは利用できない。基準を満たすよう注意して作成すること。

4.3　IP アドレス

パケット交換方式では、通信の発信源、および終着点以外のホストもネットワークに接続されることを許容することから、ネットワーク回線が共用できるという特徴がある。このため、IP では、すべての IP パケットについて、どのホスト (Source Host) から送出され、どのホスト (Destination Host) に向かって転送されるかの記述を定めている。この時にホストを示すために利用される識別子が IP アドレスである。図 1 に、IP で定義されたパケットフォーマットを記す。

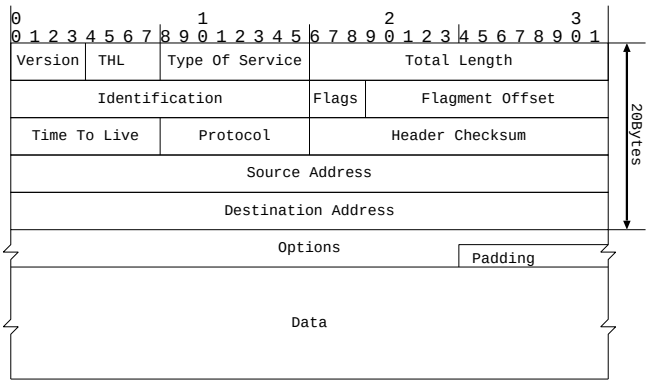

図 1: IP パケット (IPv4) のフォーマット

IP アドレスは、固定長のアドレス空間を利用している。IP バージョン 4 では、32 ビットの空間であるため、

00000000000000000000000000000000

から、

11111111111111111111111111111111

の範囲のアドレスを利用している事になる。しかし、このままでは、人間にとっての表記が困難であるため、8 ビットずつに区切り、それぞれの区切りの中で、10 進数で表記している。例えば、2 進数にて、

11000000.10101000.01100100.00000001

という表記で表される IP アドレスは、

$$
\begin{aligned}
11000000 &= 2^7 + 2^6 = 128 + 64 = 192 \\
10101000 &= 2^7 + 2^5 + 2^3 = 128 + 32 + 8 = 168 \\
01100100 &= 2^6 + 2^5 + 2^2 = 32 + 16 + 4 = 100 \\
00000001 &= 2^0 = 1
\end{aligned}
$$

より、10 進数表記では、192.168.100.1 となる。つまり、利用可能な IP アドレスの範囲を 10 進数で表記すると、0.0.0.0 から、255.255.255.255 になる。以下、IPv4 を利用した環境について説明する。

ルータはネットワーク間でパケットの転送をするため、複数のネットワークに接続されている。それぞれのネットワークをサブネットと呼び、そのネットワークに収容できるホストの数はネットマスクで定義される。[1] ネットマスクは複数の IP アドレスに対して、それぞれを 2 進数表現にした際に、最高位ビットからどのくらいの長さが同一であるかを示したものである。言い換えれば、複数の IP アドレスに対し、2 進表現で、ネットマスクと論理積演算 (&) を行った結果が同一であれば、これらの IP アドレスはそのネットマスクを利用するネットワーク内に存在すると考える。

[1] サブネットマスクと呼ぶこともある

IPv4 の場合は、アドレス長が、32 ビットであるため、ネットマスクは、0 から 32 が利用される。[2]

例えば、192.168.100.1 の IP アドレスを含む、24 ビットのネットマスクを利用するネットワーク内に存在する IP アドレスの範囲を考える。

$$\begin{aligned} 192.168.100.1 &= 11000000.10101000.01100100.00000001 \\ 24\text{ ビットマスク} &= 11111111.11111111.11111111.00000000 \end{aligned}$$

より、192.168.100. までが同一の IP アドレスである。つまり、このネットワークには 192.168.100.0 から、192.168.100.255 までの IP アドレスが存在する。これより、24 ビットのネットマスクを持つネットワークに 192.168.100.1 が存在する場合、IP アドレスの後ろに/" とネットマスクを同時に表記し、192.168.100.1/24 と記す。

また、ネットマスクを、10 進表現で示す場合もあり、24 ビットマスクの場合は、255.255.255.0 となる。

- 課題
  - それぞれの IP アドレスが同一ネットワークに存在するか計算し、確認せよ。
    - 192.168.10.130 と 192.168.10.150 は、同一の 28 ビットネットワークに存在するか
    - 10.100.5.1 と 10.100.62.180 は、同一の 18 ビットネットワークに存在するか
  - 本実験で対称とする Linux システムでは、ネットワーク関連の設定および確認を"ip" コマンドで行う。特にアドレスの設定および確認は "ip addr" コマンドで行う。[3] 各自の PC について、インターネットに接続する前後で "ip addr" を行い、違いを調べよ。[4]
  - IP アドレスの自動設定を停止させ、手動にて IP アドレスを設定せよ。

