COMPUTER NETWORKS AND INTERNET PROTOCOLS

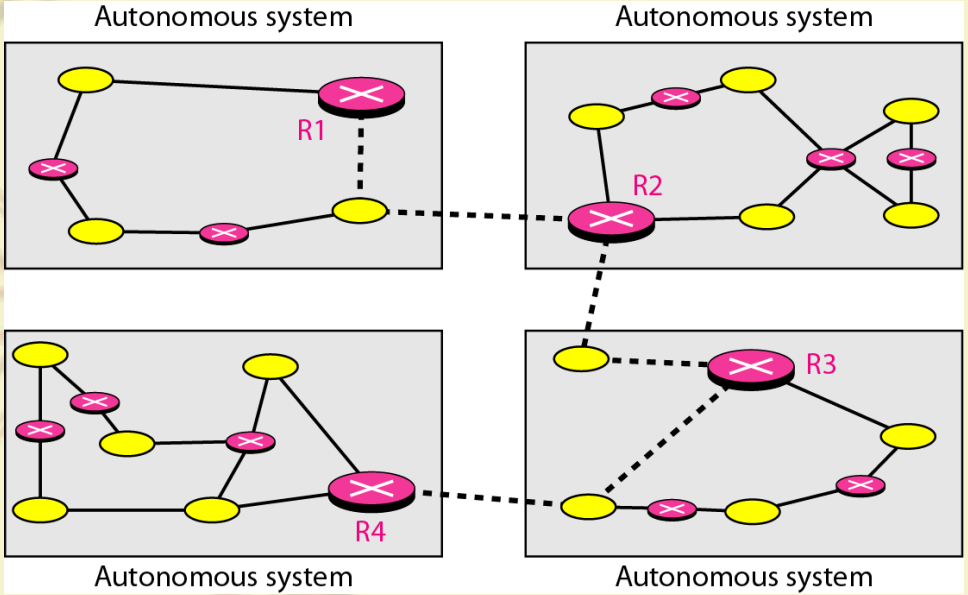
IP Routing – IV [Border Gateway Protocol - BGP]

SOUMYA K GHOSH
COMPUTER SCIENCE AND ENGINEERING
IIT KHARAGPUR

SANDIP CHAKRABORTY
COMPUTER SCIENCE AND ENGINEERING
IIT KHARAGPUR

Autonomous System (AS)

AS is a logical portion of a larger IP network. An AS normally consists of an internetwork within an organization. It is administered by a single management authority. An AS can connect to other autonomous systems managed by the same organization or other public or private networks.



Ref: Data communications and networking by Behrouz A. Forouzan; TCP/IP Tutorials and Technical Overview, IBM Redbooks

Routing Protocols in AS

Two sets routing protocols are used – (i) to determine routing paths within an AS; (ii) others are used to interconnect a set of autonomous systems:

Interior Gateway Protocols (IGPs): Interior Gateway Protocols allow routers to exchange information within an AS. Examples of these protocols are Open Short Path First (OSPF) and Routing Information Protocol (RIP).

Exterior Gateway Protocols (EGPs): Exterior Gateway Protocols allow the exchange of summary information between autonomous systems. An example of this type of routing protocol is **Border Gateway Protocol (BGP).**

