Networking Layer

Discussion on switching is continued with an increased scope.

* Layer 3 Switching

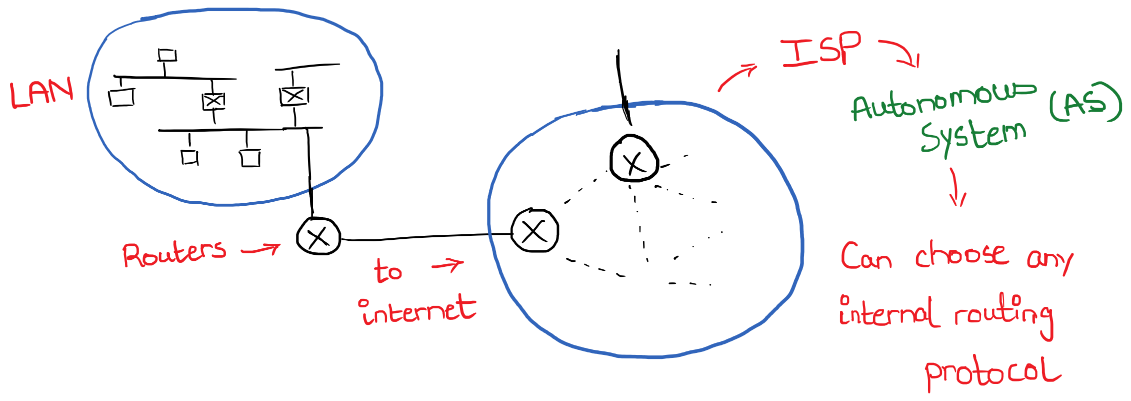
Is used to have millions of users connected to a network. This is not possible via extensive Layer 2 switching because of a number of short comings.

- The spanning tree protocol would cause the packet to take a longer path than the optimal one, as few nodes are disabled. For a large number of nodes, it would be very inefficient.
- This size of the forwarding table is proportional to number of nodes, which is very large. This is because each node has to be stored in the table. The MAC addresses have no restriction, which causes storing and lookup tough. This non-restricted storage is called as Flat addressing.

IP addresses are stored using Hierarchical Addressing in Layer 3 to combat this.

Layer2 is not stable, as the failure of the root node would require the spanning tree to be reconstructed. The root node sends out Hello messages to let the network know that it is up an running. Failure of any node would cause the hello messages to not reach all nodes, needing tree reconstrunction.

Historically, different institutions used to have their own addressing conventions which made connecting them globally tough. IP addresses take care of this issue.



We will be looking at a few protocols used by AS. These are divided into two categories -

Intra domain Routing — Protocols within AS

Inter domain Routing — protocols between ASes (lol)

↳ Border Gateway Protocol (BGP)

" The " protocol used in internet.

Intra domain protocols are of two types -

1) Distance Vector RIP - Routing Information Protocol

2) Link-State routing OSPF — Open shortest path

ISIS — Intermediate system to IS

- These three protocols try to find shortest paths while avoiding loops in the system. Famous algorithms are used for this requirement.

1) Distance Vector protocol

Distance Vector uses a "distributed" Bellman-Ford Algorithm.

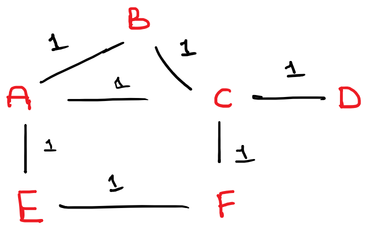
That is, only the next hop is needed, and not the entire path itself.

We utilise this property to modify the algorithm appropriately.

The algorithm's steps are described below –

- Each router in the network maintains a forwarding table, whose columns contain the possible destination, the hop need to reach it, and the cost of path to destination

- The table is initialised with itself having cost 0, and this is broadcasted such that only its neighbours can hear.



dest.	next hop	cost
A	–	0
B	B	1
C	C	1
E	E	1

A's table after one update

IP addresses

A hears (B,0), (C,0)
(E,0)

A sends out (A,0) to B,C,E

not broadcasted !