[2] ネットマスクはマスクをかける長さでもあるため、ネットマスク長と呼ばれる場合もある。

[3] 以前は Linux システムでの IP アドレスの設定および確認を"ifconfig"コマンドにて行う事が主流であった。

[4] 以降、Linux システムでのコマンドを実行する際は、事前に"man コマンド名" にて、どのようなプログラムなのかと、実行に必要なオプションを確認すること。また、コマンド自体がエラーになる場合は、プログラム自体がインストールされていない可能性がある。そのような場合は、エラー時のメッセージを参考にすること。必要なプログラム（パッケージ）は、コマンドラインにて、apt-get update && apt-get install パッケージ名" により大抵の場合はインストールが出来る。

4.4 経路表

ホストが IP パケットを送出するとき、およびルータが IP パケットを転送するときに参照される宛先データベースを経路表[5] と呼ぶ。

ホストおよびルータがネットワークを構成し、通信ができるようになるためには、

- それぞれの機器に正しい IP アドレスが設定され、
- それぞれの機器で正しい経路表が設定されている

ことが必要となる。

Linux システムの場合、経路表の参照および確認は、ip route” コマンドで行う。[6]

図 2 の構成の場合、ホスト A およびルータ B の IP アドレス、経路表は、下記の状態で通信が可能となる。

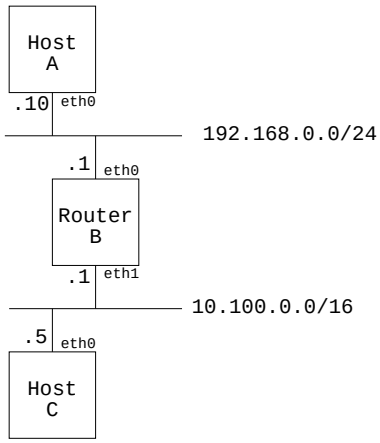

図 2: ホスト 2 台とルータ 1 台により構成されたネットワーク

```
$ ip addr show dev eth0
2: eth0: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc pfifo_fast state UP
    link/ether 00:16:3e:78:fd:01 brd ff:ff:ff:ff:ff:ff
    inet 192.168.0.10/24 brd 192.168.0.255 scope global eth0
       valid_lft forever preferred_lft forever
```

図 3: ホスト A の IP アドレス

```
$ip route show
192.168.0.0/24 dev eth0  proto kernel  scope link  src 192.168.0.10
10.100.0.0/16 via 192.168.0.1 dev eth0
```

図 4: ホスト A の経路表

[5]ルーティングテーブルと呼ぶ場合もある

[6]以前は Linux システムでの経路表の参照および確認は”route”コマンドにて行う事が主流であった。

```
$ ip addr show dev eth0
2: eth0: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc pfifo_fast state UP
    link/ether 00:16:3e:88:22:12 brd ff:ff:ff:ff:ff:ff
    inet 192.168.0.1/24 brd 192.168.0.255 scope global eth0
       valid_lft forever preferred_lft forever

$ ip addr show dev eth1
3: eth1: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc pfifo_fast state UP
    link/ether 00:16:3e:8a:d9:1c brd ff:ff:ff:ff:ff:ff
    inet 10.100.0.1/16 brd 10.100.255.255 scope global eth1
       valid_lft forever preferred_lft forever
```

図 5: ルータ B の IP アドレス

```
$ip route show
192.168.0.0/24 dev eth0  proto kernel  scope link  src 192.168.0.1
10.100.0.0/16 dev eth1 proto kernel scope link src 10.100.0.1
```

図 6: ルータ B の経路表

ホスト A の経路表 (図 4) を参照しよう。

まず 2 行目では、10.100.0.0/16 のネットワークへのパケット転送では、192.168.0.1 のルータを経由する (via) 経路を利用することを示している。[7] また”dev eth0” より、その際に利用するネットワークインターフェースが、eth0 である事が分かる。

次に、1 行目では、via” の表示がないことから、192.168.0.0/24 のネットワークへのパケット転送では、ルータを経由しない経路となることが分かる。つまり対象ネットワークがネットワークインターフェース (eth0) に直接接続されている事を示している。

- 課題
  - ホスト C の IP アドレスは、10.100.0.5/16 である。ホスト C の経路表はどのように定義されているか答えよ。
  - 各自の PC を IP アドレス自動設定にしてインターネットに接続せよ。その際の経路表を確認し、各項目の意味を説明せよ。特に、default として記載される経路が記載される場合がある。default は 0.0.0.0/0 と同値である。これは何を意味するか、ネットマスクに注目し考察せよ。

4.5 運用管理

IP ネットワークを運用および管理するためには、IP に関する知識だけでなく、その特性を活用した種々のツールを利用して、日々発生するトラブルへの対応や、利用傾向を調査し、機器やリンク、トポロジの変更を行っていかなければならない。本章では、基本的なツールの紹介を行う。

[7] ルータの別称としてゲートウェイを用いる場合もある。

4.5.1 ping

IP では、管理のために作成された ICMP(Internet Control Message Protocol) と呼ばれるプロトコルがある。あるホスト間での、到達性を確認するためには、ICMP で定義された、ICMP echo および ICMP echo reply が利用できる。これを利用したプログラムが ping である。ping -n 宛先 IP アドレス” で、宛先アドレス宛の ping が実行できる。

