Week4 Review Notes

ELEC0099: Introduce to Internet Protocol Networks 21/22

RUFENG DING

Contents

1 What is routing?

The propagation of connectivity information in order build the routing tables. Routing tables are used for packet forwarding. Final hosts usually are configured to talk with one router.

Routing Tables

Routing:

- Routing protocols act before any data packets go on the network.
- They are not involved directly in data transmission.
- They are equivalent to the people who put traffic signs on roads.

Routing/Forwarding Plane

Routers are just computers with more than one network interface. They forwarding plane forwards every packet. It need to be fast. On the Routing Plane, one or more Routing daemons(normal programs) talks with daemons on other routers, to exchange routing information and updates the routing tables in the forwarding plane.

Building the Routing Tables

Routing tables can be built by hand but

- In some cases there are many entries
- They may have to change quickly when things change in the network

Today we need Routing Protocols to automatically build and update the Routing Tables.

1.1 The Big Pictures: Intra-domain vs inter-domain

Inside each Autonomous System (AS) we run an **Intra-domain routing** protocol.

Between each **AS** we run a **Inter-domain Routing** protocol.

Graphs a useful mathematical abstraction several real-life problems can be approached by using graphs. These are used in lots of fields like road traffic optimization. Graph have: Nodes/Edges/Costs.

1.2 Dijkstra's Best Path algorithm

looks to the fig 1. The notes of Dijstra.

Notes

83.03 A teaching note on Dijkstra's shortest path algorithm*

As part of the International Baccalaureate Advanced Mathematics course, our class has been learning Graph Theory including the standard tabular method for finding the shortest path between two vertices.

The standard procedure is explained here, by doing one of the examples from the classic textbook [1, p. 167].

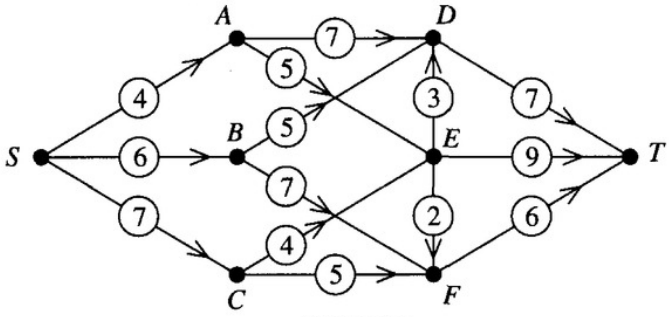

FIGURE 1

Give vertex S a potential of 0 and then label vertices reached by an arc from S by the number 0 (the potential of S) plus the distance from S to that vertex, so vertices A, B, C are labelled 4, 6, 7 respectively. Take the smallest label, 4, set the potential of all vertices labelled by it to 4, so here A obtains potential 4. Then label each vertex, reached by an arc from A, and not yet assigned a potential, with 4 (the potential of A) plus the distance from A to that vertex (unless the vertex already has a smaller label, in which case it retains that smaller label). Thus D, E receive labels of 11, 9 respectively. Since 6 is now the smallest label, 6 is taken as the potential of B and this process is repeated, continuing until T receives a potential (which is the shortest distance from S to T).

Presenting this process in tabular form gives:

Standard tabular method for the graph in Figure 1

	S	A	B	C	D	E	F	T
S	[0]	4	6	7				
A		[4]			11	9		
B			[6]		11		13	
C				[7]		9	12	
E					11	[9]	11	18
D, F					[11]		[11]	17
T								[17]

* Something went wrong with the printing of Note 83.03 so it is reprinted here.

Figure 1: Dijstra

2 RIP

Distance Vector vs Link-State

Bellman-Ford

Distribution of aggregate information

Distribution only to neighbours - Distance Vector Algorithms (eg. RIP Protocol)

Dijkstra

Distribution of information on local links

Distribution to every node - Link-state Algorithms (eg. OSPF Protocol)

RIP - Routing Information Protocol: Original routing protocol. Developted in 1969 Defined in RFC 2453. Gained popularity with inclusion in UNIX BSD in 1982. RIP itself is an Application level protocol running over UDP but it affects the network layer.

2.1 RIP Operation

a routing running RIP broadcast a routing message every 30 seconds to all it neighbours. Each update contains a pair <network,distance>. Wherer distance is usually the number of hops (in certain cases managers may increase the weight of a hop).

