

Practical Design For Enterprise IP Routing





Keith Bogart

CCIE #4923

✉ kbogart@ine.com
🐦 [@keithbogart1](https://twitter.com/keithbogart1)
in [linkedin.com/in/keith-bogart-2a75042](https://www.linkedin.com/in/keith-bogart-2a75042)



CCIE Routing & Switching

Course Topic Overview

- + Why Good Design Matters
- + The Three Tier Hierarchical Model
- + General Routing Design Tips
- + Best Practices & Design Guidelines For;
 - + EIGRP
 - + OSPF
 - + IS-IS
 - + BGP
 - + Route Redistribution



Learning Objectives:

- Explain The Importance Of Creating A Good Routing Protocol Design
- Select The Best Routing Protocol For Your Environment
- Derive Good Design Decisions For Your Routing Protocol In Order To:
 - Plan For Future Scalability
 - Apply Security Practices
 - Manipulate Path Control Logically & Consistently

****NOTE**** This course does not contain any labs

- + Good Understanding Of IPv4/IPv6 Addressing, Subnetting, & Summarization
- + Protocol Operational Knowledge Of IGPs & BGP
- + Familiarity With The Purpose & Function Of Redistribution

Course Prerequisites



Let's Get Started!



Why A Good Routing Design Matters



Everyone Believes In Good Network Design...Right?

- + There are MANY questions that must be answered prior to thinking about an IP Routing plan.
 - + Unfortunately, many companies never answer those questions.
- + Most large enterprise networks started out as small networks and grew to what they are today.
- + Lack of preparation on someone else's part unfortunately DOES constitute an emergency on your part.



Your Typical Network

- + Many greenfield networks are deployed with little consideration given to planning and design
- + Bad IP routing planning leads to:
 - + Table and database bloat
 - + Increases complexity of troubleshooting
 - + Increases the complexity of implementing packet control via;
 - + Routing manipulation
 - + Security enforcement
 - + QoS implementation
 - + Premature exhaustion of usable host space



- “Greenfield” – A brand new network that you design from the ground up.
- “Brownfield” – An existing network for which you are asked to apply changes or implement new features.
- Network design concepts almost always apply best to greenfield deployments. It can be difficult, if not impossible, to apply network design concepts (related to IP addressing and routing) to existing networks (“Brownfield deployments”)

IGP & BGP Routing Design Benefits

- + Proper design and implementation of a routing protocol yields several benefits:
 - + Facilitates easy route summarization
 - + Makes troubleshooting easier
 - + Minimizes the need for route filtering
 - + Decreases routing table bloat
 - + Provides for future scalability





Thanks for Watching!



The Three Tier Hierarchical Model

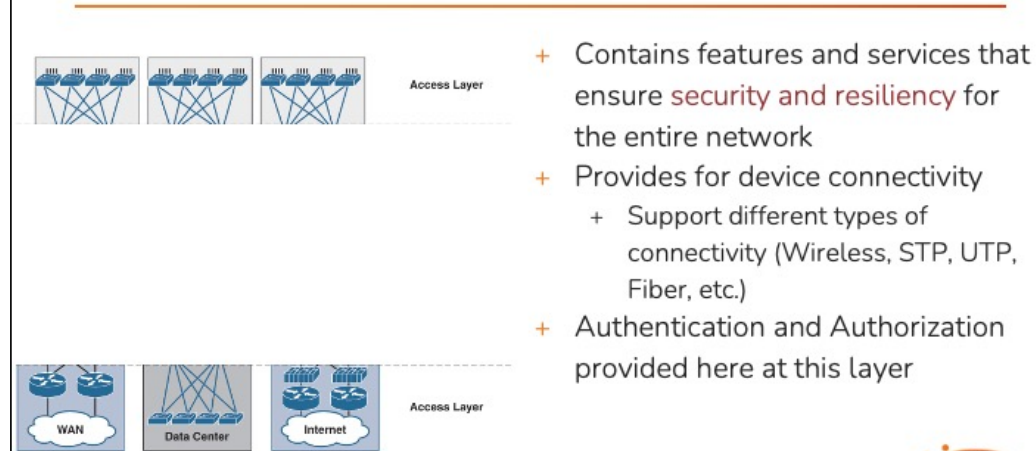
Adhering To Hierarchical Models

- + Good routing design requires implementation of a hierarchy in your network
- + Breaking the design up into layers allows each layer to implement specific functions
 - + Simplifies the network design
 - + Simplifies deployment and management of the network.
- + Modularity in network design allows you to create design elements that can be replicated throughout the network.
- + Replication provides an easy way to scale the network as well as a consistent deployment method.

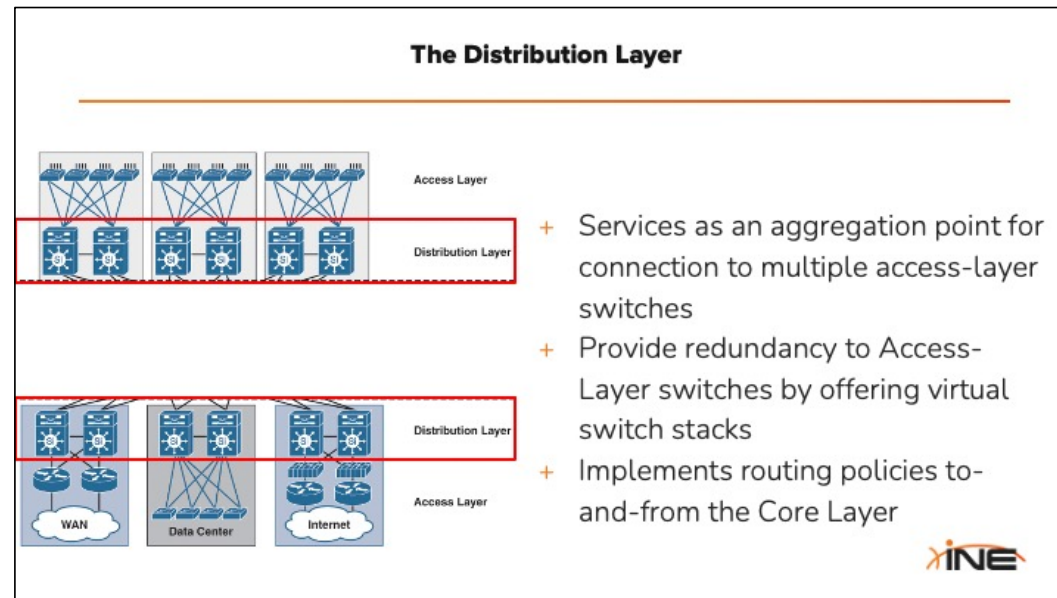


- In flat or meshed network architectures, changes tend to affect a large number of systems. Hierarchical design helps constrain operational changes to a subset of the network.
- ---- Single sites might only require a collapsed core model (Access and Distribution Layers only)
- ---- Campus and large Enterprises with multiple buildings or sites should use all three layers

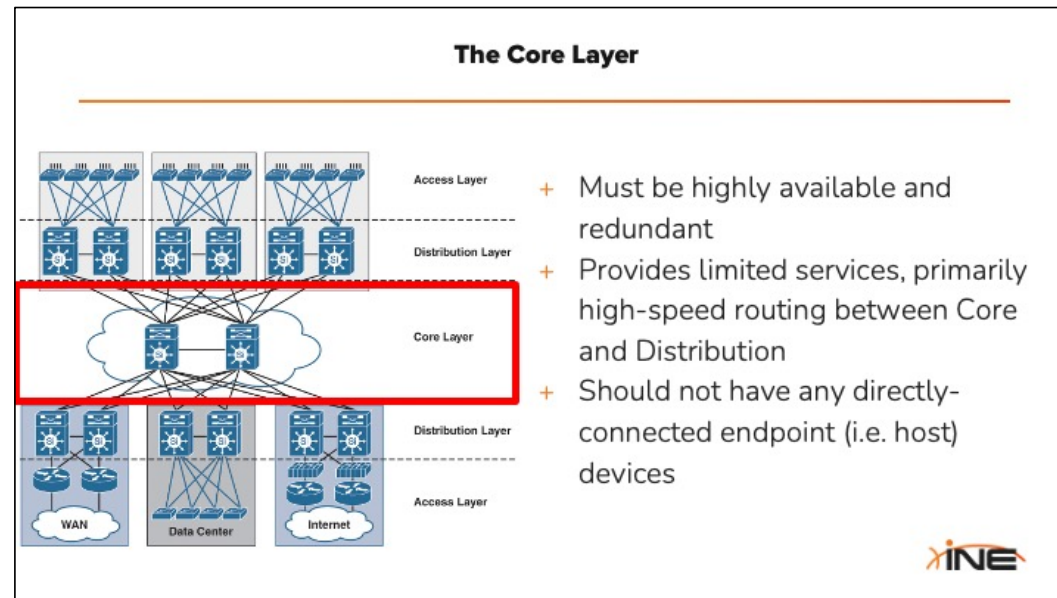
The Access Layer



- The Access Layer also:
- ---Supports PoE
- ---Applies QoS markings and (for LAN-to-LAN traffic across a single Access-Layer switch) QoS Enforcement
- Notice that Layer-3 Routing typically is NOT implemented within the Access Layer



- Distribution Layer is also responsible for enforcing security and authorization policies for traffic coming from one, connected Access block (such as a fileserver connected to an access switch in the “Data Center” block) and destined for a device in another Access Layer block that is connected to the same Distribution Block (ie. Traffic destined for a server connected to Data Center Access-Switch-2).
- The above is referred to as “East-West” traffic because, from source to destination, the traffic never needs to leave the Distribution Layer and be forwarded to the Core.



- Because everything aggregates up to the Core, keep a close eye on oversubscription levels for your traffic and bandwidth.



Thanks for Watching!



General Routing Design Tips

Routing Protocols: Design Objectives

- + Scalability
 - + Reduce the size of routing tables as much as possible
 - + Summarize wherever possible
 - + Implement filtering when routes aren't needed
 - + Implement and utilize backup paths when available
- + Security
 - + Prevent unexpected peerings and adjacencies
 - + Hide the knowledge of your routing protocols
- + IPv6 preparation...even if not using it now.



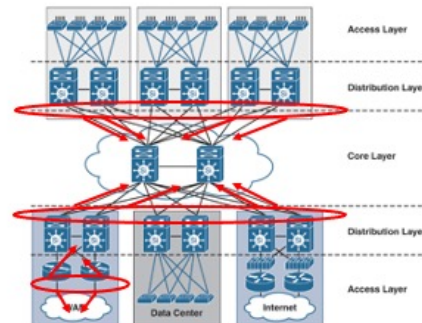
Just Configure It!

- + What happens when someone takes their knowledge of routing configuration and applies it to a network without any forethought or planning?
 - + Increased likelihood of rogue peerings
 - + Leaking of private networks into public domains
 - + Loss of routing upon an IP addressing change
 - + Sub-optimal routing
 - + Routing Loops



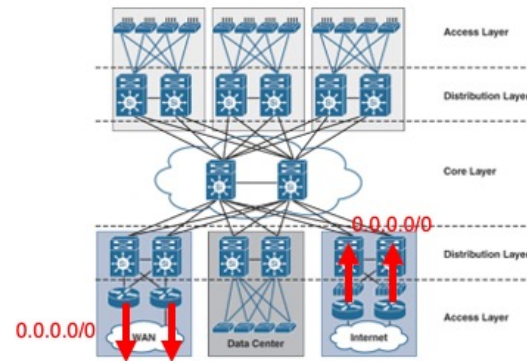
Summarization Points

- + Summarize at the Distribution Layer pointing towards the Core (to reduce Core routing tables)
- + Also summarize at the WAN connectivity points and remote-access points towards the network Core



Default Routing

- + Use default routing as much as possible
 - + WAN Edge and/or Internet-facing routers should originate the default routes



Preventing Unwanted Peers

- + Implement "passive interfaces" on Access-Layer edge interfaces
 - + Prevents unwanted/illegal routing peering on these interfaces.
 - + Prevents unauthorized visibility into your IGP
- + Implement routing protocol authentication on all Layer-3 interfaces that have been activated for routing.





Thanks for Watching!



Selecting A Routing Protocol

The Importance Of Selection

- + When designing for greenfield deployments (or overhauls of brownfields) one must select an IPv4/IPv6 routing protocol
- + Different protocols have different strengths and weaknesses
- + Some protocols were designed for specific purposes



IGP Or EGP?

- + Are you restricted to an IGP or EGP?
 - + IGP = Interior Gateway Protocol (RIP, EIGRP, OSPF, IS-IS)
 - + EGP = Exterior Gateway Protocol (BGP)
- + BGP requires an underlying IGP (for next-hop reachability)
 - + Unless peering with directly-connected neighbors
- + IGPs require no underlying protocols



- Although initially the configuration is more complex and a good design HAS to be planned, using BGP within your Enterprise offers several benefits:
<https://networkdirection.net/articles/routingandswitching/bgpintheenterprise/>

Contrasting IGP

	EIGRP	OSPF	IS-IS
Proprietary or Vendor-Agnostic?	Proprietary	Vendor-Agnostic	Vendor-Agnostic
Hierarchy Required In Design?	No	Not required but recommended	Not required but recommended
Summarization & Filtering Restrictions?	No restrictions	Yes...only ABRs and ASBRs can perform summarization/filtering	Yes...only ABRs and ASBRs can perform summarization/filtering
Supports dynamic generation of default routes?	No	Yes, with Stub areas	Yes, when using multiple areas
Resource consumption (CPU, memory, etc)	Low	High	Medium



- A non-proprietary protocol can also be important if you will be implementing automation. Many non-Cisco automation platforms only work with OSPF and not EIGRP.
- EIGRP is a flat (non-hierarchical) IGP protocol which means: Without proper summarization, route flaps could be felt everywhere in the network (Queries)

Contrasting IGP

	EIGRP	OSPF	IS-IS
Supports equal-cost multipath?	Yes	Yes	Yes
Support unequal cost multipath?	Yes	No	No
Metric accurately reflects topology?	Yes (BW and Delay...can also enable Reliability and Load)	Yes (BW)	No (Static value)
Learning Curve	Low	High	High
Supports unequal timers (Hello and Hold) between neighbors?	Yes	No	Yes



- EIGRP Variance (can be a quick fix for congested links rather than implementing complex QoS policies)
- While it is not recommended that timers be different between IGP peers, this can be useful on slow-speed WANs in which it is known that transmissions in one direction have a definite delay that is higher-or-lower than transmissions in the reverse direction (especially with multiaccess segments containing three-or-more routers)

Other IGP Differences

- + EIGRP: Enabled on all IPv6 interfaces by default when configured in named-mode.
- + EIGRP converges more quickly than OSPF
<https://scialert.net/fulltext/?doi=aujcs.2014.1.8>
- + Both EIGRP and OSPF support forming neighbors with unicast Hellos (for security purposes)
 - + IS-IS doesn't use IP for Hello packets
 - + IS-IS sets the "group" bit in the Hello packet destination MAC which results in switch flooding





Thanks for Watching!