- 課題
  - 図 2 の構成を作成せよ。ホストおよびルータから ping にて接続された他の機器への到達性を確認すること。なお、PC をルータとして動作させる場合には、経路表を正しく設定した後、

    `sysctl -w net.ipv4.ip_forward=1’’`

    のコマンドを実行し、IP パケット転送機能を有効にすればよい。

4.5.2 tcpdump

tcpdump はパケットキャプチャツールである。tcpdump を実行すると、実行した機器のインターフェースに到着したパケットを表示する。運用管理においては、ping 等のツールを使った際に表示される結果だけを当てにしてはならない。もしツールを実行し、エラーが発生したならば、どのような原因により、そのエラーが発生したのか、適宜パケットキャプチャを行い、どのような通信がされているかを確認する必要がある。

- 課題
  - 図 2 の構成にて、ping を行いながら、同時に tcpdump を実行させ、ping コマンドにより生成されたパケットがどのようなものかを確認せよ。また、ping 実行時のオプションを変化させ、パケットにどのような違いがでるかを確認せよ。
  - tcpdump をグラフィカルに実行できる強力なツールとして、wireshark がある。wireshark を実行し、tcpdump と同等の結果が表示される事を確認せよ。
  - wireshark 実行時に、宛先アドレス (dst) を限定するフィルタ、送信元アドレス (src) を限定するフィルタを適用し、結果を確認せよ。
  - フィルタは tcpdump でも適用できる。man を参照し、wireshark で適用したフィルタと同様のフィルタを適用せよ。

4.5.3 traceroute

ping では、到達性の有無のみしか確認できない。あるホストに対して、到達性が無いという問題が発生した場合、図 2 に示すような簡単な構成のネットワークや、どのような構成になっているかが把握できている場合ならば、ping を用いて、到達できる場所と到達できない場所を探し出し、問題点を探し出す事ができる。しかし、インターネットは相互接続ネットワークであり、複雑な構成であったり、自分が管理する範囲外にて、構成変更がされることも有り得る。そのような場合に役に立つツールが、traceroute である。

traceroute は、IP パケットヘッダに定義された TTL(Time To Live) 値を利用し、送信ホストから受信ホストまでの通過経路中に存在する各ルータからの ICMP パケットによって、経路の表示を行う。

- 課題
  - 図 2 の構成にさらに 1 台ルータを追加し、ホスト- ルータ-ルータ-ホストとなる構成を作成せよ。そして ping で各機器から他の機器への到達性を確認せよ。
  - 上記の構成にて、ルータがそれぞれ 3 つのネットワークを収容する構成を作成せよ。そして到達性の確認をせよ。
  - 上記の構成で、各ホスト、ルータにおいて traceroute を実行せよ。また適宜宛先を変更して試すこと。
  - traceroute を実行している際に、wireshark を実行し、TTL 値がどのように利用されているかを述べよ。
- 発展課題
  - IP ネットワークは OSI 定義でレイヤ 3 の階層に定義されている。その下層で代表的に利用される技術にイーサネット (Ethernet) が存在する。イーサネットと IP の関係を調べよ。また、実際にどのように利用されているかを、wireshark を用いて調査せよ。なお、arp,route の各コマンドがキーワードである。
  - 同一ネットワーク以外のネットワーク（ルータを経由した先にあるネットワーク）に存在するホスト宛のパケットは、そのホストの IP アドレスが宛先 IP アドレスとして記されている。IP ヘッダには 宛先ホストとして IP アドレスは一つしか記述できないが、どのようにしてパケットがルータを経由するか、調査せよ。

5　経路制御アーキテクチャ

5.1　経路の集約

IP アドレスは連続した、アドレス空間を利用しているため、経路表で利用する経路において、隣接するアドレス範囲を示す経路においては、ネットマスクを利用して、複数のエントリを一つのエントリに集約することができる。例えば、192.168.100.0/25 と、192.168.100.128/25 が同一のルータに向かう場合は、192.168.100.0/24 として表記することができる。これを経路の集約という。

- 課題
  - 1,2 日目の最後の課題で作成した構成について、アドレスアサインを工夫し、最低一つのルータでは、経路集約ができるようにせよ。

5.2　最長一致 (ロンゲストマッチ) の原則

経路表を参照する際、宛先アドレスに対し、複数の項目に該当 (マッチ) する場合がある。例えば、図 7 の経路表では、10.100.10.10 宛の経路では (*A)(*B) の 2 つの経路が該当する。

```
$ip route show
192.168.0.0/24 dev eth0  proto kernel  scope link  src 192.168.0.1
192.168.10.0/24 dev eth1  proto kernel  scope link  src 192.168.10.1
10.100.10.0/24 via 192.168.0.1 dev eth0                                (*A)
10.100.0.0/16 via 192.168.10.1 dev eth1                                (*B)
```