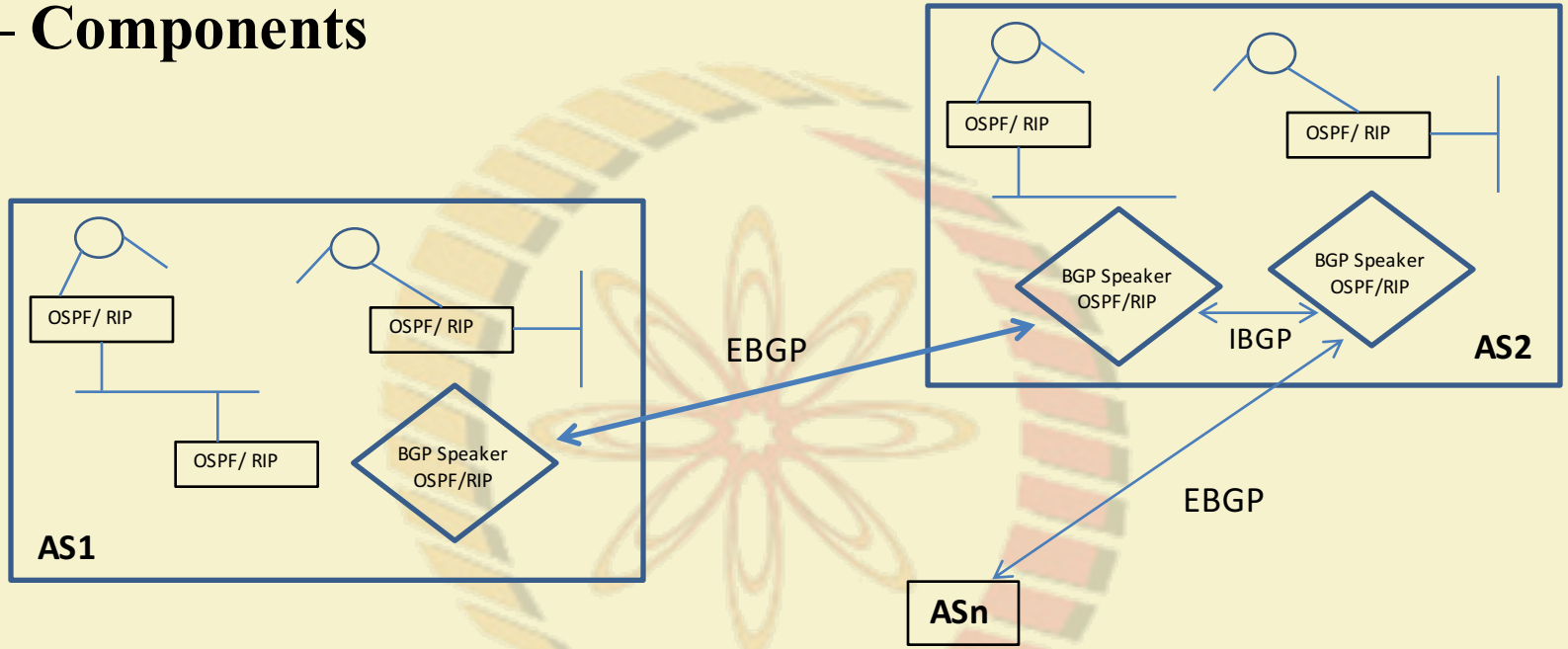
Ref: TCP/IP Tutorials and Technical Overview, IBM Redbooks

Border Gateway Protocol (BGP)

- Border Gateway Protocol (BGP) is an exterior gateway protocol.
- It was originally developed to provide a loop-free method of exchanging routing information between autonomous systems.
- BGP has since evolved to support aggregation and summarization of routing information.
- BGP is an IETF draft standard protocol described in RFC 4271 (BGP Version 4).

Ref: TCP/IP Tutorials and Technical Overview, IBM Redbooks

BGP – Components



Ref: TCP/IP Tutorials and Technical Overview, IBM Redbooks

BGP – Basic Concepts

- **BGP speaker**: A router configured to support BGP.
- **BGP neighbors** (peers): A pair of BGP speakers that exchange routing information.
- Two types of BGP neighbors:
  - Internal (**IBGP**) neighbor: A pair of BGP speakers within the same AS.
  - External (**EBGP**) neighbor: A pair of BGP neighbors, each in a different AS. These neighbors typically share a directly connected network.
- **BGP session**: A TCP session connecting two BGP neighbors. The session is used to exchange routing information. The neighbors monitor the state of the session by sending *Keepalive* messages.
- **AS number**: A 16-bit number uniquely identifying an AS.
- **AS path**: A list of AS numbers describing a route through the network. A BGPneighbor communicates paths to its peers.