- After the update, A sends out the table to its $\boxed{\text{neighbours}}$. Similarly, it also hears additional information from its neighbours.

  In the above example:-

  A hears from B — (B,0), (A,1), (C,1)

  from C — (C,0), (B,1), (A,1), (F,1), (D,1)

  from E — (E,0), (A,1), (F,1)

Updated table — assuming C is heard first

dest.	Next hop	Cost
A	-	0
B	B	1
C	C	1
E	E	1
! F	C	2 $(\overset{A\to C}{1} + \overset{C\to F}{1})$
! D	C	2 (1 + 1)

An entry is updated if $w(A\to x) + w(x\to Dest) < w(A\to Dest)$

* Reliability - what to do when a link fails?

Assume that the A→C link fails due to a practical error in the network. A's forwarding table is updated to contain (C,∞) as a row, which is then shared to E and B. Assume that reaching C from E required passing through A previously.

dest.	Next hop	Cost
C	A	2

old

$\xrightarrow[\text{from A}]{\text{E receives (C,∞)}}$

dest.	Next hop	Cost
C	A	∞

It then receives (C, 1) from F, which updates the routing table to (C, F, 2). This updated path ripples backwards, modifying the entry appropriately.

If a packet addressed to C is received by A when the row is (C, ∞), the packet is dropped. We would thus want the update period to be short. The updates are also called as Routing Advertisements.

There are two types of updates :-

1) Triggered updates –

Are triggered due to an event, such as disconnection in this case. They usually have a high priority. A sending out (C, ∞) is an example of this.

2) Periodic Updates

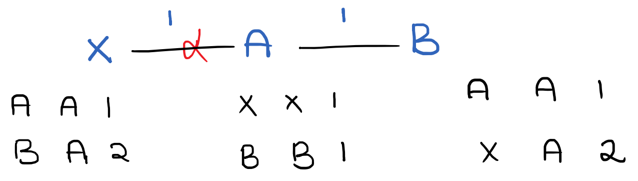
Tell neighbours about the shortest path and the cost. These are sent out periodically during the construction of the forwarding table.

Note that the reconstrunction of the topology of the entire network isnt possible by using the information of a single node. Although not a problem, this might be a cause for concern.

A few problems with the distance vector protocol are discussed below.

* Count to Infinity problem

This problem is caused due to a periodic update being received before a triggered update, and due to the topology being unknown.



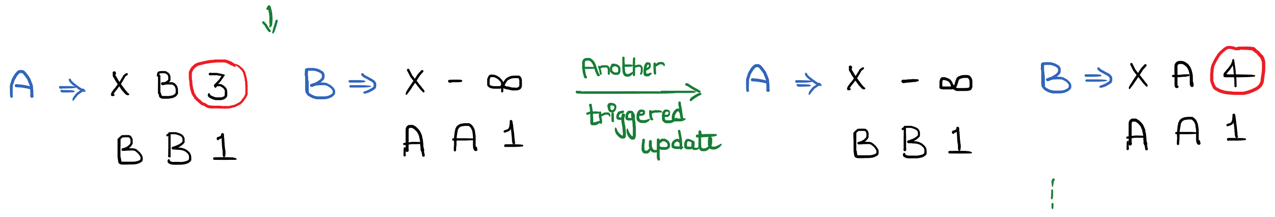
A	A	1
B	A	2

X	X	1
B	B	1

A	A	1
X	A	2

In the above example, the failure of X-A link triggers an update (x,∞) from A to B. Suppose that B sends a periodic update $(x,2)$ to A before the triggered update is received. This would cause A to wrongly think a path B-X of length 2 exists, which it updates in its table. B receives the triggered update and puts (x,∞).