図 7: 複数の経路が該当する場合

このような状態では、ネットマスクが長い経路が優先される。これを最長一致 (ロンゲストマッチ:longest match) の原則という。上記の場合では、経路エントリ 1 行目の eth0 宛となる。

- 課題
  - 図 7 の経路表を含む構成を作成せよ。図 7 の経路表を持つルータにて、

    `ip route show` 宛先アドレス

    および

    `ip route show match` 宛先アドレス

    のコマンドを実施し、結果が異なることを確認せよ。
  - 図 7 の経路表を含む構成にて、宛先アドレスと送信元アドレスに注意し、ping および traceroute を実施せよ。その上で、各インターフェースで同時に tcpdump を実施し、途中経路で最長一致が実現されることを確認せよ。

5.3 経路の対称性

先の課題までで確認したように、ある宛先までの経路は、その経路中のルータによって、逐次的に次の宛先が決まり、決定していく。これは、つまりホスト A からホスト B までの経路と、ホスト B からホスト A までの経路が必ずしも一致しない事を示している。インターネットでの通信形態から分かるように、クライアントからサーバに、リクエスト（少量のパケット）を送って、サーバからクライアントまでデータ（大量のパケット）が返ってくることもある。このような非対称なデータ転送量から、ネットワーク管理者が意図的に経路を非対称にすることもあれば、パケットが複数のネットワークを横断する際には、それぞれのネットワークの管理方針（ポリシ）の違いから、非対称な経路となることもある。

- 課題
  - インターネット中には、研究や運用管理に活用するため、任意のホストに対し、ping や traceroute を行うインターフェースを公開しているサーバ (router server) がある。web 検索のキーワードで、looking glass” および”traceroute” を用いて、その公開サーバを探しだし、自分の PC とそのサーバ間、および複数のサーバ間の traceroute を実行し、対称性を確認せよ。

5.4 動的経路制御

前述までの課題では、ルータを手動で設定し、ネットワークを構築した。この手法を静的経路制御法 (スタティックルーティング) という。しかし、この手法には大きな欠点がある。この手法を用い、ネットワークを拡張し、ルータを追加した場合は、既存のルータ群に対し、該当経路を追加する必要がある。ルータ数台規模の小規模ネットワークならば、手動での設定でも管理が行き届くが、ルータの数が増加し、10 台、50 台となった場合の手動設定は、管理コストが高すぎ、人間の手に負えない、管理不十分のネットワークになってしまうという問題がある。

また、耐障害性が低いという問題もある。例えば、図 8 の、物理的にリング状に構成されたネットワークにおいて、ホスト A ⇒ ルータ B ⇒ ルータ C ⇒ ルータ E ⇒ ホスト F” の経路があるとする。この場合に、故障や誤設定に起因して、ルータ C が使えなくなると、ホスト A - ルータ B - ルータ D - ルータ E - ホスト F” の繋がりがあっても、その経路が無いためにパケットが届かなくなる。