Ref: TCP/IP Tutorials and Technical Overview, IBM Redbooks

BGP – Basic Concepts (contd…)

- **Traffic type**: BGP defines two types of traffic:
  - **Local**: Traffic local to an AS either originates or terminates within the AS. Either the source or the destination IP address resides in the AS.
  - **Transit**: Any traffic that is not local traffic is transit traffic. One of the goals of BGP is to minimize the amount of transit traffic.
- **AS type**: BGP defines three types of autonomous systems:
  - ***Stub***: A stub AS has a single connection to one other AS. A stub AS carries only local traffic.
  - ***Multihomed***: A multihomed AS has connections to two or more autonomous systems. However, a multihomed AS has been configured so that it does not forward transit traffic.
  - ***Transit***: A transit AS has connections to two or more autonomous systems and carries both local and transit traffic. The AS can impose policy restrictions on the types of transit traffic that will be forwarded.
  - The autonomous system can be either a multihomed AS or a transit AS.

Ref: TCP/IP Tutorials and Technical Overview, IBM Redbooks

BGP – Basic Concepts (contd…)

- **Routing policy**: A set of rules constraining the flow of data packets through the network. Routing policies are not defined in the BGP protocol. Rather, they are used to configure a BGP device. For example, a BGP device can be configured so that:
  - A multihomed AS can refuse to act as a transit AS. This is accomplished by advertising only those networks contained within the AS.
  - A multihomed AS can perform transit AS routing for a restricted set of adjacent autonomous systems. It does this by tailoring the routing advertisements sent to EBGP peers.
  - An AS can optimize traffic to use a specific AS path for certain categories of traffic.
- **Network layer reachability information (NLRI)**: NLRI is used by BGP to advertise routes. It consists of a set of networks represented by the tuple <length,prefix>. For example, the tuple <14,202.124.116.0> represents the CIDR route 202.124.116.0/14.
- **Routes and paths**: A route associates a destination with a collection of attributes describing the path to the destination. The destination is specified in NRLI format. The path is reported as a collection of path attributes. This information is advertised in UPDATE messages.

Ref: TCP/IP Tutorials and Technical Overview, IBM Redbooks

EBGP – IBGP Communications

- BGP does not replace the IGP operating within an AS. Instead, it cooperates with the IGP to establish communication between autonomous systems.
- BGP within an AS is used to advertise the local IGP routes. These routes are advertised to BGP peers in other ASs.