RIP participants updatae their routing tables based on these broadcasts.

RIP example
Importtant: routers never know the topology of the network. **Don't forget**: routers will also send message to same node which will also propagate it. Routers will choose the best metric to reach. **All routers** send equivalent messages.

Hysteresis - A RIP router does not change the route if the distance received is the same. If a router does not received an update after 180 seconds the router is deleted. The maximum distance for RIP is 16. If an administrator wnats to have bigger value it has to parrtition the network.

2.2 Split horizon

CURE: In RIP routers do not annouce routes on the links these were received - Split horizon.

2.3 RIP message

command: 1 = request 2 = reply.

- updates are replies whether asked for or not
- initializing node broadcasts request
- requests are replied to immediately

Version: 1 or 2.
Address family: 2 for IP.
IP address: non-zero network portion, zero host portion.

- identifies particular network

Metric:

- Path distance from this router to network
- Typically 1, so metric is hop count

3 OSPF

OSPF: open shortest path first. To encourage the use of link state routing protocols a working group of the IETF has designed OSPF. Defined in RFCs 1245, 1246, 1247, 2328(v2), 5340(v3 for IPv6).

3.1 OSPF operation

differences form RIP:

- lt is link-state based. Every router knows about every link in the network.
- incude type of service routing.
- authenticaiton of routing messages
- definition of areas
- allow load-balancing of traffic
- serveral minor improvements...

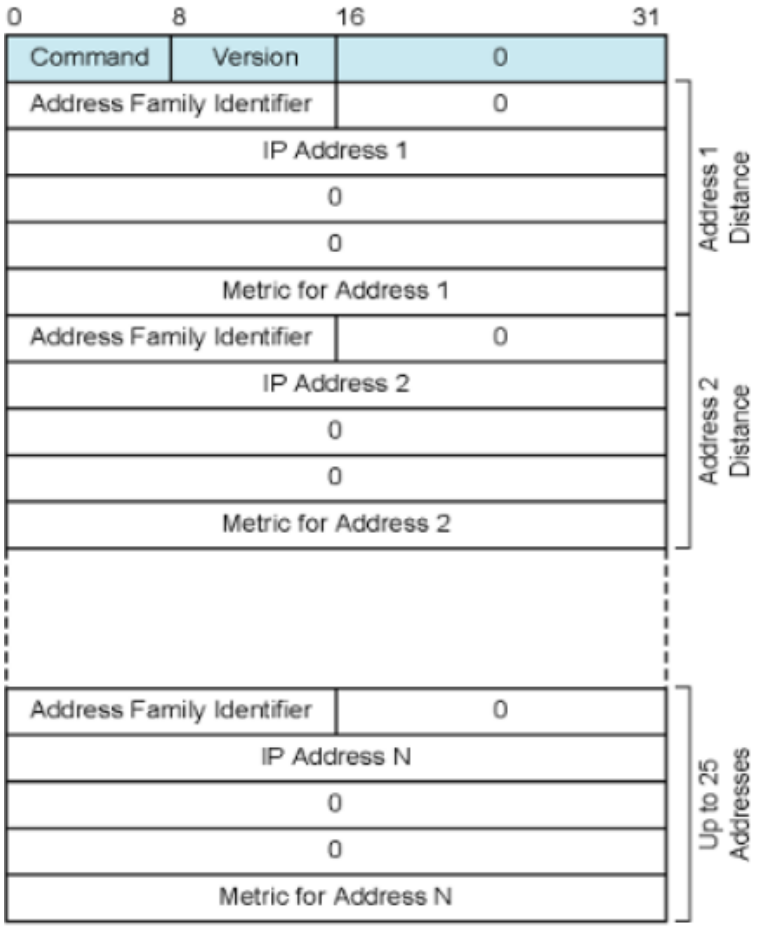

Figure 2: RIP message

3.2 OSPF Overview(OSPF = Flooding + Dijkstra)

Router maintains descriptions of state of local links as a directed graph:

1. 1 - Transmits updated state information to all routers it knows about using Flooding It sneds a message about each link to all the neighbours. These replicate that message to all the neighbours except the one that sent the message. If routers receive a message they already broadcasted they just drop it.
2. 2 - Router receiving update must acknowledge lots of traffic generated
3. 3 - after receeiving all the message form the network OSPF routers calculate what is the shortest path to reach all the destination.
4. 4 - with that information they calculate what is the link to the used for every network
5. 5 - Because all the routers have the same information and calculate the same paths using the same algorithm all the forwarding decision are consistent.