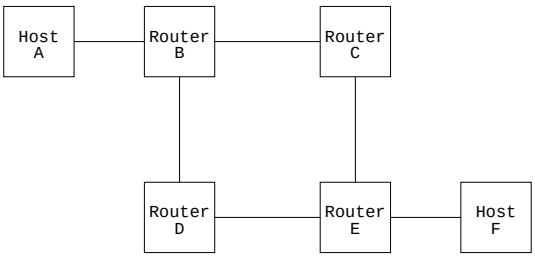

図 8: ホスト 2 台とルータ 4 台により構成されたネットワーク

このような問題を解消するために提案された技術が動的経路制御法 (ダイナミックルーティング) である。

動的経路制御法の基本的な機能は、下記である。

- ルータは自身が直接収容するネットワークの経路を他のルータに教える（経路広告）
- ルータは他ルータから教えられた経路を他のルータに伝える（経路配布）

上記の機能を実現する際に、ルータ間でのやりとりが必要になる。この際に利用されるプロトコルをルーティングプロトコルという。

- 課題
  - 図 8 において、静的経路制御法を用い、ホスト A から送信されたパケットがループした経路を通過するよう、各ルータの経路を設定せよ。その場合に、ホスト A からホスト F 宛、およびホスト F からホスト A 宛の ping および traceroute を実行した場合の結果を予想し、確認せよ。また同時に tcpdump も実行し、パケットの状態も確認せよ。なお、以降は特に指示が無い場合は、利用する IP アドレスを 192.168.0.0/16 の範囲から選択し、IP ルーティングが正しく動作するようにネットワーク設計を行うこと。
  - 上記の状態で、実験 TA の指示に従い、ルーティングプロトコルの一実装である RIP(Routing Information Protocol) を設定せよ。RIP の利用には、ルーティング用ソフトウェアである quagga を用いること。”router rip”内で、”timers basic 5 15 10”を実施し、タイマの値を変更しておくこと。なお quagga 動作中は、”ip monitor” コマンドを実施するターミナルを一つ用意し、ホスト内での経路登録の状況を監視すること。
  - RIP が動作している構成で、ルータ C またはルータ D を停止させ、経路の切り替わりを確認せよ。
  - RIP が動作している構成で、あるルータでのネットワークの増減、ネットワークアドレスの変更を行い、どのように情報が伝播するか、パケット転送にどのように影響があるかを確認せよ。なお、quagga にはデバッグモードも存在するので活用すること。
- 発展課題
  - RIP が動作しているネットワークにおいて、既に利用している IP アドレスを含むネットワークを異なるルータに設定し、経路広告を行うと何が発生するか予想し、実際に動作を確認せよ。
  - RIP の他の代表的な動的経路制御プロトコルには、OSPF(Open Shortest Path First) がある。プロトコルの動作時に利用されるアルゴリズムの違いを調べよ。また、このアルゴリズムに起因して、どのような長所、短所があるのか、違いを調べよ。
  - RIP で構成されたネットワークとは別に、同一トポロジで OSPF ネットワークを構築せよ。
  - RIP を利用するネットワークと OSPF を利用するネットワークを相互接続する場合は、プロトコルの違いのため、直接接続せず、両方のプロトコルが利用できるルータ準備し、通訳のように使う。この時に、経路は異なるプロトコルに配布されるため、「経路が再配布 (redistribute)) される」、という。RIP と OSPF を接続し、再配布の設定をせよ。

6 DNS

IP アドレスは数字の羅列であるため、アドレスとホストの対応関係を人間が把握するのは難しい。特に多くの IP アドレスを扱うようになると、識別が困難になってくる。このため、人間が覚えやすい文字列 (名前) と IP アドレスのマッピングを行う仕組みが開発された。これが DNS(Domain Name System) である。例えば、電気系の web ページを扱うホストの名前は、

www.ee.t.u-tokyo.ac.jp

であるが、DNS を用いてこの IP アドレスを検索すると、

157.82.13.244

が得られる。このとき実行される検索は名前解決と呼ばれる。

6.1 ドメイン

名前表記は、いくつかの.”（ドット）で区分けされた部分により構成される。この表現は名前で利用される範囲” を示しており、これをドメインと呼ぶ。あるドメインについて、その部分集合をサブドメインと呼ぶ。ドメインは表記中の最も右の部分が最上位となり、階層が深くなるにつれて、左側に表記していく。

例えば、www.ee.t.u-tokyo.ac.jp は、.jp” ドメインの一部であり、また、.ac.jp” ドメインの一部でもある。ac.jp ドメインは jp ドメインのサブドメインである。

すべてのドメインは、.” というドメインの配下にサブドメインとして存在する。www.ee.t.u-tokyo.ac.jp というホスト名は、このルールに基づき正しい表記を行うと、www.ee.t.u-tokyo.ac.jp. となる。最も右に存在するべきの”.” を省略した、www.ee.t.u-tokyo.ac.jp のような書式は慣例的に許されている記法なのである。

DNS の構造において、最上位に存在するドメインは、トップレベルドメイン (TLD:Top Level Domain) と呼ばれ、.jp”、.com”、.net” のようなものがある。その次の階層のドメインは、セカンドレベルドメイン (SLD:Second Level Domain) と呼ばれ、.ac.jp”、.google.com”、.ntt.com” のようなものがある。それ以降のドメインは順次、第 3 レベルドメイン、第 4 レベルドメインのように呼ばれる。

DNS を用いた検索システムは、サーバクライアント方式で構成されている。つまり検索要求を発行するクライアントと、クライアントへの応答として、データ (ドメイン情報) を返すサーバ群である。これらのサーバは、インターネット中に分散しており、各サーバが提供するデータベースは、ドメインとサブドメインの関係より、ツリーを用いた階層構造状に連携している。

一つのホスト名を解決する場合にも、分散している各階層のデータを受け持つサーバを逐次的に検索して、目的のデータにたどり着く仕組みになっている。この時、各サーバが担当する階層データをゾーンと呼ぶ。[8]

[8] ゾーンは各サーバが保持する設定情報であり、ドメインは、”.”で区切られる階層情報であるため、ゾーンとドメインは異なる意味を持つ。

- 課題
  - インターネットに接続されたホストにて、web ブラウザを用いて、URL のホスト部の最も右に”.” を付与し、web ページ閲覧の際に、ブラウザの挙動が異なるかを確認せよ。また同様に ping やその他のアプリケーションを用いた場合の動作も確認せよ。

6.2 検索の仕組み

DNS の検索システムには、下記の 3 つの機能が登場する。

- リゾルバ (Resolver)

  リゾルバは、DNS 情報を必要とするホストで稼働し、キャッシュサーバに DNS 検索を依頼する。ホスト内では通常は、アプリケーションからの要求により、オペレーティングシステム内のリゾルバプログラムが実行される。このホストは主に末端の機器であり、普段我々が利用する PC 等がそれに該当する。なお、キャッシュサーバの IP アドレスは、IP アドレス自動設定時に自動的に設定されるか、ネットワーク管理者からの情報を元に、手動で設定する。