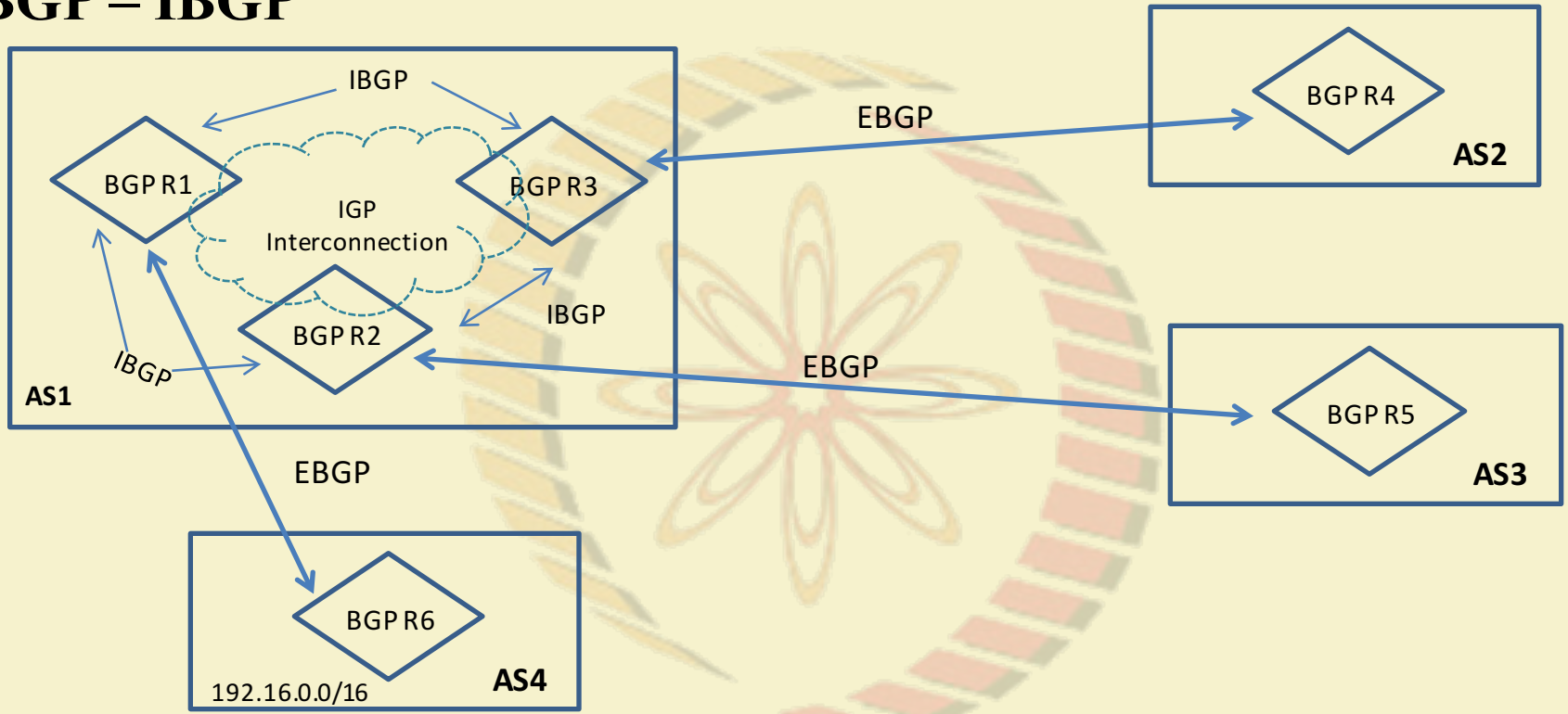
Ref: TCP/IP Tutorials and Technical Overview, IBM Redbooks

EBGP – IBGP Communications

- Role of BGP and the IGP: Both BGP and the IGP are used to carry information through an AS.
- Establishing the TCP session between peers: Before establishing a BGP session, a device verifies that routing information is available to reach the peer:
  - **EBGP peers**: EBGP peers typically share a directly connected network. The routing information needed to exchange BGP packets between these peers is trivial.
  - **IBGP peers**: IBGP peers can be located anywhere within the AS. They do not need to be directly connected. BGP relies on the IGP to locate a peer. Packet forwarding between IBGP peers uses IGP-learned routes.
- Full mesh of BGP sessions within an AS: IBGP speakers assume a full mesh of BGP sessions have been established between peers in the same AS.
- When a BGP speaker receives a route update from an IBGP peer, the receiving speaker uses EBGP to propagate the update to external peers. Because the receiving speaker assumes a full mesh of IBGP sessions have been established, it does not propagate the update to other IBGP peers

Ref: TCP/IP Tutorials and Technical Overview, IBM Redbooks

EBGP – IBGP



Ref: TCP/IP Tutorials and Technical Overview, IBM Redbooks

BGP Protocol

- BGP establishes a reliable TCP connection between peers. Sessions are established using TCP port 179.
- BGP assumes the transport connection will manage fragmentation, retransmission, acknowledgement, and sequencing.
- When two speakers initially form a BGP session, they exchange their entire routing table. This routing information contains the complete AS path used to reach each destination. This information avoids the routing loops and counting-to-infinity behavior observed in RIP networks.
- After the entire table has been exchanged, changes to the table are communicated as incremental updates.