6. 6 - If for some reason information is not consistent packets may get looped. eg. link goes down and that information has not arrived to all the routers. Remember TTL!!!

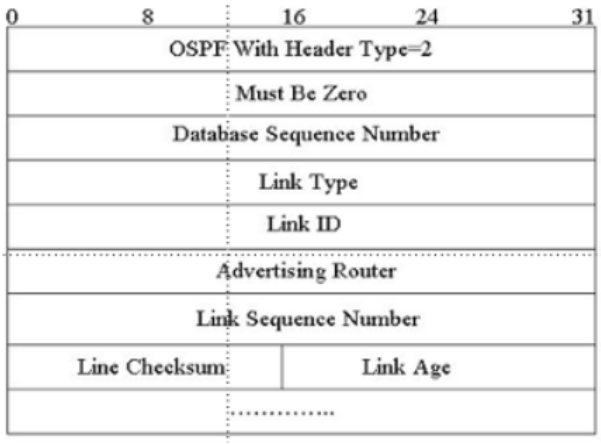

Figure 3: OSPF message

In reality OSPF announces links or networks. The type is indicated in the 'Link type' of the Database Description message

3.3 Link costs

cost of each hop in each direction is called routing metric. OSPF provides flexible metric scheme based on type of service (TOS)

- Normal (TOS) 0
- Minimize monetary cost (TOS 2)
- Maximize reliability (TOS 4)
- Maximize throughput (TOS 8)
- Minimize delay (TOS 16)

Each router generates 5 spanning tree (and 5 routing tables)

3.4 Area

mark large internets mroe managable. Configure as backbone and multiple areas.

- Area - colleciton of contiguous network and hosts plus routers connected to any included network.
- Backbone - contiguous collection of networks not contained in any area, their attached routers and routers belonging to multiple areas.

3.5 Operation of Areas

each area runs a separate copy of the link state algorithm.

- topological database and graph of just that area
- link state information broadcast to other routers in area
- reduced traffic
- intra-area routing relies solely on local link state information

3.6 OSPF Packet Header

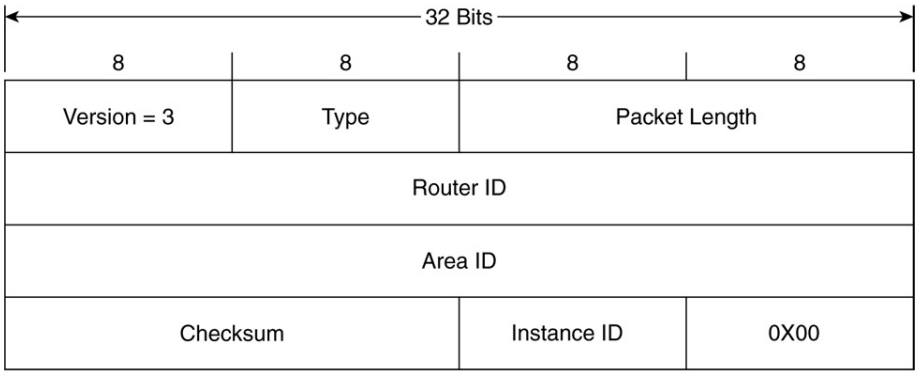

Figure 4: OSPF packet header

Packet format Notes:

- version number 2 is current
- type one of 5
- packet length: in octets including header
- router id: this packets source 32 bit
- area id: Area to which source router belongs
- Authentication type: null,simple password or encryption
- Authentication data: used by authenticaiton procedure

OSPF packet types

- hello: used in neighbour discovery

- databse description: defines set of link state information present in each router's database
- Link state request
- Link state update
- Link state acknowledgement
- Runs directly over IP

3.7 IS-IS

Has it roots in OSI. Similar to OSPF: also link state. also uses Dijstra. Runs directly over layer 2(paralled to IP).

Several minor differences:

- OSPF supports point to multipoint links
- OSPF supports virtual link
- an OSPF router can belong to multiple areas whereas an IS-IS router can belong to only one area.