- キャッシュサーバ (Cache Server)

  キャッシュサーバは、リゾルバから依頼された DNS 検索について、自身のキャッシュにデータがあればそれを応答として返答する。もし該当データが無ければ、各コンテンツサーバに再帰的に検索をかける。検索結果は随時キャッシュとしてサーバ内に保存される。別名として、フルリゾルバと呼ばれる場合もある。専用サーバ上のアプリケーションとして動作することが多い。

- コンテンツサーバ (Contents Server)

  コンテンツサーバは、自身が所持するゾーン情報の中から、キャッシュサーバからの検索に応答する。キャッシュサーバと同様、専用サーバ上のアプリケーションとして動作することが多い。

[9]

例えば、リゾルバが www.ee.t.u-tokyo.ac.jp について IP アドレスの検索を行う際は、以下の手順になる。この様子を図 9 に示す。

[9] 6.1 節では、DNS 検索システムがサーバクライアント方式であると記したが、これは実装機能とデータ転送の関係を示したものである。つまり検索データのクエリとレスポンスの転送される方向が、クライアントからサーバにクエリ、サーバからクライアントにレスポンスと定まっている事を示す。一方で、実装機能が等しい中で、クエリとレスポンスが等しく転送される関係は、Peer-to-Peer(P2P) 方式と呼ばれる。

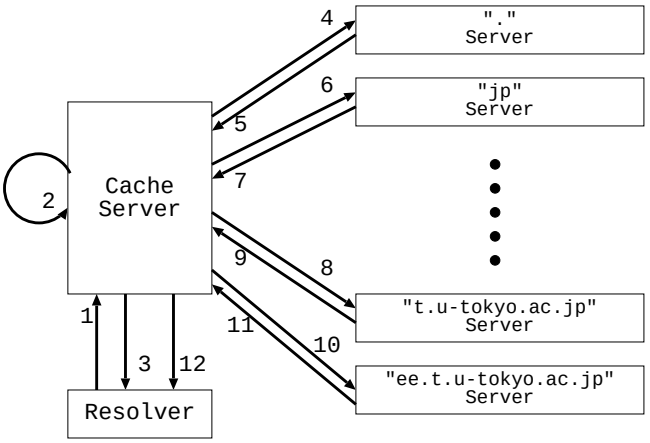

図 9: 名前解決の順序

1. リゾルバはキャッシュサーバに、www.ee.t.u-tokyo.ac.jp の IP アドレスを教えて欲しいと依頼する。
2. キャッシュサーバは、自身のキャッシュ内に www.ee.t.u-tokyo.ac.jp の IP アドレスデータがあるかを検索する。
3. 自身のキャッシュでヒットしたら、そのデータをリゾルバに返答する。
4. もし無ければ、.” を扱うコンテンツサーバに対して、jp” を扱うコンテンツサーバがどこにあるか（IP アドレスが何であるか）の検索クエリを送信する。
5. .” を扱うコンテンツサーバは、キャッシュサーバからの問い合わせに対し、自身が保持するゾーン情報を参照し、jp” を扱うコンテンツサーバの IP アドレスを返答する。
6. キャッシュサーバは、jp” を扱うサーバに、ac.jp” を扱うコンテンツサーバの IP アドレスについての検索クエリを送信する。
7. jp” を扱うコンテンツサーバは、自身が保持するゾーン情報を参照し、ac.jp” を扱うコンテンツサーバの IP アドレスを返答する。

   （以下、再帰的に繰り返し、）
8. キャッシュサーバは、ee.t.u-tokyo.ac.jp” を扱うコンテンツサーバの IP アドレスの問い合わせを行う。
9. t.u-tokyo.ac.jp” を扱うコンテンツサーバは、ee.t.u-tokyo.ac.jp” を扱うコンテンツサーバの IP アドレスを返答する。
10. キャッシュサーバは、ee.t.u-tokyo.ac.jp” を扱うコンテンツサーバに対し、www.ee.t.u-tokyo.ac.jp” の IP アドレス問い合わせを行う。
11. ee.t.u-tokyo.ac.jp” を扱うコンテンツサーバは、www.ee.t.u-tokyo.ac.jp” の IP アドレスをキャッシュサーバに返答する。

12. キャッシュサーバは、www.ee.t.u-tokyo.ac.jp” の IP アドレスをリゾルバに返答する。

キャッシュサーバからコンテンツサーバに検索を行う際に、最初の検索クエリを投げかける対象は、.” を担当するコンテンツサーバ (ルートサーバ) である。この IP アドレスは、あらかじめ自身のデータベースに記録してあり、このファイルはヒントファイルと呼ばる。ヒントファイルは、世界中に分散された 13 個のルートサーバの IP アドレスが記されている。[10]