Ref: TCP/IP Tutorials and Technical Overview, IBM Redbooks

BGP Packet types

- OPEN: This message type establishes a BGP session between two peer nodes.
- UPDATE: This message type transfers routing information between GP peers.
- NOTIFICATION: This message is sent when an error condition is detected.
- KEEPALIVE: This message determines if peers are reachable.

Ref: TCP/IP Tutorials and Technical Overview, IBM Redbooks

BGP Function

- **Opening and confirming a BGP connection**: After a TCP session has been established between two peer nodes, each router sends an OPEN message to the neighbor.
- **Maintaining the BGP connection**: BGP does not use any transport-based keepalive to determine if peers are reachable. Instead, BGP messages are periodically exchanged between peers. If no messages are received from the peer for the duration specified by the hold timer, the originating router assumes that an error has occurred. When this happens, an error notification is sent to the peer and the connection is closed.
- **Sending reachability information**: Reachability information is exchanged between peers in UPDATE messages. An UPDATE message is used to advertise feasible routes or withdraw infeasible routes.
- **Notification of error conditions**: A BGP device can observe error conditions impacting the connection to a peer. NOTIFICATION messages are sent to the neighbor when these conditions are detected. After the message is sent, the BGP transport connection is closed. This means that all resources for the BGP connection are de-allocated. The routing table entries associated with the remote peer are marked as invalid. Finally, other peers are notified that these routes are invalid.

Ref: TCP/IP Tutorials and Technical Overview, IBM Redbooks

BGP Selection

- BGP is a path vector protocol. In path vector routing, the path is expressed in terms of the domains (or confederations) traversed so far.
- The best path is obtained by comparing the number of domains of each feasible route.
- There are no universally agreed-upon metrics that can be used to evaluate external paths.
- Each AS has its own set of criteria for path evaluation.

Path attributes

- Path attributes are used to describe and evaluate a route.
- Peers exchange path attributes along with other routing information.
- When a BGP router advertises a route, it can add or modify the path attributes before advertising the route to a peer.
- The combination of attributes are used to select the best path.

Ref: TCP/IP Tutorials and Technical Overview, IBM Redbooks

BGP Path Attribute

Four path attribute categories:

- Well-known mandatory: The attribute must be recognized by all BGP implementations. It must be sent in every UPDATE message.
- Well-known discretionary: The attribute must be recognized by all BGP implementations. However, it is not required to be sent in every UPDATE message.
- Optional transitive: It is not required that every BGP implementation recognize this type of attribute. A path with an unrecognized optional transitive attribute is accepted and simply forwarded to other BGP peers.
- Optional non-transitive: It is not required that every BGP implementation recognize this type of attribute. These attributes can be ignored and not passed along to other BGP peers.