3.8 OSPF weight calculation

There is no standard. $10^8/(interfacebandwidth)$.

More elaborate methods:

- Take into account traffic matrix
- spread the traffic over all links:minimizing the most used link

3.9 OSPF Configuration

OSPF works by itself...

But needs to be configured system administrators login into routers(these are just computers) and through Command line configure the parameters.

3.10 Summary

- Routing tables are built by applications running on the routers that communicate using routing protocols

- Graphs are mathematical abstraction used to calculate routing tables
- Each Autonomous System(AS)runs the same protocol across all its routers
- There are two types of routing protocols:link-state and distance-vector
- OSPF is the most used interior routing protocol
- OSPF uses a flooding algorithm to propagate link-state information and the Dijkstra algorithm to calculate the best path to every other node in the network
- Given the best path,calculating the routing table is straightforward

4 BGP

4.1 Routing continued

At this point you should know how IP packets traverse an Autonomous system to reach a destination inside taht AS.Here all the routers are under the same administration.

Now we will look at how are packets routed over several Autunomous systems?

- Many more networks
- Owned by different organizations

4.2 Autonomous System

The Internet is not controlled by any centeral authority. If an organization is big enough it can form an Autonomous System. These Autonomous Systems make bilateral agreements between themselves. There isn't any other form of organization!!!

An AS can be a big company, it can be a Research network and it can be an ISP. Each AS, runs one instance of an IGP (OSPF,RIP,etc).

4.3 BGP details

Border Gateway Protocol:

- connects Autonomous Systems
- Transmits reachability to networks (or prefixes)

allows routers (gateways) in different ASs to exchange routing information. Messages sent over TCP. Three funcitonal procedures:

- Neighbour acquisition
- Neighbour reachability
- Network reachability

4.4 BGP Messages

- Open Start neighbour relationship with another router.
- Update
  - Transmit information about single route.
  - List multiple routes to be withdrawn.
- Keepalive
  - Acknowledge open message.
  - Periodically confirm neighbour relationship.
- Notification Send when error condition detected.

4.5 BGP Message Formats

four types of mesages:

- OPEN - start a BGP peering
- KEEPALIVE - keeps alive the connection
- UPDATE - updates any information about network reachability using
  - withdraw routes
  - attributes
  - network layer reachability (network / prefixes)
- NOTIFICATION - notifies of any protocol errors. Must close TCP connection after it is sent.

Most used attributes

- ORIGIN - Tells how a route was learnt (1 - IGP/2 - iBGP/3 - Other)
- ASPATH - List of AS numbers that packets will traverse for this announcement.

- NEXT-HOP - IP address of the router that packets need to be sent.
- MUTL-EXIT DESCRIMINATOR(MED) - when a AS has several options to a given destination, it may include information for preferring one of them.
- LOCAL-PREF - allow a specific AS to signal internal preference for a route to a destnation (when there is nore than one)

4.6 BGP and CIDR

CIDR allows BGP to transmit much less information due to Aggregation of routes.

4.7 BGP = iBGP + eBGP

when we talk about BGP we usually mean eBGP.

iBGP it is the part of iBGP concerned with transmitting external connectivity information inside the Autonomous System. IT IS NOT an Interior Routing Protocol. Routers will not learn how to reach a router inside their network through iBGP.

4.8 BGP Route Reflection

some AS have a big number of external connections.

And If every router transmits every message revceived to every router in the AS there will be lots of traffic generated. iBGP uses sometimes **Route Reflectors** These reveice the messages and send summaries to all the routers. (For Ases with more than 100 nodes, it is recommended to use a router reflector instead of a full mesh)

4.9 Route selection

every router in an AS needs now to decide which route to take to all prefixes based on following rules(by order):

1. looks at LOCAL-PREF
2. the route with less ASes in the AS list
3. looks at the MED (Multi-Exit Discriminator)
4. IGP distance to the NEXT-HOP
5. if all routes were learned through iBGP, choose the neighbour with the lowest BGP identifier

4.10 Transit vs non-Transit

when a ISP is providing connectivity to its clients, it is also providing connectivity to clients of other ISPs!!!