Best Practices For EIGRP Route Tuning

What Is EIGRP Route Tuning?

- + EIGRP route tuning is the process of applying adjustments to your EIGRP configuration (in one or more routers) so that traffic to certain destinations takes predefined paths.
- + Numerous ways to accomplish this:
 - + Manipulating EIGRP metrics
 - + Implementing creative route summarization
 - + Using EIGRP variance



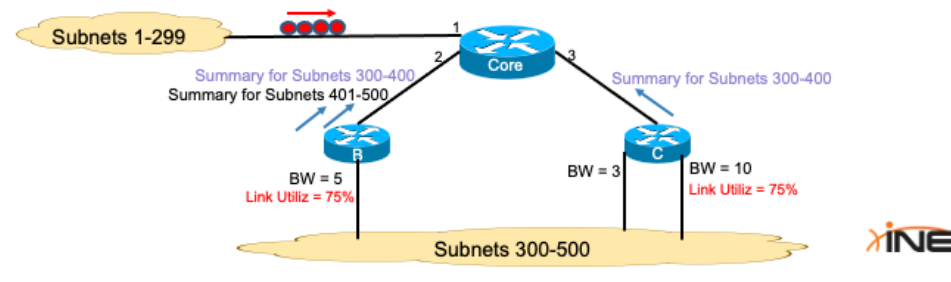
The Goals Of Good Route Tuning

- + Ensure changes made to EIGRP don't affect other protocols or features
- + Ensure summarized routes are used as intended
- + Spread traffic among available paths to avoid congestion points



Route Tuning Best Practices

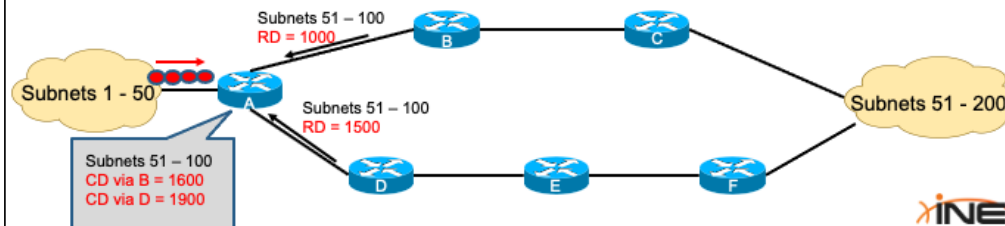
- + Only adjust delay...not bandwidth
- + If summarization is used to perform path manipulation, utilize the "summary-metric" command to give the summary route a consistent and unchangeable metric.



- Without manually setting the metric of a summarized route, the best metric of all subset routes will be used...which could change if that subset route goes away.
- In this diagram, the network policy is that the Core router should ALWAYS forward packets through Rtr-C if ANY route exists for subnets 300-400.

Route Tuning Best Practices

- + Variance review
 - + Set a Variance of at least "2" on all routers so that each router can take advantage of load-balancing and not overload single links.





Thanks for Watching!



EIGRP Scalability Best Practices

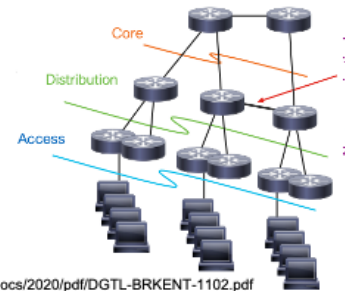
Designing For EIGRP Scalability

- + A good EIGRP design ensures that EIGRP can grow as your network grows.
- + Poorly designed EIGRP networks will experience:
 - + Routing and forwarding table bloat
 - + EIGRP Stuck in active situations
 - + Insufficient resources to handle massive EIGRP tables
- + The EIGRP “Active” process should be limited in scope



EIGRP Scalability Best Practices

- + Summarize as much as possible, in the appropriate places:
 - + Do NOT use auto-summarization but use manual summarization instead.
 - + Summarize up-and-down layers of the design model, not between adjacent layers



Graphic courtesy of
<https://www.ciscoplive.com/c/dam/r/ciscoplive/us/docs/2020/pdf/DGTL-BRKENT-1102.pdf>



- When you implement summarization between routers that are at the same level of the model (for example, between two connected routers at the Distribution Layer) it can make troubleshooting difficult and lead to other problems.

EIGRP Scalability Best Practices

- + Prevent EIGRP Queries from propagating over WAN links.
 - + Utilize EIGRP Stub and Summarization wherever possible to prevent this.
- + Unless all routers need to know about all routes, utilize filtering to also reduce routing tables and query scope.





Thanks for Watching!



EIGRP Security Guidelines

The Goals Of Securing EIGRP

- + Prevent unwanted, rogue or malicious peerings
- + Prevent the accidental leakage of routes
- + Prevent the poisoning of the IP Routing Table
- + Hide EIGRP protocol information from end-users

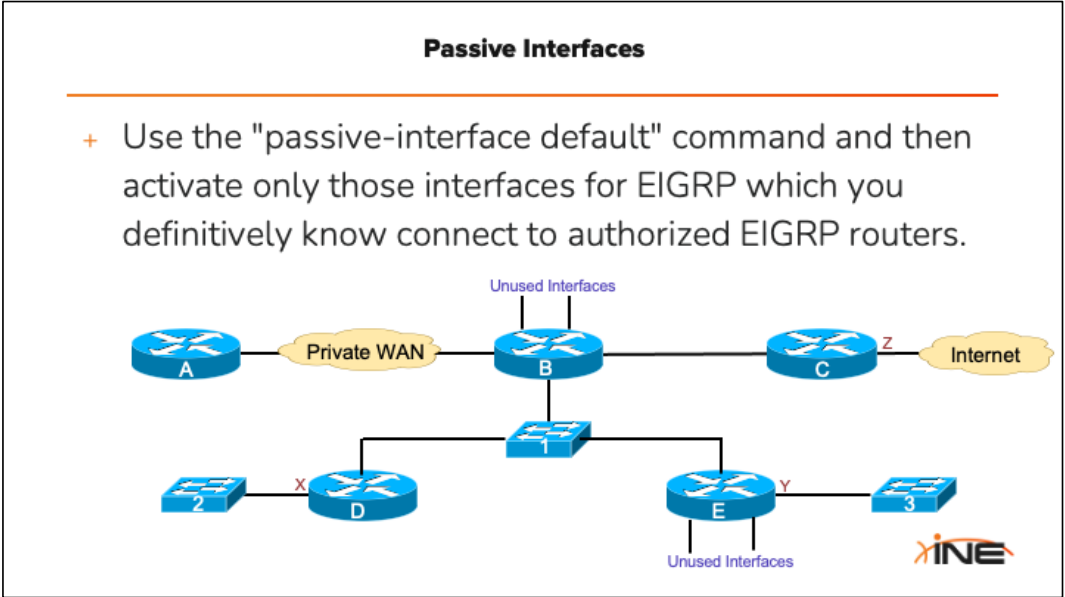


Passive Interfaces

- + Use the "passive-interface default" command and then activate only those interfaces for EIGRP which you definitively know connect to authorized EIGRP routers.

The diagram shows a network topology with five routers (A, B, C, D, E) and three switches (1, 2, 3). Router A is connected to Router B via a 'Private WAN' link. Router B is connected to Router C via a link labeled 'Z'. Router C is connected to the 'Internet'. Router B is also connected to a central switch labeled '1'. Router D is connected to switch '1' via a link labeled 'X'. Router E is connected to switch '1' via a link labeled 'Y'. Router E is also connected to switch '3'. Router A has two 'Unused Interfaces' indicated by lines pointing to the top of the router. Router B has two 'Unused Interfaces' indicated by lines pointing to the top of the router. Router D has one 'Unused Interface' indicated by a line pointing to the bottom of the router. Router E has two 'Unused Interfaces' indicated by lines pointing to the bottom of the router. The switches are labeled '2' and '3' respectively. The 'Internet' is represented by a cloud icon. The logo 'INE' is in the bottom right corner.

- ## Passive Interfaces
- + Use the "passive-interface default" command and then activate only those interfaces for EIGRP which you definitively know connect to authorized EIGRP routers.
-
- The diagram shows a network topology with five routers (A, B, C, D, E) and three switches (1, 2, 3). Router A is connected to Router B via a 'Private WAN' link. Router B is connected to Router C via a link labeled 'Z'. Router C is connected to the 'Internet'. Router B is also connected to a central switch labeled '1'. Router D is connected to switch '1' via a link labeled 'X'. Router E is connected to switch '1' via a link labeled 'Y'. Router E is also connected to switch '3'. Router A has two 'Unused Interfaces' indicated by lines pointing to the top of the router. Router B has two 'Unused Interfaces' indicated by lines pointing to the top of the router. Router D has one 'Unused Interface' indicated by a line pointing to the bottom of the router. Router E has two 'Unused Interfaces' indicated by lines pointing to the bottom of the router. The switches are labeled '2' and '3' respectively. The 'Internet' is represented by a cloud icon. The logo 'INE' is in the bottom right corner.



Controlling EIGRP Route Advertisement

- + For IPv4, configure specific EIGRP “network” commands utilizing specific wildcard masks.
- + Prevents private networks from accidentally being advertised.



Bad!!

```
Router eigrp INE
address-family ipv4 autonomous-system 20
network 10.0.0.0
```

Better

```
Router eigrp INE
address-family ipv4 autonomous-system 20
network 10.10.2.0 0.0.0.255
```

Best!!

```
Router eigrp INE
address-family ipv4 autonomous-system 20
network 10.10.2.0 0.0.0.3
network 10.10.2.4 0.0.0.3
```

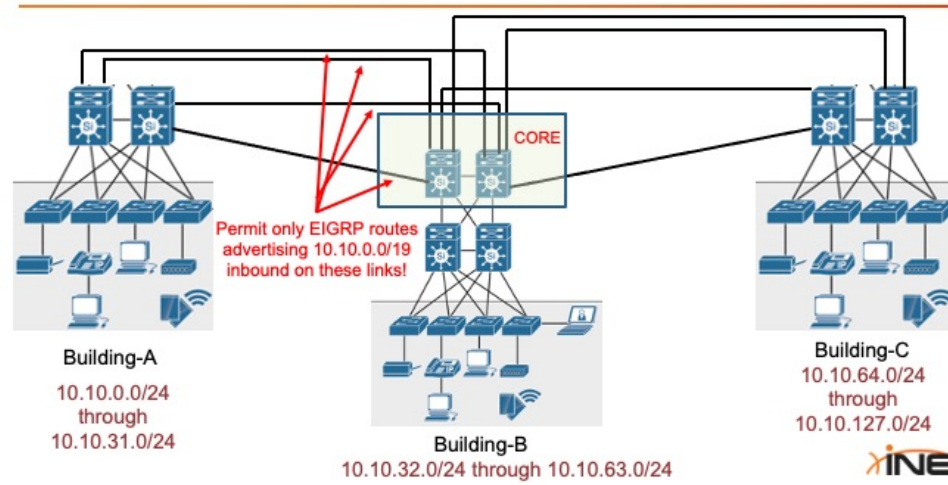


Using EIGRP Filters For Security

- + Within Core routers, implement inbound EIGRP filters to allow only known internal routes.
- + This will prevent rogue routers from injecting unauthorized and false routes which could overwhelm routing tables.

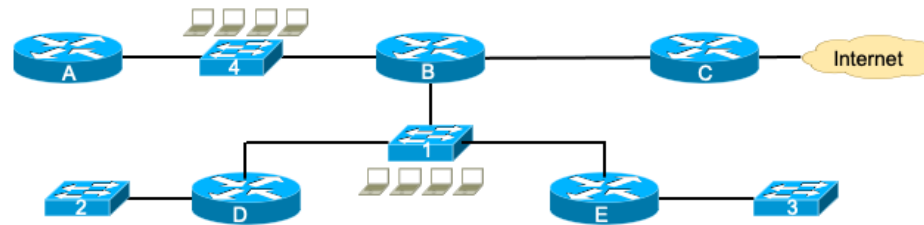


Filter Example



Unicast EIGRP Neighbors

- + Use the EIGRP "neighbor" command to form unicast peerings for security purposes.





Thanks for Watching!



Other General EIGRP Design Guidelines

EIGRP Timers

- + If possible, ensure all routers are using the same EIGRP Hello and Hold timers
- + Do not reduce the Hello-timer below 2-seconds otherwise links experiencing delays or congestion could cause dropped neighbors.
 - + Utilize BFD if you wish to have sub-second dead-intervals



- If possible, ensure all routers are using the same EIGRP Hello and Hold timers (otherwise, peers with mismatched timers may time-out and declare each other dead).
- Remember that the command, "ip hold-time eigrp" does not set your OWN LOCAL hold-time but rather advertises to your peer what HIS hold-time should be for your transmitted Hello packets).