Ref: TCP/IP Tutorials and Technical Overview, IBM Redbooks

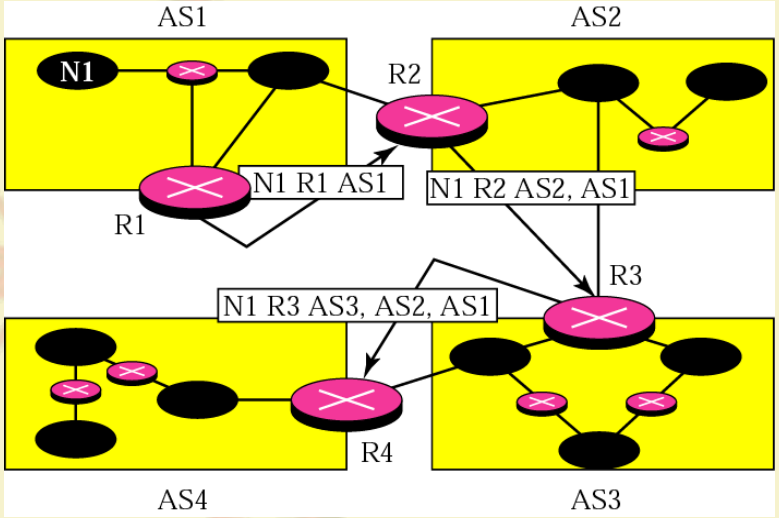
BGP

Example of Network Reachability

Network	Next router	Path
N1	R1	AS14,AS23,AS67
N2	R5	AS22,AS67,AS5,AS89
N3	R6	AS67,AS89,AS9,AS34
N4	R12	AS62,AS2,AS9

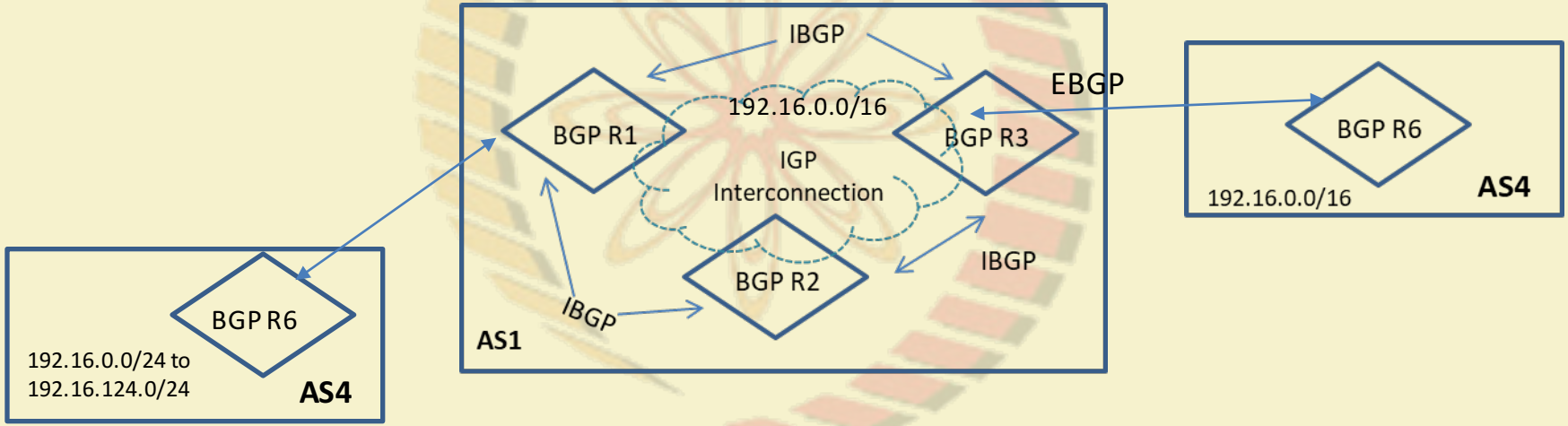
- Loop Prevention in BGP:
  - Checks the Path before updating its database. (If its AS is in the path ignore the message)
- Policy Routing:
  - If a path consist of an AS against the policy of the current AS, message discarded.

Message Advertisements



BGP Aggregation

- The major improvement introduced in BGP Version 4 was support for CIDR and route aggregation.
- This feature allows BGP peers to consolidate multiple contiguous routing entries into a single advertisement.
- It significantly enhances the scalability of BGP into large internetworking environments.



Ref: TCP/IP Tutorials and Technical Overview, IBM Redbooks

BGP Confederations

- BGP requires that all speakers within a single AS have a fully meshed set of IBGP connections.
- This can be a scaling problem in networks containing a large number of IBGP peers.
- A BGP confederation creates a set of autonomous systems that represent a single AS to peers external to the confederation.
- This removes the full mesh requirement and reduces management complexity.

Ref: TCP/IP Tutorials and Technical Overview, IBM Redbooks


thank you!