Imaging that traffic form Customer 1 was going to customer 3 via ISP2. Neither C1 or C3 are paying ISP2 .. ISP2 would be carrying traffic that was not paid. Two options: Transit Service and Non-transit service.

4.11 BGP Polocies

In theory, BGP allows each domain to define its own routing policy. In practice, there are two common policies:

- customer-provider peering: customer C buys Internet connectivity from provider P
- shared-cost peering: Domains x and y agree to exchange packets by using a direct link or through an interconnection point

4.11.1 Customer-provider peering

Customer sends to its provider its internal routes and the routes learned form its own customers - Provider will advertise those routes to the entire Internet to allow anyone to reach the Customer.

Provider sends to its customers all know routes - Customer will be able to reach anyone on the Internet

4.11.2 Shared-cost peering

Peer X sneds to Peer Y its internal routes and the routes learned from its own customers.

- Peer Y will use shared link to reach Peer X and Peer Y's customer
- Peer X's providers are not reachable via the shared link

4.12 Traffic cost models

fixed cost : one fixed payment per month
linear cost: payment proportinal to traffic
95th percentile:

- Bandwidth is measured (or sampled) from the switch or router and recorded in a log file. In most casses, this is done every 5 minutes.
- At the end of the month, the sample are sorted from highest to lowest, and the top 5% (which equal to approximately 36 hours of a 30-day billing cycle) of data is thrown away.
- The next highest measurement becomes the billable use for the entire month.

4.13 Routing Policies

Routing policies implement business relationships between domains.

The routing policy of a domain is implemented via the route filtering mechanism on BGP routers:

- Inbound filtering: Upon reception of a route from a peer, a BGP router decide whether the route is acceptable, and if so whether to change some of its attribtes.
- Outbound filtering: Before sending its best route towards a destination, a BGP router decides which peers should receive this route and whether to change some of its attributes before sending it.

4.14 Route Instabilities

several types of Instabilities affect the behaviour of BGP:

- Instabilities of the IGP
- Hardware Failures
- Software problems
- Insufficient Horsepower
- Insufficient memory
- Network Upgrades
- Human Error
- Backup Link Overloads

4.15 Route Flap Dampening

imagine that the link AS1-AS2 goes up and down frequently. This generates disturbances in the entire Internet. This could be aggravated if this is done intentionally. In this case AS2 does Route Flap Dampening. That is it stops advertising routes to AS1 to the rest of the Internet.

4.16 Redundancy with Multi-homing

Multi-homing refers to a single network having more then one connection to the Internet. This can be done mainly for two reasons:

- Load-balancing
- Fault-tolerance

Two types:

- Single provider
- Multiple provider

Multi-homing with Multiple Providers more interesting and useful is when a customer has two providers. Multi-homing brings some problems:

- Destroys route symmetry
- May cause packet reordering
- Makes firewalls more difficult
- impact on the BGP tables: CIDR become difficult

4.17 Load Balacing

Inbound Load Balancing can be achieved by announcing some of our routes in one Link and some in the other(s).

An easy way of achieving Outbound Load Balancing is to use a Route Reflector. The Route Reflector can be configured to announce some routes to some of our routers and other routes to the others.

4.18 Ineraciton with the IGP

The interaction of the IGP with BGP needs to be careful thought by the network administrator. You can have the default that all you networks are annouced on all the BGP peering or you can decide which areas are announced to. You can even decide not to annouance some networks (eg. networks used on the administration of the AS) In certain circumstances networks obtained from BGP can be inserted into the IGP: Risky!!!

4.19 Anycast

anycast is another message paradigm: A packet is sent to the closest member of a group.

Today in the IPv4 internet this is achieved by announcing the same address/prefix from several locations in the Internet. BGP will redirect packets from different sources to the closed destinations. eg. DNS root servers.

4.20 confederations

another way of dealing with big ASes is to break them into confederations. A confederation is an AS broken into sub-ASes. These can run different IGPs uesd in company acquisitions.

4.21 Forwarding summary: longest Prefix Match

since with prefix/CIDR there is not a fixed distinction between ”network” and ”host” routes select the next hop using longest Prefix Match.

Imaging that the routing table has two entries:
138.39.0.0/16 - Link 1
138.39.16.0/24 - Link 2

if a packet arrives with destination address 138.39.16.32 the router uses link2 because 24 bits match against 16.