The updated tables are :-



$A \Rightarrow$

X	B	$\boxed{3}$
B	B	1

$B \Rightarrow$

X	-	∞
A	A	1

$A \Rightarrow$

X	-	∞
B	B	1

$B \Rightarrow$

X	A	$\boxed{4}$
A	A	1

As the rows of A, B have changed; another update pair similar to above is triggered. This would inturn cause the cost (highlighted in red) to count upto ∞. (Hence the name)

Fix 1 A simple fix is to have a maximum value that the cost can attain. Once attained, we assume that the node is not reachable.

Max cost in RIP is 16.

Fix 2 Split-horizon

We do not want the information about a destination to be to a node which is the next hop to reach that destination.

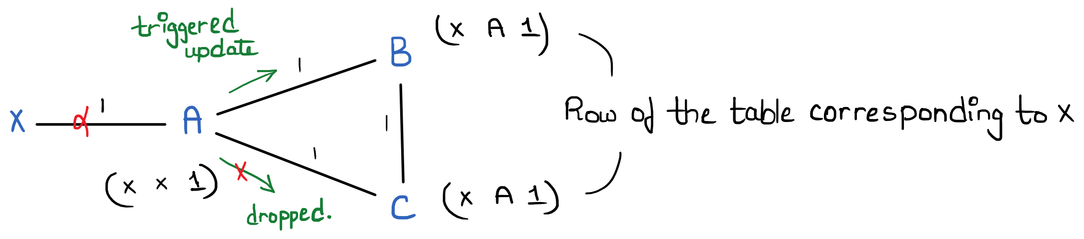
In the above example, the periodic update from B about X is not received by A which breaks the cycle and solves the problem.

Fix 3 Split-Horizon with Poisson reverse

A node tells the next hop to its destination that the cost of reaching that destination is ∞. (Instead of staying quiet)

That is, B sends (x, ∞) to A by this fix, which also solves the problem.

While these fixes work for this case, we cannot be sure that they work everywhere. Consider the following example.



The X-A link failure triggers an update from A to B and C. Now, assume that the update never reaches C as the signal is lost. The respective rows would be :-

A ⇒ (x,-,∞) B ⇒ (x,-,∞) C ⇒ (x, A, 2)

Let us assume that split-horizon is followed. Then, C sends out an update regarding x only to B. The row is updated to be (x, C, 3) now. Again, via Split horizon, B sends out an update to only A, whose row is updated to be (x, B, 4). A sends a message to B and forms a cycle, with the new row being (x, A, 5).

↳ Count to Infinity even with Split-Horizon!

Using split-horizon with poisson reverse also faces this issue, but ∞ (16) is reached faster in its case leading to lesser time usage. These both ideas solve the problem only to some degree.

Routing Information Protocol (RIP)

This is based on distance vector protocol. We assume that cost of all links are 1, and 16 is taken as ∞.

- Having $16 = \infty$ causes the scope of the graph to be limited.

Advantages and Disadv.

Distance Vector protocol is very easy to implement. However, it takes time for the routing tables to converge. Also, count-to-Infty has not been solved elegantly.

2) Link-State Routing

Each node broadcasts the cost to its immediate neighbours. That is, the local topology is broadcasted. The topology of the entire graph can be reconstructed by any node using the local topology information

The dijkstra's algorithm is used in this case as the topology is known. The node itself is taken as the source in the algorithm, after which a minimum-distance tree is created. The routing table can be computed from this tree.

Count-to-Infinity is not a problem here. If the link between any two nodes fails, the nodes send out a signal calling for the reconstruction of the trees.

Advantages and Disadv.

There are no routing loops and count-to-infinity problem. The construction is also fast as each node needs to transmit information only once. However, the algorithm is quite complex.

Preferences

Intra-domain network is small where count-to-infinity is not significant —— Distance Vector protocol (RIP)

Large Intra-domain network causing frequent failure of links —— Link State Routing