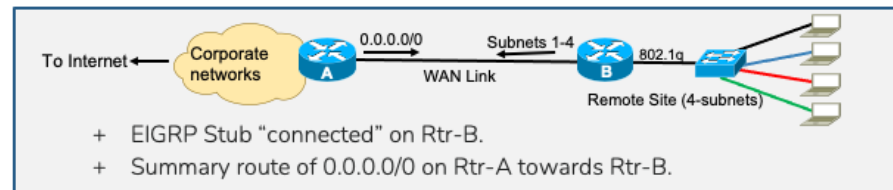
Use Named-Mode EIGRP

- + Configure EIGRP in named-mode to provide:
 - + Access to wide metrics
 - + Access to more complex authentication methods (HMAC with SHA)
 - + Access to protocol families (for IPv4, IPv6, VRF, etc configuration)
 - + Automatically activates EIGRP for IPv6 on all IPv6-enabled interfaces



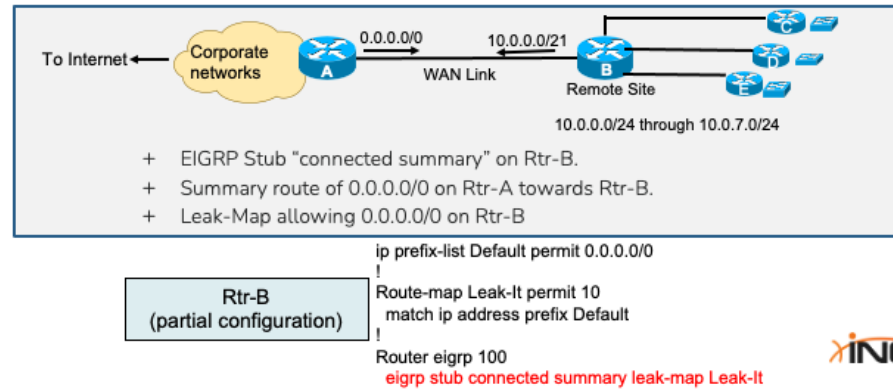
EIGRP Stubs

- + Configure remote sites with the EIGRP “stub” feature to limit EIGRP Query scope.



EIGRP Leak-Maps

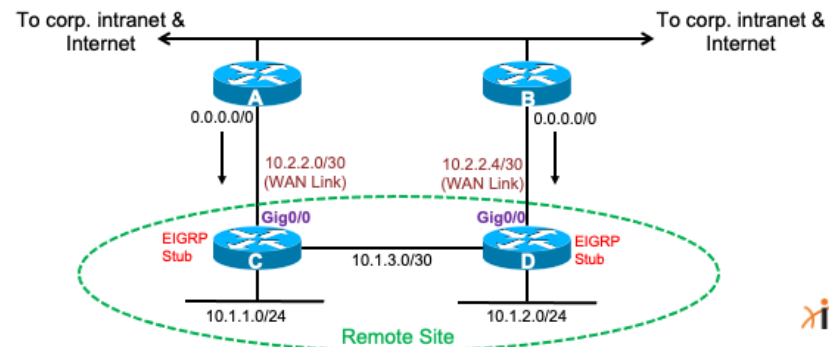
- + Utilize Leak-Maps when specific EIGRP (internal) routes need to be leaked into (or out of) Stub sites



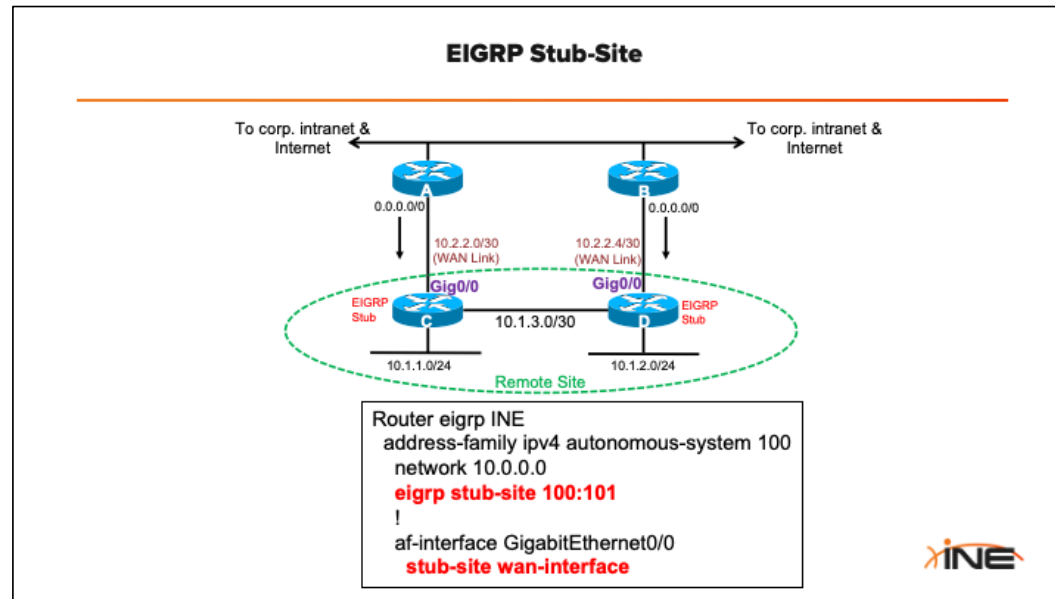
- Leak-maps can also be utilized within the Enterprise when you need to leak specific subnets that would normally be suppressed via summarization.

EIGRP Stub-Site

- + Alternatively, use the EIGRP “stub-site” feature to avoid using leak-maps



- Normally, if "C" and "D" were configured with EIGRP Stub then neither router would pass along the default route to the other. They would also not pass along each other's WAN subnets or their local 10.1.x.x/24 subnets. Leak-maps would be required for this.



- When configured as a “stub-site”, any routes that “C” or “D” learn over the interface identified as the “wan-interface” get tagged (inbound) with a special EIGRP community that contains the site-id (100:101 in this case) and then are allowed to be forwarded to any stub-site peers.
- The stub-site peer will install the route but will NOT advertise the route outbound on any of its “wan-interface(s)”.
- Any other routes covered by “network” or “summary-address” statements on “C” or “D” will be advertised to each other as well as advertised on the WAN link to “A” and “B”.
- So no leak-maps are required yet you still get the benefit of EIGRP Stub in that routers “A” and “B” will never forward EIGRP Queries to “C” or “D”.
- Also ensures that “A” and “B” will never use the Remote Stub Site as a transit path.

EIGRP K-Values

- + Some design documents recommend that you disable the K-Value applied to “bandwidth” leaving only the “delay” K-value active.
- + Others recommend that you enhance the default K-Values by enabling the K-Values used against “load” and “reliability” to obtain a more accurate EIGRP distance value.
- + No matter what you do, remember that K-Values must match between EIGRP neighbors.



Use The EIGRP Event Log

- + Before enabling any EIGRP debugs, try searching through the EIGRP event log!
 - + Displays the most recent 500 EIGRP events
 - + Displays most recent events first

```
R3#sho ip eigrp events
Event information for AS 100:
1  20:58:24.316 Poison squashed: 0.0.0.0/0 reverse
2  20:58:24.290 Change queue emptied, entries: 1
3  20:58:24.290 Metric set: 0.0.0.0/0 metric(1966080)
4  20:58:24.290 Update reason, delay: new if delay(Infinity)
5  20:58:24.290 Update sent, RD: 0.0.0.0/0 metric(Infinity)
6  20:58:24.290 Update reason, delay: metric chg delay(Infinity)
7  20:58:24.290 Update sent, RD: 0.0.0.0/0 metric(Infinity)
8  20:58:24.290 Route installed: 0.0.0.0/0 13.13.13.1
9  20:58:24.290 Route installing: 0.0.0.0/0 13.13.13.1
```





Thanks for Watching!



Best Practices For OSPF Path Manipulation

The Goals Of Good OSPF Path Manipulation

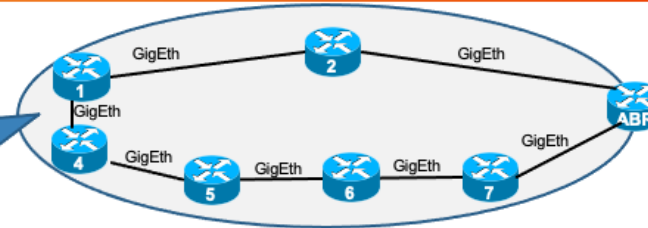
- + Ensure changes made to OSPF don't affect other protocols or features
- + Don't sacrifice OSPF scalability for path manipulation
- + Spread traffic among available paths to avoid congestion points
- + Don't allow traffic paths to become unpredictable when adding redundancy



Manipulating Intra-Area Routes

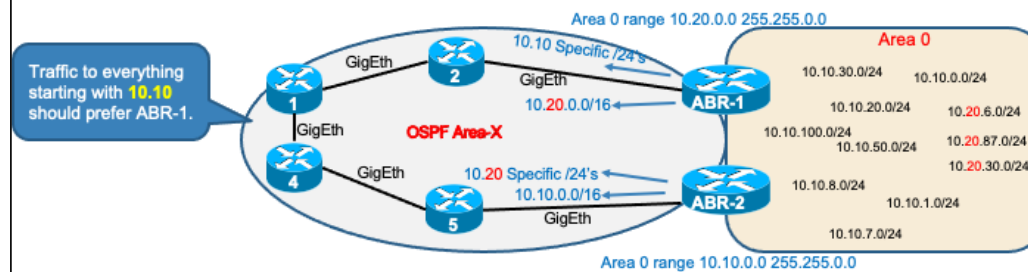
All traffic from R1 and R4 to the ABR is taking the "1-2-ABR" path. This path is congested.

Objective: Make R4 prefer the "5-6-7-ABR" Path.



- + Summarization and filtering not possible for intra area routes.
- + When manipulating intra-area traffic paths, influence the OSPF metric using the "*ip ospf cost*" command on interfaces rather than modifying "bandwidth"

Manipulating Inter-Area Routes

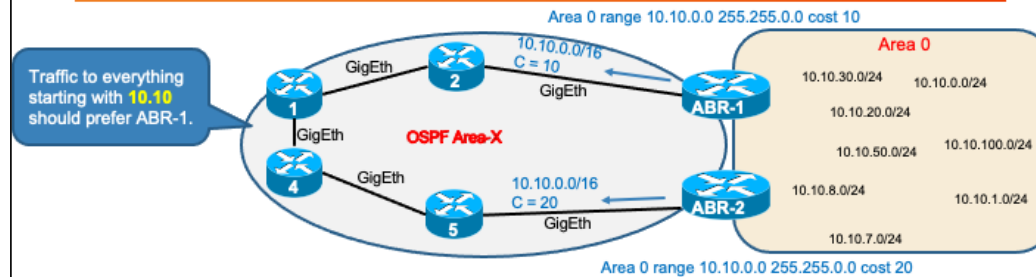


- + When manipulating inter-area traffic paths (with two-or-more ABRs connecting Area-X to the backbone);
 - + Utilize OSPF summarization as much as possible on the ABRs;
 - + Longest match rule will avoid summarized prefix



- Using this method, the internal cost to reach the ABRs are irrelevant as /24's will always be selected over a matching /16.

Manipulating Inter-Area Routes



- + It is more scalable to have both ABRs summarize
 - + Path determination manipulated by cost
 - + Summarized prefix cost should be manually set so it is deterministic
 - + area x range y.y.y.y z.z.z.z cost <cost>



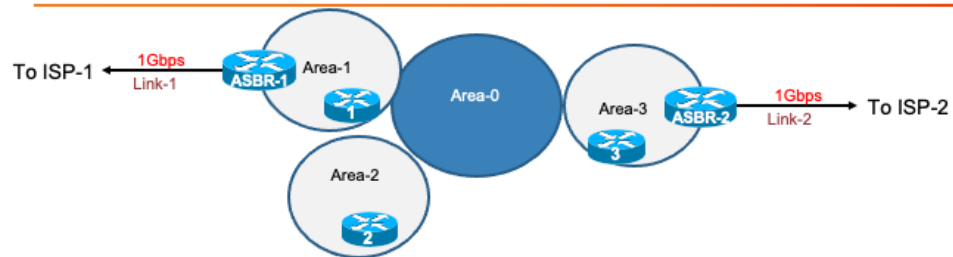
- While this method is more scalable as it follows the general rule of "summarize as much as you can, wherever you can" it is less deterministic than the previous method. If paths change within Area-X (or new paths are introduced) the cumulative cost might become lower for internal routers to reach ABR-2 instead of ABR-1.

Manipulating External Routes

- + Ideally, an ASBR would generate a default route and avoid redistribution.
- + If redistribution is required, utilize OSPF external metric types instead of cost to influence external path selection
 - + O E1 is always preferred over O E2



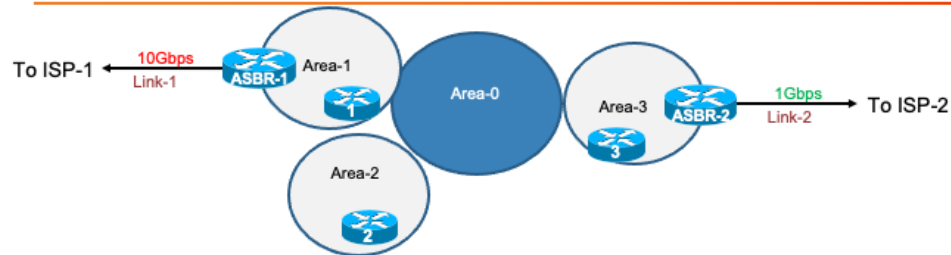
Redundant Externals: Same Link Speed



If	Then	How (ASBR1)	How (ASBR2)	Cost Comparison
External ISP links are equal	Internal costs to reach external networks is more important than external link costs . Send packets to closest ASBR.	Redistribute as O E1 s (cost = x)	Redistribute as O E1 s (cost = x)	X + internal cost to ASBR (closest ASBR wins)

- OSPF RFC 2328 (in section 2.3) identifies why External Type-1 and External Type-2 were created.
- External Type-1's are to be used when the path THROUGH YOUR OWN OSPF DOMAIN is of more concern to you than the path packets take once they LEAVE your domain.