- 課題
  - リゾルバ上のキャッシュサーバの IP アドレスは、Linux システム上では、/etc/resolv.conf” に登録されている。このファイルを確認せよ。また、このファイルを編集し、誤った IP アドレスが指定された場合には、通常のインターネット利用にどのような影響が発生するかを予想し、実際に確認せよ。

[10] 13 という個数は、DNS で利用される UDP パケットに収まるクエリ数から定義されている。

6.3 階層構造

6.2 節に示したように、DNS は階層構造を利用して、より深いレベルのゾーンに関するデータを、別のサーバに任せる (委譲する:delegate) 形態になっている。このデータ委譲は、DNS の管理権限 (management priviledge) の委譲も可能にしている。

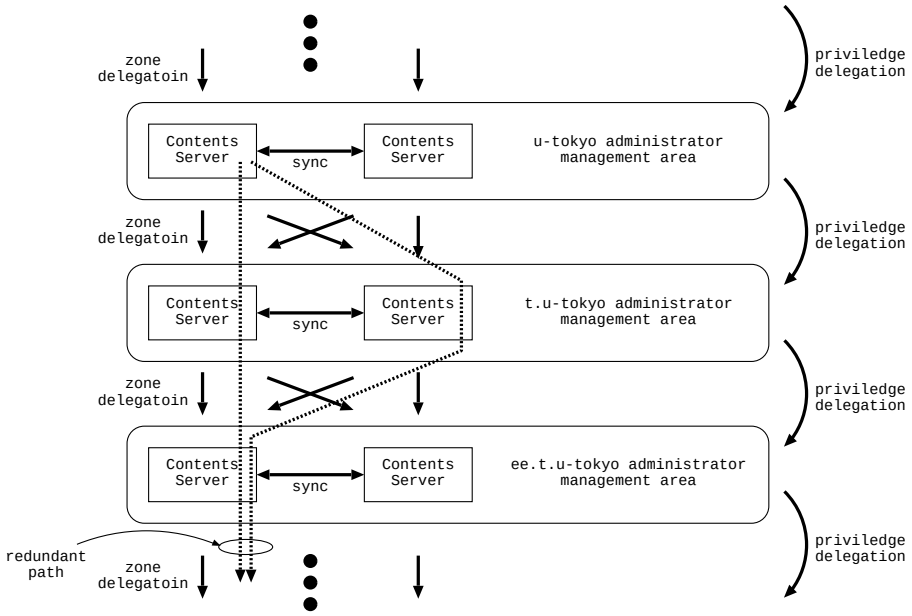

図 10: DNS の階層構造

例えば、u-tokyo.ac.jp” を扱うコンテンツサーバは、東京大学の情報基盤センターが管理しており、t.u-tokyo.ac.jp” では工学部情報システム室が、ee.t.u-tokyo.ac.jp” では、電気系管理者がそれぞれ管理を行っている。この場合に、www.ee.t.u-tokyo.ac.jp” の IP アドレスを扱っているのは、電気系管理者の配下にある ee.t.u-tokyo.ac.jp” を扱うコンテンツサーバである。もし、www.ee.t.u-tokyo.ac.jp” として登録してある IP アドレスを変更する場合には、電気系管理者が ee.t.u-tokyo.ac.jp” を扱うコンテンツサーバにて、設定変更をするだけで十分であり、情報基盤センターや、工学部情報システム室に作業依頼をすることは無い。同様に、もし ee.t.u-tokyo.ac.jp” よりさらに深いレベルのゾーンを作成する事も、電気系管理者のみの判断により実現できる。このように、インターネットに存在する様々な DNS 情報は、階層構造を利用した設定権限の委譲により、その管理を（一極集中せずに、）分散させている。

また、DNS では、ある一つのドメインにおいて、複数のコンテンツサーバを登録することができる。この複数のコンテンツサーバ間で、登録情報を複製しておくことで、冗長性 (redundancy) の確保が実現できる。つまり、上記の階層構造では、あるレベルのドメインを扱うコンテンツサーバから、一つ下位レベルのコンテンツサーバへのリンクを複数持つことになり、障害であるリンクが使えなくなっても、別リンクを利用して、検索を続けることができるのである。

このように、DNS は階層構造を利用することで、管理の分散や冗長性の確保が実現できる。

- 課題
  - リゾルバとして稼働するホスト上で、専用アプリケーションを利用することにより、手動にて DNS の検索クエリを発行することができる。dig” コマンドを利用して、リゾルバとキャッシュサーバ間の通信を確認せよ。同時に wireshark も実施し、どのようなパケットになっているかの確認も行うこと。
  - リゾルバとして稼働する Linux ホスト上で一時的にキャッシュサーバとしての機能を扱えるツールに、dnstracer” がある。このツールを利用して、視覚的に DNS の検索状態を把握し、再帰的な検索が行われることを確認せよ。
  - リゾルバ、キャッシュサーバ、コンテンツサーバは機能であるため、例えば、リゾルバ、キャッシュサーバを同一ホストで稼働させることも可能である。もちろん、リゾルバ、キャッシュサーバ、コンテンツサーバすべてを動かすこともできる。各自の PC に DNS サーバをインストールし、リゾルバとキャッシュサーバが同時に稼働する構成にせよ。DNS サーバは bind を利用すること。そして、リゾルバより、普段使わないが存在するホスト名（例えば、www.whitehouse.gov”）に対する検索クエリをキャッシュサーバに投げかけ、キャッシュサーバが再帰的に検索する事を確認せよ。また、一度検索したクエリはキャッシュサーバのキャッシュに一定期間蓄積されることを確認せよ。
  - DNS は階層構造を利用することで、規模性の確保も可能となる。どのようにこれを実現しているか考察せよ。