Redundant Externals: Different Link Speeds (Option-A)

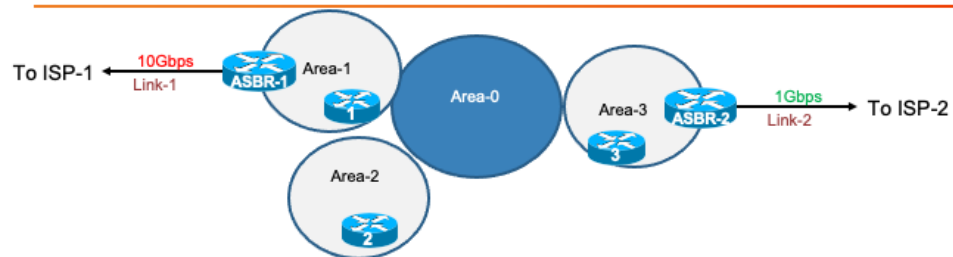


If	Then	How (ASBR1)	How (ASBR2)	Cost Comparison
External ISP Link-1 is the preferred link (over Link-2)	Send packets to ASBR1 which is connected to Link-1	Redistribute as O E2s (cost = X)	Redistribute as O E2s (cost = Y)	X < Y Internal cost irrelevant (All packets forwarded to ASBR1)



- External Type-2's (according to RFC 2328) are to be used when the path packets take once they LEAVE your domain is of MORE IMPORTANCE than the path packets take through your own OSPF domain.
- The downside to this approach is that the route that each router has will always show up as "x" as a cost in the routing table and OSPF LSDB.
- So this gives you zero visibility into the path that is actually being selected in your network to reach ASBR-1.
- If links are flapping within your network, and congestion is occurring because packets are being forwarded on non-optimal links to reach ASBR1, you will have no visibility into this by simply viewing the OSPF cost value of your routes.

Redundant Externals: Different Link Speeds (Option-B)



If	Then	How (ASBR1)	How (ASBR2)	Cost Comparison
External ISP Link-1 is the preferred link (over Link-2)	Send packets to ASBR1 which is connected to Link-1	Redistribute as O E1s (cost = X)	Redistribute as O E2s (cost = X)	X = X (initially) O E1 > O E2 ASBR1 always preferred



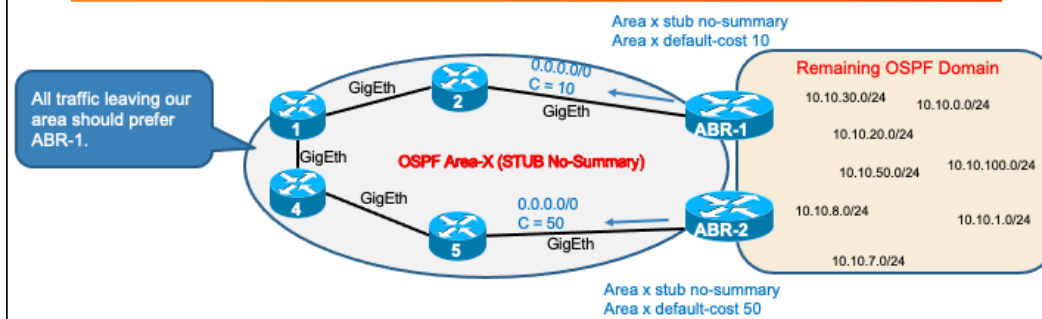
- With this approach you still meet your objective of having all routers prefer ASBR-1 because O E1 is always preferred over O E2 (regardless of cost)
- However, each router's routing table will now show you a cost value that more accurately reflects the internal cost to reach ASBR1 and this can be tracked.
- Now, if internal paths to reach ASBR1 change, you can easily see this as a new/different costs reflected in your external routes.

Influencing ABR Selection In Stub Areas

- + By default, ABRs create a dynamic default route when connected to Stub areas.
- + If two ABRs exist, both will create the default route and both set it to the same cost value.
 - + To influence the cost value of the default-route that is generated into Stub areas use command,
 - + "*area x default-cost y*"
 - + Useful if you want all routers in the Stub area to only prefer ABR-1, and only use ABR-2 if ABR-1 goes down.



Influencing ABR Selection In Stub Areas



- + Be aware that the cost to reach each ABR is still a factor here and will be added to whatever cost the ABRs are advertising within their default LSA.



Thanks for Watching!



Ensuring OSPF Scalability



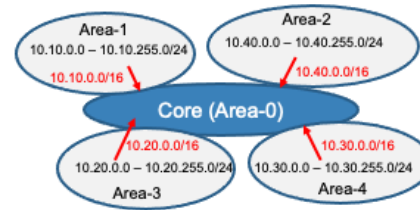
Designing For OSPF Scalability

- + Reduce the size of OSPF databases and routing tables by;
 - + Liberal use of summarization
 - + Effective design of OSPF areas
 - + Use of OSPF Stub areas
 - + Logical placement of OSPF ABRs



Implement Summarization

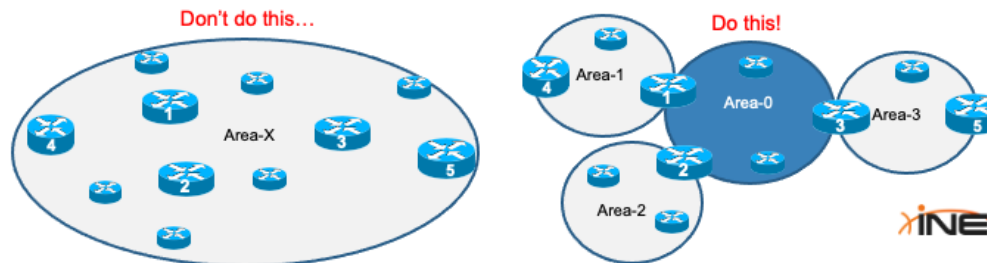
- + Summarization is not only useful for traffic manipulation, but also a powerful tool to ensure OSPF scalability
- + OSPF summarization:
 - + Should be done on ABRs
 - + At the Distribution Level
 - + Towards the Core



- Ensure that a matching summary route to Null0 exists to prevent routing loops (this should happen by default but be sure to doublecheck).

Utilize Areas

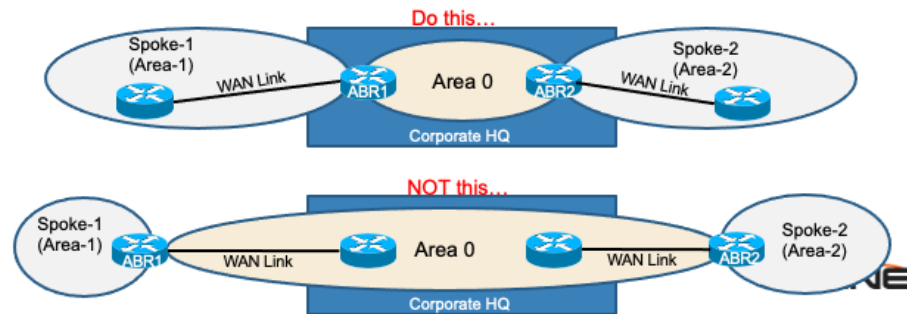
- + Segment your topology into multiple areas. Benefits include:
 - + Reduces quantity of SPF calculations
 - + Reduces LSA flooding
 - + Allows for summarization and filtering at ABRs



Hub & Spoke Topologies

+ For hub-and-spoke topologies:

- + Each spoke should reside in its own unique area
- + ABRs should be located within the hub site (not at a spoke site)
- + Configure spoke sites as Totally Not-So-Stubby Areas



- Scalability means limiting the size of OSPF LSDBs...and restricting how far LSAs can travel.
- In this scenario, scalability involves eliminating as much as possible the OSPF traffic flowing across the WAN links.



Thanks for Watching!



OSPF Security Guidelines

The Goals Of Securing OSPF

- + Prevent unwanted, rogue or malicious adjacencies
- + Prevent the accidental leakage of routes
- + Hide OSPF information from end-users



Preventing Rogue Adjacencies

- + Utilize authentication between every pair of adjacencies
 - + Do not reuse the same password on all authentications
 - + Use the most secure form of authentication possible (HMAC-SHA)
- + Use "passive-interface default" and then activate only known, critical interfaces.



Controlling Route Advertisement

- + Ensure that OSPF "network" statements are as precise as possible to prevent OSPF from activating on unwanted interfaces
- + Alternatively, implement OSPF with interface-level "ip ospf" commands



Bad!!

```
Router ospf 1
network 10.0.0.0 0.255.255.255 area 1
```

Better

```
Router ospf 1
network 10.10.2.0 0.0.0.3 area 1
network 10.10.2.4 0.0.0.3 area 1
```

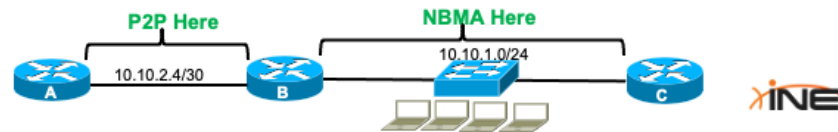
Best!!

```
Interface GigabitEthernet0
ip ospf 1 area 1
Interface GigabitEthernet1
ip ospf 1 area 1
```



OSPF Network Types

- + When appropriate, consider changing the OSPF network type to NBMA and implementing the OSPF "neighbor" command to form unicast peerings for security purposes.
- + If security isn't a concern, set the network-type to "point-to-point" to prevent DR/BDR elections, speedup formation of adjacencies, reduce OSPF LSA generation, and (Slightly) reduce OSPF compute requirements.



- Another benefit of P2P network type is that it eliminates Type-2 LSAs thus reducing the overall size of OSPF LSDBs.



Thanks for Watching!



Other OSPF Guidelines

Use OSPFv3

- + For greenfield environments, implement OSPFv3 instead of OSPFv2
 - + Facilitates an easier transition to IPv6
 - + OSPFv3 can carry both IPv4 and IPv6 prefixes
 - + OSPFv3 supports full IPsec authentication AND encryption
- + Even if you don't have a current IPv6 addressing scheme, enable it on all devices and interfaces to at least obtain Link-Local addresses



- LIMITATION: OSPFv3 (Cisco) doesn't currently support virtual-links
- If you suspect that an area may (in the future) contain an edge router that must connect to a non-contiguous OSPF area that would normally require a virtual link (eg. Partner, Acquisition, Merger, etc) create a secondary OSPF process on the edge router and redistribute between the OSPF processes.
- This allows you to retain the benefits of OSPFv3

Other Scalability Recommendations

- + When using OSPF always update reference bandwidth on all routers to 100G
auto-cost reference-bandwidth 100000
- + Try to constrain areas to a maximum of 40-50 routers
 - + This largely depends on the types of routers deployed
 - + More powerful routers means an area can have more of them in quantity.



Guidelines To Simplify Troubleshooting

- + Set OSPF router-IDs manually to predictable and meaningful values
- + Use the same OSPF process-ID on all routers to avoid confusion.



- Remember that NSSA areas with two-or-more ABRs connecting NSSA AREA-X to Normal Area-Y, only ABR with highest router-ID will translate from Type-7's to Type-5's.

Recommended Area Types

- + When possible, make all non-backbone areas into NSSA or Totally NSSA areas
 - + Same benefits as Stub or Totally Stubby Areas
 - + Contains the additional benefit of allowing you to add a local ASBR in the future if you need to.
 - + Remember that ABRs connected to NSSA do NOT automatically generate dynamic default routes.





Thanks for Watching!



IS-IS Review

A Brief Review Of IS-IS

- + Link State IGP
- + The entire router chassis is in a single area
- + Two types of IS-IS adjacencies
 - + Level-1
 - + Level-2
- + Level-2 routers can form adjacencies with other Level-2 routers in different areas



- As a Link State protocol, this means that all routers within an area must share a common IS-IS LSP database.

A Brief Review Of IS-IS

- + Level-1 routers only know about:
 - + Prefixes in their own area
 - + A default route
- + Level-2 routers know about all prefixes by default
- + Inter-area prefixes must be exchanged along the IS-IS backbone
 - + Contiguous string of Level-2 routers
- + Summarization and filtering only performed when creating Level-2 LSPs between areas.





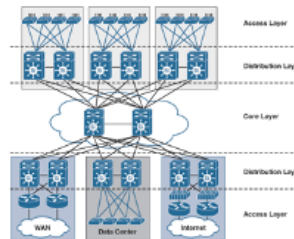
Thanks for Watching!



IS-IS Area Determination

Maintain The Three Layer Hierarchy

- + Networks which will run IS-IS should still follow the 3-layer hierarchy if possible (Access, Distribution, Core)
- + Networks should still be designed with a good IP addressing scheme



IS-IS Flat Topologies

- + IS-IS isn't nearly as complex as OSPF when it comes to building adjacencies or the propagation of topology and prefix information
 - + For this reason, IS-IS is much more tolerant of large areas with many routers (i.e. "nodes") than OSPF
 - + Many companies choose to design IS-IS with a simple flat topology (all routers in a single area) for this reason.



Designing For IS-IS Flat Topologies

- + When starting with a smaller network, all routers can be placed into a single area
- + Cisco routers are Level-1/Level-2 by default, which means they automatically create two databases and two SPF trees
 - + This consumes resources
 - + This may become problematic as the network scales



Challenges With IS-IS Flat Topologies

- + Problems with IS-IS flat topologies
 - + Doesn't allow for any kind of route aggregation or filtering
 - + Instabilities in the network are felt by all nodes
- + Initial IS-IS design should start with determining where the IS-IS backbone will be.



Determining IS-IS Levels

- + If the entire group of current routers are in a single geographic area connected by high-speed links, then configure them as Level-2 only routers.
 - + This will become your IS-IS "backbone" upon which you will build as you expand
- + Modify the LSP flooding/refresh intervals to reduce CPU load on routers

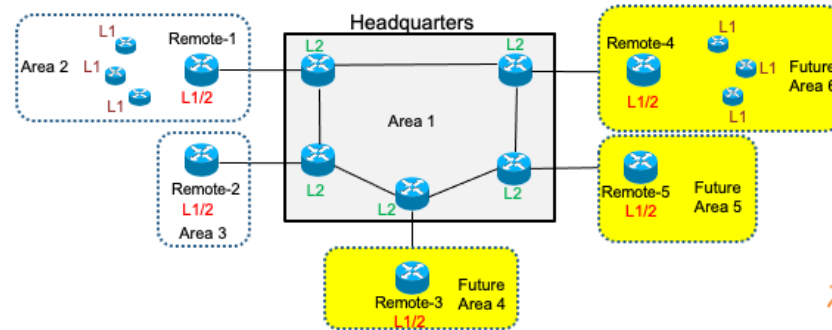
```
router isis
  lsp-refresh-interval 65000
  max-lsp-lifetime 65535
```



- max-lsp-lifetime 65535 --> maximum value (approximately 18.2 hours)

Multi-Area IS-IS

- + If the existing set of routers is geographically dispersed (or separated by slow-speed connections) then multi-area IS-IS is more appropriate:



- Determine the set of routers that will become your IS-IS backbone and configure these as Level-2 only
- Configure P2P routers at the remote ends of WAN connections as Level-1-2. These will become your aggregation points
- Future routers added to existing remote sites should be configured as Level-1 only.
- With this design, prefix summarization can be performed on the L1/L2 routers into the Core (Headquarters).



Thanks for Watching!



IS-IS Scalability Design

Scalability Factors

- + As with any IGP, a good design must account for scalability of the network. The main factors that contribute to network scalability are as follows:
 - + Number of routers
 - + Number of links
 - + Number of internal and external routes
 - + Stability of links
 - + Flooding
 - + Memory
 - + Processing capacity (CPU)



IS-IS SPF Calculations

- + SPF algorithm used to compute a shortest-path tree to every destination prefix
- + Good IS-IS design minimizes the size of the SPF tree and how frequently it must be recomputed.
- + The greatest limitation to IS-IS scalability is excessive LSP flooding
 - + Like OSPFv3, only changes in the topology (i.e., link additions or subtractions, new routers introduced, etc.) will affect SPF calculation, NOT changes to IP addressing.



How Do We Reduce IS-IS LSP Flooding?

- + Like OSPF, IS-IS routers sharing a common area must have common databases
 - + Level-1-only routers have a single Level-1 LSP database
 - + Level-2-only routers have a single Level-2 LSP database
 - + Level-1-2 routers have TWO databases and must keep track of TWO SPF trees
- + Minimizing LSP flooding necessitates the following:
 - + Dividing an IS-IS routing domain into multiple areas
 - + Utilizing Summarization and/or Filtering to limit LSP creation and propagation



Summarization Points

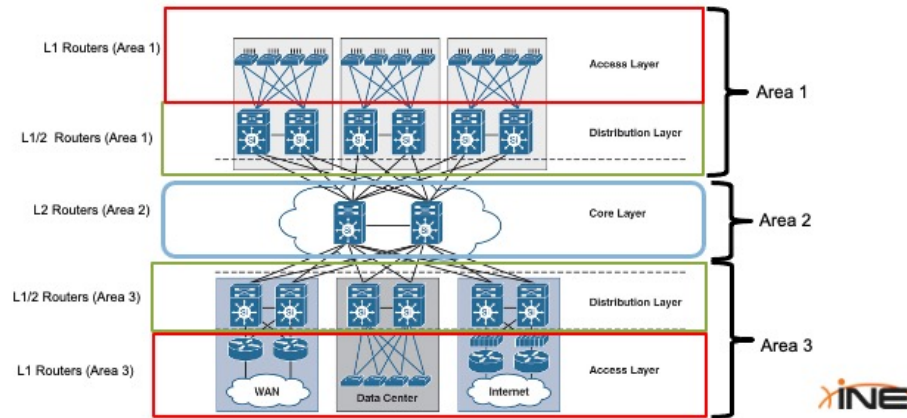
- + Summarization should still be performed at the Distribution Layer
 - + Towards the Core (to reduce the size of routing tables in the Core)
 - + Towards the Access Layer (to reduce the size of routing tables in this layer)
- + Summarization can only be performed at area boundaries



- IS-IS uses the following command for summarization: summary-address <summarized prefix> <dotted-decimal mask>

Summarization Points

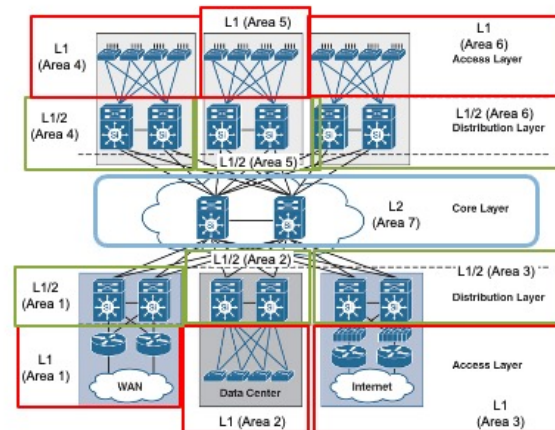
+ Strategy #1 (minimal IS-IS areas)



- Remember that IS-IS L1 routers only create a default route to their nearest L1/L2 router IF that router has generated an LSP with the “Attached” bit set. And that bit is ONLY set when a L1/L2 router is “attached” to two or more areas.
- An L1/L2 router (or a Level-2 only router) that has formed an adjacency with another L1/L2 router (or an L2-only router) in another area is considered an ABR (Area Border Router).

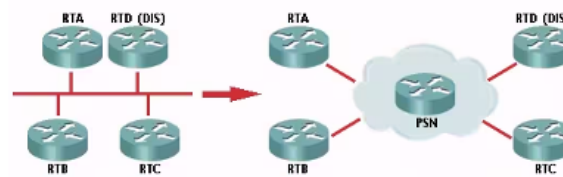
Summarization Points

+ Strategy #2 (several IS-IS areas)



Eliminate The Pseudo Node!

- + IS-IS only has two network types
 - + Broadcast
 - + Point-to-Point
- + Broadcast networks elect one router as the DIS (Designated Information System) which creates the IS-IS Pseudo Node



Graphic courtesy of Cisco, "Understanding IS-IS Pseudonode LSP"

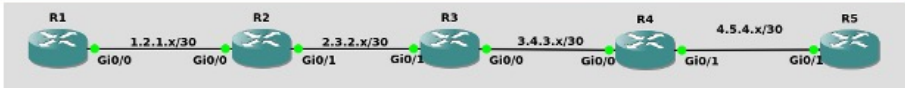


Eliminate The Pseudo Node!

- + Pseudo Node generates its own LSP in addition to all other LSPs generated by nodes on the broadcast network.
 - + LSP generated for Level-1 and Level-2 (if L1/L2 router)
 - + This can inflate the size of the IS-IS database in large networks with many broadcast segments
- + For routers connected with point-to-point Ethernet, change the network type to Point-to-Point to eliminate the Pseudo Node and its LSPs



Database Size Example With Pseudo Nodes



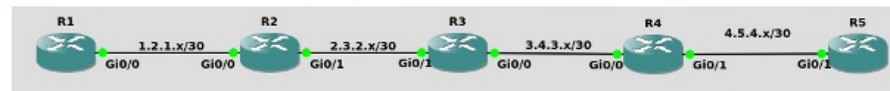
LSPs beginning as Rx.01 or .02
are generated from Pseudo
Nodes

```
R1#sho isis database
IS-IS Level-1 Link State Database:
LSPID      LSP Seq Num
R1.00-00   * 0x00000003
R2.00-00   0x00000005
R2.01-00   0x00000001
R3.00-00   0x00000004
R3.01-00   0x00000001
R3.02-00   0x00000001
R4.00-00   0x00000004
R5.00-00   0x00000003
R5.01-00   0x00000001
IS-IS Level-2 Link State Database:
LSPID      LSP Seq Num
R1.00-00   * 0x00000006
R2.00-00   0x00000007
R2.01-00   0x00000001
R3.00-00   0x00000006
R3.01-00   0x00000001
R3.02-00   0x00000001
R4.00-00   0x00000006
R5.00-00   0x00000004
R5.01-00   0x00000001
R1#
```

18-LSPs (total)



Database Size Example Without Pseudo Nodes



```
R3(config)#int range gig 0/0 - 1
R3(config-if-range)#isis network point-to-point
R3(config-if-range)#end
```

```
R1#sho isis database

IS-IS Level-1 Link State Database:
LSPID      LSP Seq Num
R1.00-00   * 0x000000DE
R2.00-00   0x000000DE
R3.00-00   0x000000DE
R4.00-00   0x000000DF
R5.00-00   0x000000D9
IS-IS Level-2 Link State Database:
LSPID      LSP Seq Num
R1.00-00   * 0x000000DD
R2.00-00   0x000000E2
R3.00-00   0x000000DE
R4.00-00   0x000000DE
R5.00-00   0x000000DA
R1#
```

10-LSPs (total)





Thanks for Watching!



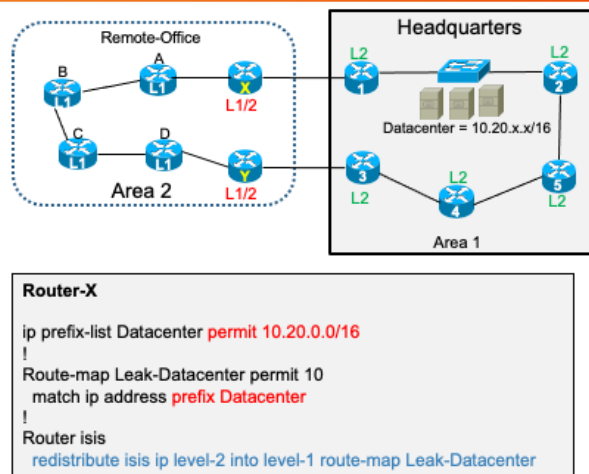
IS-IS Path Manipulation

Level-1 and Level-2 Router Interactions

- + Routers running as Level-1 only do not learn of Level-2 routes. They simply (dynamically) create a default route based on the closest L1/L2 routers.
 - + If multiple L1/L2 routers connect to an Area, this can lead to sub-optimal routing
 - + IS-IS LSP leaking (L2 routes leaked into the L1 domain) can alleviate this.



IS-IS LSP Leaking



- In this scenario, without IS-IS LSP Leaking configured on router-X, routers C and D would send all of their Inter-Area traffic to router-Y because it was closer than router-X. This would lead to non-optimal traffic forwarding.

IS-IS Default Metrics

- + The default IS-IS metric is "10" for any kind of interface rendering IS-IS more akin to a Distance-Vector protocol than a Link-State protocol.
- + For networks that contain a range of link bandwidths, predetermine what IS-IS cost values you want for link representations and document it.

Interface Bandwidth	IS-IS Metric Value
1 Gigabit	1000
10 Gigabit	100
40 Gigabit	10

For example, only

- + Default IS-IS metrics only allow a maximum metric of 1023 for total path metric.



Widening IS-IS Metrics

- + In extremely large topologies the maximum metric value may not be large enough
- + Consider the use of IS-IS "wide metrics" to gain access to a larger range of metric values

router isis

metric-style wide





Thanks for Watching!



IS-IS Security Design

IS-IS Authentication

- + Although IS-IS supports the use of a Key-Chain, it will not recognize any "cryptographic" commands within the keychain
 - + Keychain is ONLY used as an alternative placement of the password
 - + IS-IS supports both plain-text and MD5 authentication
 - + Although the IOS commands only say "MD5" this is actually HMAC-MD5



IS-IS Authentication

- + IS-IS supports area-wide (best for network automation) as well as interface-level authentication (best for security with using different passwords per neighbor)

```
!
key chain INE
key 1
  key-string ine123
!
interface GigabitEthernet0/0
ip address 2.1.2.1 255.255.255.0
ip router isis
  isis authentication mode md5
  isis authentication key-chain INE level-2
!
```



IS-IS Authentication Sniffer Capture

```

> IEEE 802.3 Ethernet
> Logical-Link Control
> ISO 10589 ISIS InTRA Domain Routeing Information Exchange Protocol
> ISIS HELLO
  ... ..10 = Circuit type: Level 2 only (0x2)
  0000 00.. = Reserved: 0x00
  SystemID (Sender of PDU): 0011.1111.1111
  Holding timer: 10
  PDU length: 1497
  .100 0000 = Priority: 64
  0... .... = Reserved: 0
  SystemID (Designated IS): 0011.1111.1111.01
  > Authentication (t=10, l=17)
    Type: 10
    Length: 17
    hmac-md5 (54), message digest (length 16) = e0d78cb3d570860482af1e1d5d3f3d05
  > Protocols Supported (t=129, l=1)
  > Area address(es) (t=1, l=2)
  > IP Interface address(es) (t=132, l=4)
  > Restart Signaling (t=211, l=3)
  > IS Neighbor(s) (t=6, l=6)
  > Padding (t=8, l=255)

```





Thanks for Watching!



Optimizing BGP For ISP Connections

Why Use BGP In The Enterprise?

- + Required for connectivity to ISP or virtualized cloud networks
 - + Pro: Dynamic routing tracks reachability to ISP
 - + Con: Extra complexity of BGP administration
- + Core network scalability
 - + Pro: BGP was designed for scalability
 - + Con: BGP has slow convergence
- + A requirement exists for different administrative domains
 - + Pro: BGP provides more variables for path selection control than any other protocol
 - + Con: BGP has extensive learning curve



Why Use BGP In The Enterprise?