6.4　リソースレコード

DNS データベースの中で、検索結果として得られる各データをリソースレコード (RR:Resource Record) と呼ぶ。リソースレコードの属性には種類があり、これをレコードタイプ (Record Type) と呼ぶ。DNS 検索システムは非常に拡張性が高く、様々なレコードタイプが定義されているが、その中でも代表的なものを以下に示す。

6.4.1　A レコード

6.2 節での例では、「ee.t.u-tokyo.ac.jp” を扱うコンテンツサーバは、www.ee.t.u-tokyo.ac.jp” の IP アドレスをキャッシュサーバに返答」している。この場合には、A(Address) レコードが利用されている。なお、IPv4 アドレスは A レコードに登録されているが、IPv6 アドレスは AAAA レコード (Quad A レコードと呼ばれる) に登録されている。

6.4.2　NS レコード

同様に、6.2 節での例では、「jp” を扱うサーバに、ac.jp” を扱うコンテンツサーバの IP アドレスについての検索クエリを送信」している。このように、より深いレベルのドメインを扱うコンテンツサーバを指し示す場合には、NS(Name Server) レコードが利用される。NS レコードで定義されているデータは、IP アドレスではなく、ホスト名である。よって、上記の例をレコードタイプを用いて厳密に説明すると、「NS レコード” の問い合わせを行い、応答として得られるホスト名について、A レコード” の問い合わせを行うこと」となる。

6.4.3　SOA レコード

SOA(Start Of Authority) レコードとは、コンテンツサーバが扱う各ゾーン情報についてのパラメータを扱うリソースタイプである。このレコードには、キャッシュ可能な時間 (TTL:Time To Live) や、管理者の連絡先等が定義されている。

これらのレコードタイプは、例えば以下のように使われる。

- www.ee.t.u-tokyo.ac.jp の A レコードを教えてほしい (i.e. ホスト名が www.ee.t.u-tokyo.ac.jp として登録された IP アドレスを教えてほしい)
- ee.t.u-tokyo.ac.jp の NS レコードを教えてほしい (i.e. ee.t.u-tokyo.ac.jp のゾーン情報を持つコンテンツサーバのホスト名を教えて欲しい)
- 課題
  - 実験 TA の指示に従い、各自の PC にて、コンテンツサーバを稼働させ、DNS の階層構造を作成せよ。
  - 6.4 節で挙げた他に、代表的なレコードタイプには、MX レコード、CNAME レコードがある。これらの定義を調査し、実際に各自のコンテンツサーバに設定し、利用せよ。
- 発展課題
  - ホスト名から IP アドレスを検索することを (A レコードを解決すること) を、正引きと呼ぶ。この逆で、IP アドレスからホスト名も検索できる。この操作を逆引きと呼び、PTR レコードを利用する。この動作を調べ説明せよ。

7　相互接続

前章までの課題において、日常的に利用されているインターネットアーキテクチャの、基本部分を確認した。これらの構成は、実際のインターネット環境に接続しても正しく稼働するはずである。

- 課題
  - 実験 TA の指示に従い、動的経路制御を用いて、実際のインターネット環境に接続し、実験網からインターネット、インターネットから実験網の双方向にて到達性の確認せよ。
  - 実験 TA の指示に従い、上記のインターネットに接続された各機器について、DNS サーバの設定を行い、実験網外から正しく参照できることを確認せよ。
- レポート課題
  - 相互接続実験で作成したネットワークについて解説せよ。解説では、一連の実験で確認した要素技術が、どこに含まれているか具体的に示すこと。また、他チームとの相互接続、インターネット接続をしている点に着目し、各要素技術の動作状況を示すこと。
  - 各自の自宅にて、PC が電源投入され、ネットワーク接続を行い、東大の web ページ (www.u-tokyo.ac.jp) の web 閲覧を行う時点までに、どのような通信が行われているかを、可能な限り詳細に述べよ。その際には、各部分において利用されるプロトコル、および通過する経路も調査せよ。

参考文献

[1] The Internet Engineering Task Force (IETF), Request for Comments (RFC), http://www.ietf.org/rfc.html

[2] JPRS, DNS 関連情報, http://jprs.jp/tech/

[3] WIDE University, School of Internet, http://www.soi.wide.ad.jp/

[4] Internet Assigned Numbers Authority (IANA), Domain Name System (DNS) Parameters, http://www.iana.org/assignments/dns-parameters