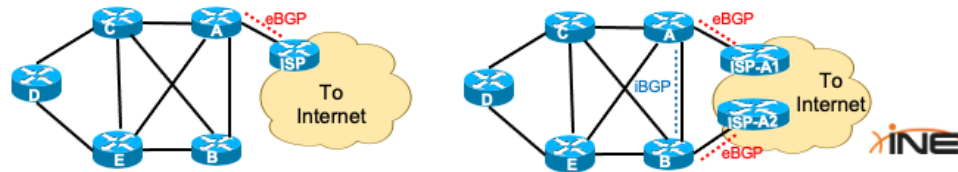
- + Required for connectivity to ISP or virtualized cloud networks
 - + Pro: Dynamic routing tracks reachability to ISP
 - + Con: Extra complexity of BGP administration
- + Core network scalability
 - + Pro: BGP was designed for scalability
 - + Con: BGP has slow convergence
- + A requirement exists for different administrative domains
 - + Pro: BGP provides more variables for path selection control than any other protocol
 - + Con: BGP has extensive learning curve

This is what we'll discuss in this section.



BGP For Single ISP Connectivity

- + For single-homed connectivity;
 - + BGP only required if using PI (Provider Independent) IP addressing
 - + SP transmits default route via BGP
 - + Enterprise advertises global networks to ISP via BGP
- + Dual-homed to a single ISP
 - + BGP can be used for ingress path control
 - + Receive default route from both ISP connections via BGP



BGP For Dual ISP Connectivity

- + Ensure that only locally-originated routes are advertised to each ISP

```
R4(config)#ip as-path access-list 2 permit ^$  
R4(config)#route-map Local-Only permit 10  
R4(config-route-map)#match as-path 2  
R4(config-route-map)#exit  
R4(config)#router bgp 300  
R4(config-router)#neighbor 7.7.7.7 route-map Local-Only out  
R4(config-router)#neighbor 5.5.5.5 route-map Local-Only out
```

- + Avoid redistributing your IGP routes into BGP. Take control using the "network" command instead.

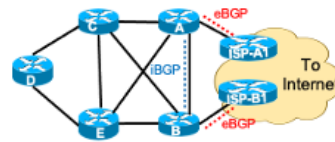


- Implement route filtering with as-path access-lists to ensure that only locally-originated routes are advertised to each ISP (prevents your ASN from becoming a Transit ASN).

Don't Forget To Summarize

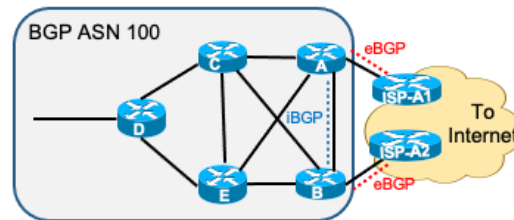
- + Implement route summarization to minimize the size of BGP tables and routing tables.

```
R4(config-router-af)#aggregate-address 165.50.32.0 255.255.224.0 summary-only
```



Manipulating ISP Inbound Path Selection

- + Understand your ISPs expectations and rules
 - + Do they allow AS-Path Prepending?
 - + Do they apply Weight or Local-Preference on their routers?





Thanks for Watching!



Enterprise BGP; Introduction to Core Network Design

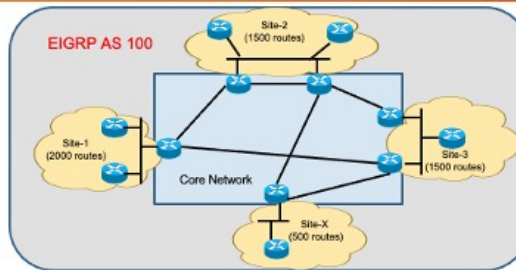
Why Use BGP In The Enterprise?

- + Required for connectivity to ISP or virtualized cloud networks
 - + Pro: Dynamic routing tracks reachability to ISP
 - + Con: Extra complexity of BGP administration
- + **Core network scalability**
 - + Pro: BGP was designed for scalability
 - + Con: BGP has slow convergence
- + **A requirement exists for different administrative domains**
 - + Pro: BGP provides more variables for path selection control than any other protocol
 - + Con: BGP has extensive learning curve

These are what we'll discuss in this section.



The Problem



- + One single IGP domain for everything
- + All networks propagated to all sites
- + 1000's of networks causing IGP instability

BGP Core Designs

- + Three design topologies exist for BGP core implementation
 - + Internal BGP (iBGP) architecture
 - + External BGP (eBGP) architecture
 - + Hybrid (iBGP/eBGP) architecture
- + Each is designed to satisfy certain requirements
- + We'll discuss each of these coming up.



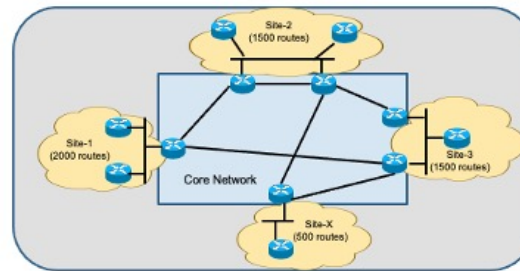


Thanks for Watching!



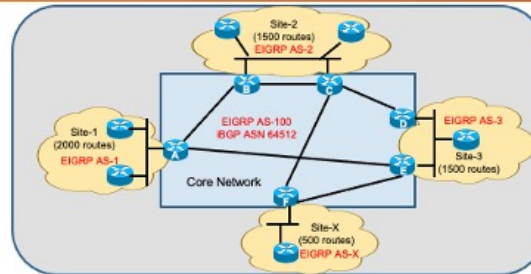
iBGP Core Network Design

Initial Assumptions



- + One administrative authority controls it all
- + Primary concern is to achieve route stability

iBGP Core Architecture Overview

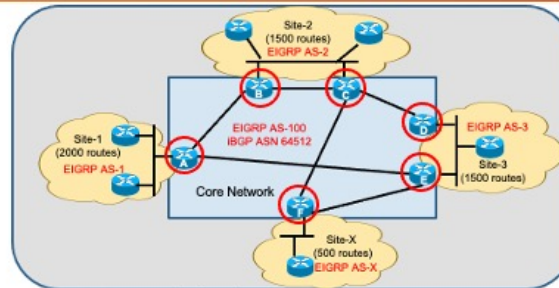


- + Uses a single BGP AS in the network core.
 - + All core routers run iBGP
 - + All core routers have a full iBGP mesh (unless route reflectors are implemented)
 - + Core routers peer via Loopbacks for stability



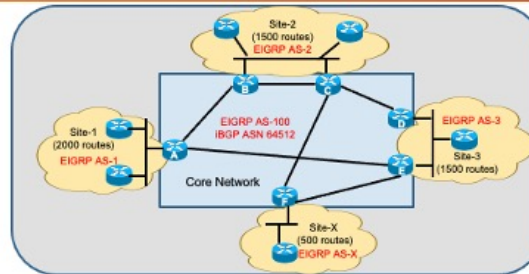
- The Cisco Press book recommends that you implement a full-mesh of the core iBGP routers, but I see no reason why route reflectors couldn't be implemented if the size of the core demands it.
- Because the IGP running in a single region doesn't have any IGP peerings with different regions, technically all regions COULD run the same EIGRP ASN. While this could potentially simplify automation, it might also introduce additional complexity into troubleshooting and potentially lead to unwanted IGP peerings between regions (i.e., "Sites") if you're not careful.

iBGP Core Prefix Propagation



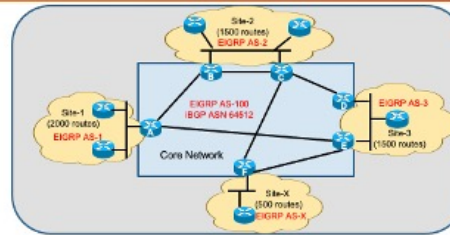
- + Core routers inject default route into regions
- + Core routers summarize (if possible) regional routes into core
- + Core routers must still contain full set of routes for entire domain
- + Utilize “network” statements, if possible, on Core routers
 - + Otherwise redistribute with careful filtering

iBGP Core Best Path Selection



- + Best path through the core determined by;
 - + IGP metric to iBGP next-hop
 - + Lowest BGP router-ID (if a tie in IGP metric)

iBGP Core Pros & Cons



+ Pros:

- + IGP metric tracks actual core topology better than BGP metrics which results in good best-path selection.

+ Cons:

- + Edge/Core routers must run three routing processes
- + Not suitable when a need exists for multiple routing policy authorities



- BGP was designed that a single BGP ASN should be under one source of administrative control. So this topology would NOT be suitable if each Region contains their own administrative control, claiming sole control over all aspects of routers within their region.

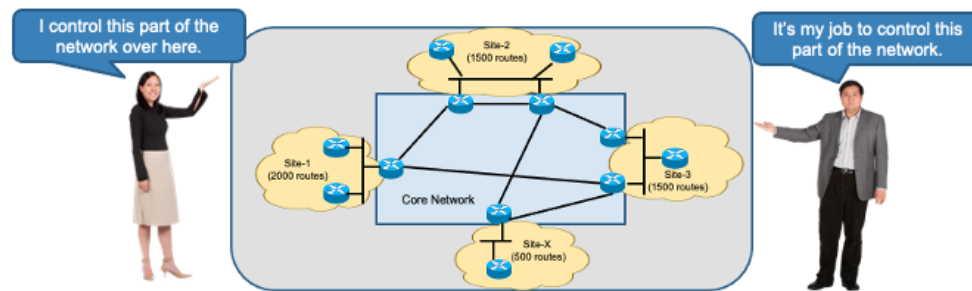


Thanks for Watching!



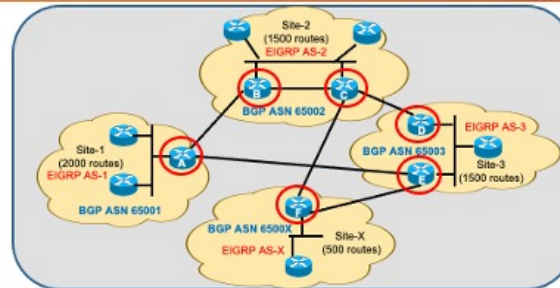
eBGP Core Network Design

Initial Assumptions



- + Network control needs to be dispersed
- + Must achieve route stability and scalability

eBGP Core Architecture Overview

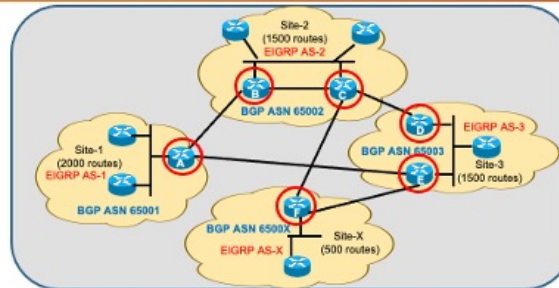


- + Each region a unique BGP ASN.
 - + All core routers run eBGP between different regions
 - + iBGP possible if a single region contains two-or-more Core routers
 - + No IGP necessary in Core: eBGP peerings between connected interfaces



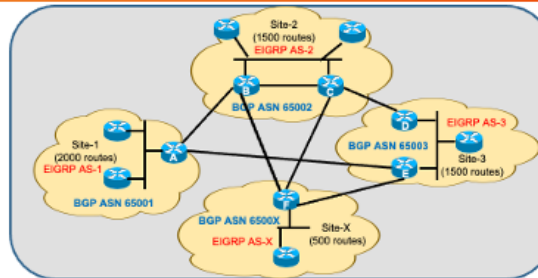
- Since each regional router is located within its own unique BGP ASN, it makes more sense here that the IGP be configured with its own unique AS (or process-ID) as well.

eBGP Core Prefix Propagation



- + Core routers inject default route into regions
- + Core routers summarize (if possible) regional routes into core
- + Core routers must still contain full set of routes for entire domain
- + Utilize “network” statements, if possible, on Core routers
 - + Otherwise redistribute with careful filtering

eBGP Core Best Path Selection

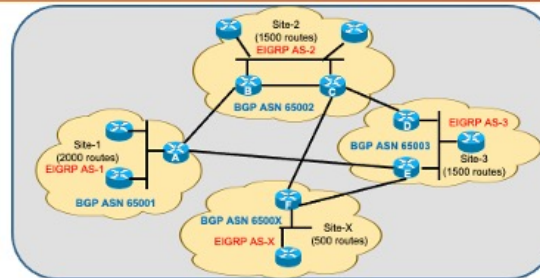


- + Best path through the core determined by;
 - + Shortest AS-Path Length
 - + Lowest BGP router-ID (if a tie in path length)
 - + It is recommended to enable "`bgp best path compare-routerid`" command
 - + Provides for a more deterministic path selection



- Without the command "`bgp best path compare-routerid`", if there is a tie in the AS-Path length then the default behavior (for eBGP) is to select the path that was learned first (the oldest path).
- If you know (when configuring BGP in the first place) that eBGP path selection may come down to router-ids, this allows you to deterministically set your router-ids in advance.

eBGP Core Pros & Cons



+ Pros:

- + Core routers only need to run two routing processes
- + No need for an IGP in the Core
- + Suitable when a need exists for multiple routing policy authorities

+ Cons:

- + Best path selection run without visibility into the physical topology and link bandwidth.



- BGP was designed that a single BGP ASN should be under one source of administrative control. So this topology would be ideal if each Region contains their own administrative control, claiming sole control over all aspects of routers within their region.
- Each region would be responsible for the BGP policy of their own ASN.

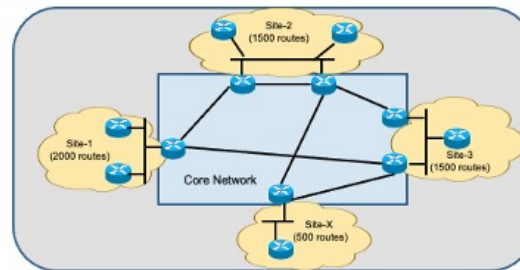


Thanks for Watching!



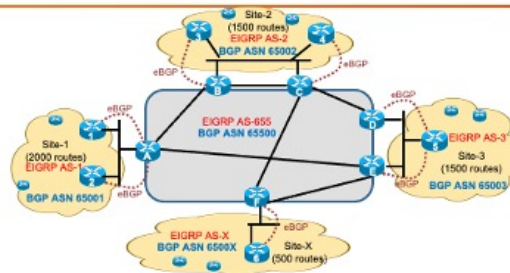
BGP Hybrid Core Network Design

Initial Assumptions



- + Network control needs to be divided
- + Core network path selection should be based on actual topology and path bandwidth
- + Must achieve route stability and scalability

iBGP/eBGP Core Architecture Overview

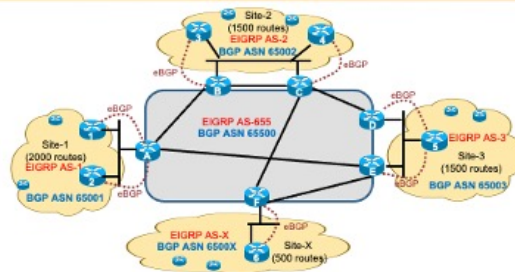


- + Each region a unique BGP ASN.
 - + If multiple regional border routers exist in a single region, they run iBGP between themselves
- + Core is a unique BGP ASN
 - + All Core routers run iBGP with other Core routers
 - + eBGP between Regional Border routers and Core routers
 - + IGP necessary in Core: iBGP peerings between Loopbacks



- Remember to have a full-mesh of iBGP within the Core, unless you plan on implementing Route Reflectors.
- If multiple Regional Routers exist within a region (like Router-3 and 4) they would run iBGP between themselves.

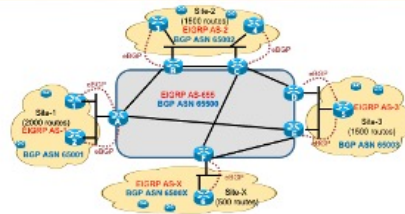
iBGP/eBGP Core Prefix Propagation



- + Core routers inject default route into regions via eBGP
- + Regional Border routers summarize (if possible) regional routes into core
- + Core routers must still contain full set of routes for entire domain
- + Utilize "network" statements, if possible, on Regional Border routers
 - + Otherwise redistribute with careful filtering



iBGP/eBGP Core Best Path Selection

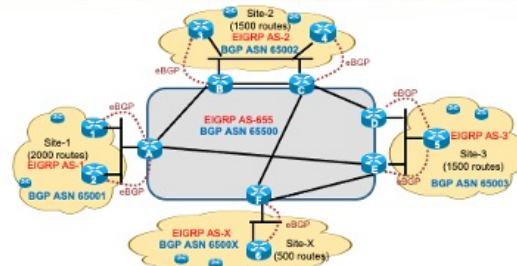


- + Best path through the core determined by;
 - + Lowest IGP metric to BGP next-hop
 - + Lowest BGP router-ID (if a tie in path length)
 - + It is recommended to enable "**bgp best path compare-routerid**" command
 - + Provides for a more deterministic path selection
- + eBGP path attributes can be utilized to influence inbound path selection from the Core into a Region (when multiple Regional Border routers exist)



- Because this topology is well-suited to providing administrative control to groups of engineers who are each in charge of their own Region, it allows engineers controlling Site-1 to determine the inbound BGP policy between their routers and Router-A.

iBGP/eBGP Core Pros & Cons



+ Pros:

- + Core routers only need to run two routing processes
- + Suitable when a need exists for multiple routing policy authorities
- + Path topology and bandwidth determines best path through the Core

+ Cons:

- + More upfront configuration complexity





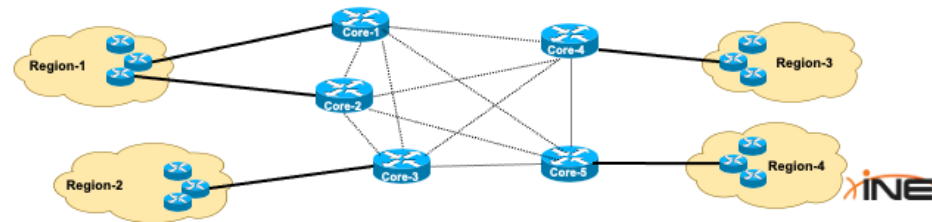
Thanks for Watching!



BGP Core Scalability

Core iBGP Full Mesh

- + Within an Enterprise, ideally the iBGP core should support a full-mesh
 - + Provides for maximum redundancy
 - + Allows routers to learn of all alternative paths and select the best path for themselves

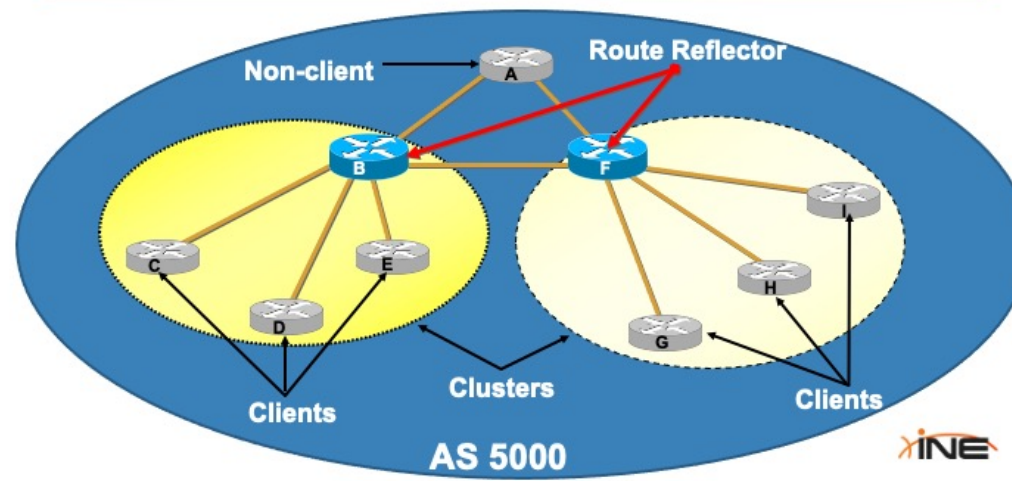


Scaling Beyond A Full Mesh

- + Sometimes the size of the BGP core (or router placement) might make a full iBGP mesh unfeasible
- + To maintain end-to-end iBGP reachability without a full mesh one can implement:
 - + Route Reflectors
 - + Confederations
- + Must still ensure iBGP topology contains redundancy



Route Reflector Terminology Review



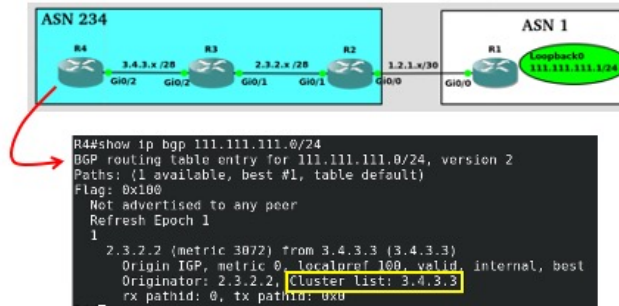
Route Reflector Attributes

- + Originator_ID attribute
 - + Carries the RID of the originator of the route in the local AS (created by the RR)
- + Cluster_list attribute
 - + The local cluster-id is added when the update is sent to clients (added by the RR).
 - + Cluster-ID used by RR to detect loops
 - + Default is to use the Router-id
 - + Configurable using the command, "*bgp cluster-id x.x.x.x*"



Route Reflector Cluster-IDs

- + The BGP Cluster-ID attribute is used by Route Reflectors to stop the looping of BGP routing updates



- + This value (by default) takes the form of the BGP Router-ID but can be configured to a different value.

Route Reflector Forwarding Rules

- + Once the best path is selected:
 - + From non-client iBGP peer → reflect to all clients
 - + From client iBGP peer → *reflect to all non-clients* AND other clients
 - + From eBGP peer → reflect to all clients and non-clients

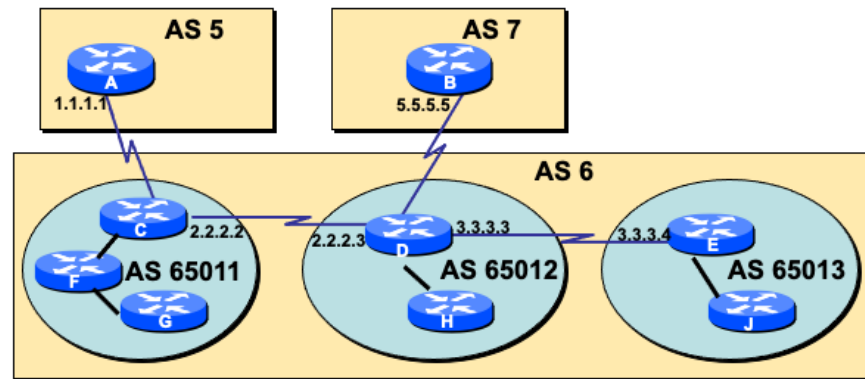


Review Of BGP Confederations

- + Solves iBGP mesh problem
- + Divide the AS into sub-AS's
 - + It is recommended to use private AS#s for Sub AS's
- + Visible to outside world as single AS
- + Preserve local preference, MED, and NEXT_HOP
- + iBGP speakers within a sub-AS are fully meshed
- + Route-reflectors can be used within a Sub AS



Confederations Visualized



Peering When Using Confederations

- + eBGP used between sub-AS's
- + iBGP used within sub-AS's
- + Next_hop from remote AS is preserved as update is passed between Sub-AS's
- + **DESIGN NOTE:** Confederations can be more difficult to design in a way that facilitates an incremental design.
 - + Better to avoid them and just use Route Reflectors if possible.





Thanks for Watching!



BGP Route Reflector Best Practices

Objectives For iBGP Scalability

- + Within an Enterprise, ideally the iBGP core should support a full-mesh
- + Sometimes the size of the BGP core (or router placement) might require Route Reflectors
- + Properly implement route reflectors and/or confederations
- + Ensure iBGP topology contains redundancy



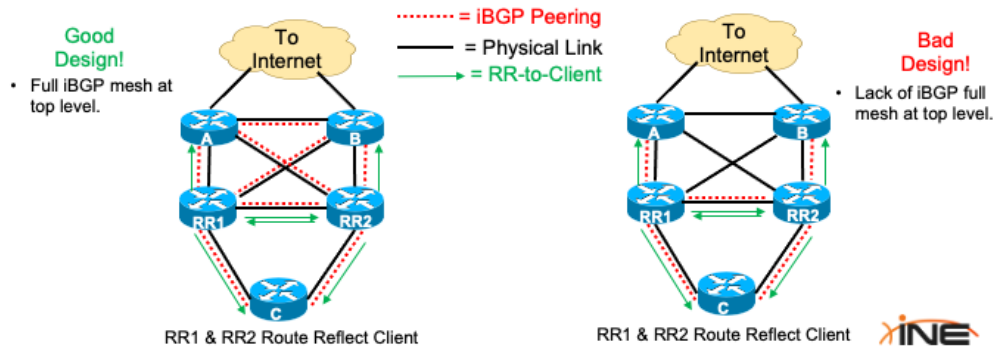
Route Reflector Usage

- + Remember that RRs only reflect their best path to any given prefix.
- + Remember the additional BGP Path Attributes that RRs utilize:
 - + Originator-ID
 - + Cluster List
- + Multiple hierarchies of RRs can be implemented



Where A Full Mesh Is Needed

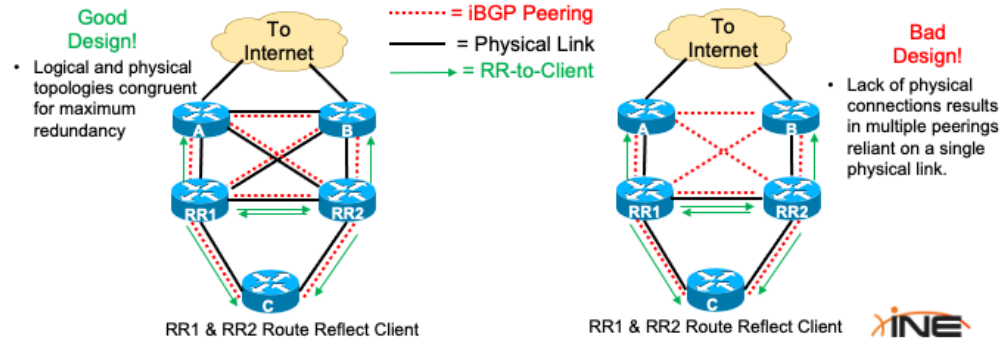
- + Edge routers should have full iBGP mesh with each other as well as with Route Reflectors.



- Whether or not routers “A” and “B” are RR-Clients of the two Route Reflectors depends on just how much redundancy you want in your network.
- When it comes to BGP, higher redundancy comes with a cost (greater BGP resource consumption in the form of more duplicated BGP prefix entries and more process time spent on best-path selection).

Importance Of Congruent Links

- + In an RR environment, ideally one will keep physical and logical topologies congruent.



Effectively Utilizing Cluster-IDs

- + For redundancy, RR clients should be served by two RRs
- + Then the question becomes...
 - + Should all clients, and both reflectors, be within the same Cluster?



- + Or different clusters?



- + Two RRs serving the same set of clients should be configured with **identical Cluster-IDs IF:**
 - + BGP resource usage is a concern on the RRs
 - + Neither RR is ever expected to need BGP paths that were advertised from the other RR.



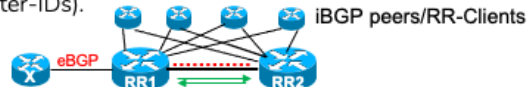
- Remember that if a RR receives a BGP update from one of its Clients...that RR will reflect that update to other Clients as WELL AS non-Clients. So in this case if RR1 receives an iBGP update from its client, if an iBGP peering exists between RR1 and RR2, (whether these two routers are clients of each other or not) RR2 will have the update reflected from RR1.
- If you WANT RR2 to store these reflected updates, it received from RR1 (as potential backup paths for RR2) then both RR1 and RR2 need to have different Cluster-IDs.

Dual RR Peering



+ Should RR1 and RR2 be iBGP peers with each other?

+ If either RR1 or RR2 have an eBGP peer, then **yes** (can have the same, or different, Cluster-IDs).



+ If either RR1 or RR2 have non-redundant paths to iBGP peers, then **yes** (but require **different** Cluster-IDs).



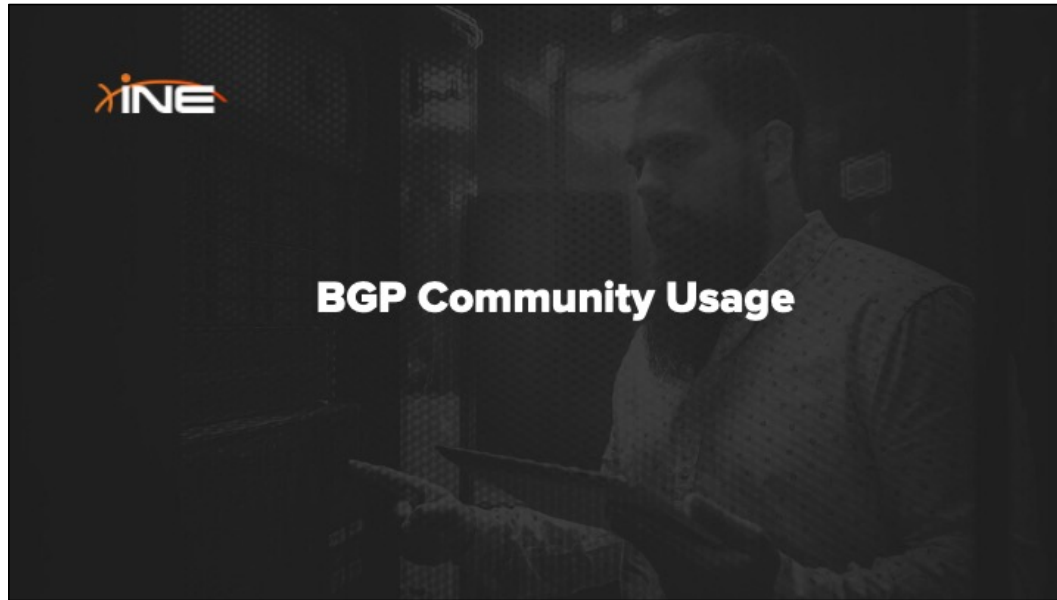
- If neither of the above conditions are true, then there is no need to have an iBGP peering between RR1 and RR2.



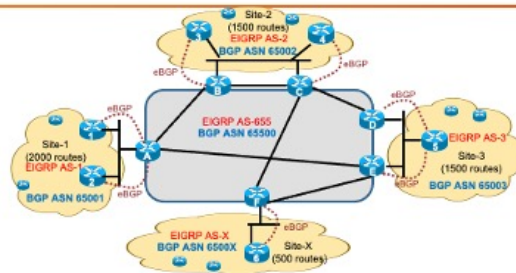
Thanks for Watching!



BGP Community Usage



Controlling Prefix Propagation



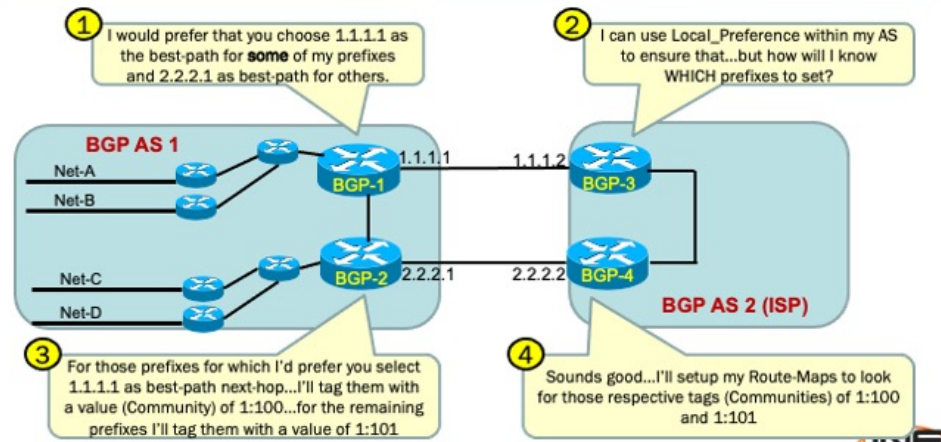
- + One can utilize well-known BGP Communities to control BGP peer route propagation.
 - + Useful if you want to advertise prefixes to peers but control how far they go from there.
 - + Peers automatically recognize and follow well-known community rules

BGP Communities Overview

- + Descriptive number/value that is applied as a “tag” to a route.
 - + Standard Community (RFC 1997) = 32-bits/4-bytes long
 - + Extended Community (RFC 4360) = 64-bits/8-bytes long
 - + Typical format AS:value
- + Used to group destinations and apply a common policy
- + Each prefix can belong to multiple communities
- + Some values are “Well-Known” and are understood to have special pre-defined meanings.



BGP Communities Use-Case



Well-Known BGP Communities

- + **internet** = (value = 0x0) all routes are members of this community
- + **no-export** = (value = 0xFFFFFFFF01) do not advertise to eBGP peers
- + **no-advertise** = (value = 0xFFFFFFFF02) do not advertise to any peer
- + **local-AS** = (value = 0xFFFFFFFF03) do not advertise outside local AS (used with confederations)



Effective Use Of Communities

- + To control which routes your peer is allowed to transmit to other (downstream) peers consider the use of the BGP "no-export" and "no-advertise" communities
 - + **No-export** ensures that your eBGP peer won't propagate (i.e., export) the prefix via eBGP to another AS.
 - + **No-advertise** ensures that your peer can only keep the prefix for itself. Not allowed to advertise it to other iBGP or eBGP peers.



- The "no-export" community only works if your eBGP peer also has another eBGP peer. It will NOT prevent your eBGP peer from propagating the route to his iBGP peers
- The "no-export" community is non-transitive so when your eBGP peer receives the community it will REMOVE IT prior to forwarding the route to his iBGP peers.

BGP Community Configuration

```
router bgp 333
neighbor x.x.x.x send-community both
neighbor x.x.x.x route-map Community out
!
access-list 1 permit 100.100.0.0 0.0.255.255
!
route-map Community permit 10
match ip address 1
set community no-export
!
route-map Community permit 20
```

Must explicitly give BGP permission to send Communities to BGP neighbors.

In this example, a BGP Community is added to outbound routes as stated in a Route map

This well-known community prevents the receiver from advertising matching routes to any of its eBGP peers.





Thanks for Watching!



Controlling BGP Update Generation



BGP Update Control

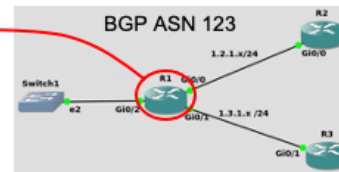
- + BGP updates from your peers may contain prefixes you don't want...or shouldn't receive
- + Routers typically create one BGP update per neighbor.
- + The quantity and content of BGP updates should be controlled;
 - + Prevent the consumption of WAN bandwidth by reducing unnecessary BGP updates
 - + Reduce the quantity of BGP updates when peered with many iBGP peers



Utilizing Peer Groups

- + BGP Peer-Groups should be utilized within an iBGP core that contains many fully-meshed peers
- + Allows the router to only generate a single BGP update for the peer-group...instead of one update per peer.

```
router bgp 123
  bgp log-neighbor-changes
  neighbor iBGP peer-group
  neighbor iBGP remote-as 123
  neighbor iBGP capability orf prefix-list send
  neighbor iBGP prefix-list MyNetworks in
  neighbor 1.2.1.2 peer-group iBGP
  neighbor 1.3.1.3 peer-group iBGP
!
ip forward-protocol nd
!
no ip http server
no ip http secure-server
!
ip prefix-list MyNetworks seq 5 permit 10.0.0.0/8 ge 9
```



- From Cisco documentation:
- “The major benefit you achieve when you specify a BGP peer group is that a BGP peer group reduces the amount of system resources (CPU and memory) necessary in an update generation. In addition, a BGP peer group also simplifies the BGP configuration. A BGP peer group reduces the load on system resources by allowing the routing table to be checked only once, and updates to be replicated to all peer group members instead of being done individually for each peer in the peer group.”

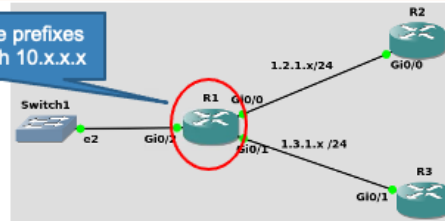
Eliminating Unwanted Prefixes

- + Receiving unwanted routes can consume unnecessary WAN bandwidth
- + Route filtering should be configured to prevent sending unwanted (or poisoned) prefixes
- + Prefix-Lists can be lengthy and prone to configuration error.
- + Considering using ORF (Outbound Route Filtering) to dynamically push Prefix-Lists to BGP peers.



BGP ORF Configuration Example

Only send me prefixes
beginning with 10.x.x.x

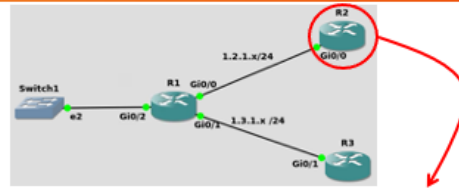


```
router bgp 123
  bgp log-neighbor-changes
  neighbor 18GP peer-group
  neighbor 18GP remote-as 123
  !
  !
  !
  ip forward-protocol nd
  !
  !
  no ip http server
  no ip http secure-server
  !
  !
  !
```

3
2

1

BGP ORF Verification



```
R2#show ip bgp neighbor 1.2.1.1
BGP neighbor is 1.2.1.1, remote AS 123, internal link
BGP version 4, remote router ID 111.111.111.1
BGP state = Established, up for 00:04:05
```

```
For address family: IPv4 Unicast
Session: 1.2.1.1
BGP table version 14, neighbor version 14/0
Output queue size : 0
Index 3, Advertise bit 0
3 update-group member
```

```
AF-dependant capabilities:
  Outbound Route Filter (ORF) type (128) Prefix-list:
    Send-mode: received
    Receive-mode: advertised
  Outbound Route Filter (ORF): received (1 entries)
```



Advertisement Of ORF Capability

19	8.680459	1.2.1.1	1.2.1.2	TCP	68 44882 → 179 [SYN] Seq=
20	8.682127	1.2.1.2	1.2.1.1	TCP	68 179 → 44882 [SYN, ACK]
21	8.682883	1.2.1.1	1.2.1.2	TCP	68 44882 → 179 [ACK] Seq=1
22	8.684733	1.2.1.1	1.2.1.2	BGP	122 OPEN Message
23	8.686895	1.2.1.2	1.2.1.1	TCP	68 179 → 44882 [ACK] Seq=1
24	8.686852	1.2.1.2	1.2.1.1	BGP	122 OPEN Message
25	8.687394	1.2.1.2	1.2.1.1	BGP	73 KEEPALIVE Message
26	8.687999	1.2.1.1	1.2.1.2	TCP	68 44882 → 179 [ACK] Seq=6
27	8.689236	1.2.1.1	1.2.1.2	BGP	73 KEEPALIVE Message
28	8.610206	1.2.1.1	1.2.1.2	BGP	73 KEEPALIVE Message
29	8.610685	1.2.1.1	1.2.1.2	BGP	77 UPDATE Message
30	8.610737	1.2.1.2	1.2.1.1	TCP	68 179 → 44882 [ACK] Seq=8
31	8.645432	1.2.1.1	1.2.1.2	BGP	98 ROUTE-REFRESH Message
<p> * Frame 31: 90 bytes on wire (720 bits), 90 bytes captured (720 bits) on interface -, id 0 * Ethernet II, Src: 0c:c1:fb:e6:00:00 (0c:c1:fb:e6:00:00), Dst: 0c:9a:b2:66:00:00 (0c:9a:b2:66:00:00) * Internet Protocol Version 4, Src: 1.2.1.1, Dst: 1.2.1.2 * Transmission Control Protocol, Src Port: 44882, Dst Port: 179, Seq: 138, Ack: 88, Len: 36 * Border Gateway Protocol - ROUTE-REFRESH Message Marker: ffffffffffffffffffffffffffffffff Length: 36 Type: ROUTE-REFRESH Message (5) Address family identifier (AFI): IPv4 (1) Subtype: Normal route refresh request [RFC2918] with/without ORF [RFC5291] (0) Subsequent address family identifier (SAFI): Unicast (1) - ORF information ORF flag: Immediate (1) ORF type: Cisco PrefixList ORF-Type (128) ORF length: 9 - ORFEntry PrefixList 00. = ORFEntry Action: Add (0) ..0. = ORFEntry Match: Permit (0) ORFEntry Sequence: 5 ORFEntry PrefixMask length lower bound: 9 ORFEntry PrefixMask length upper bound: 0 10.0.0.0/8 </p>					





Thanks for Watching!



Route Redistribution Design

The Main Goals Of Redistribution

- + Redistribution is the process of taking routes learned from one method and sending them in routing updates via a different method/protocol.
- + When configuring redistribution, one should consider:
 - + Security ("am I sending too much information?")
 - + **Loop avoidance** ("how can I prevent routing loops?")
 - + Scalability ("how do I prevent from overwhelming the receiving protocol?")



Redistribution With Security In Mind

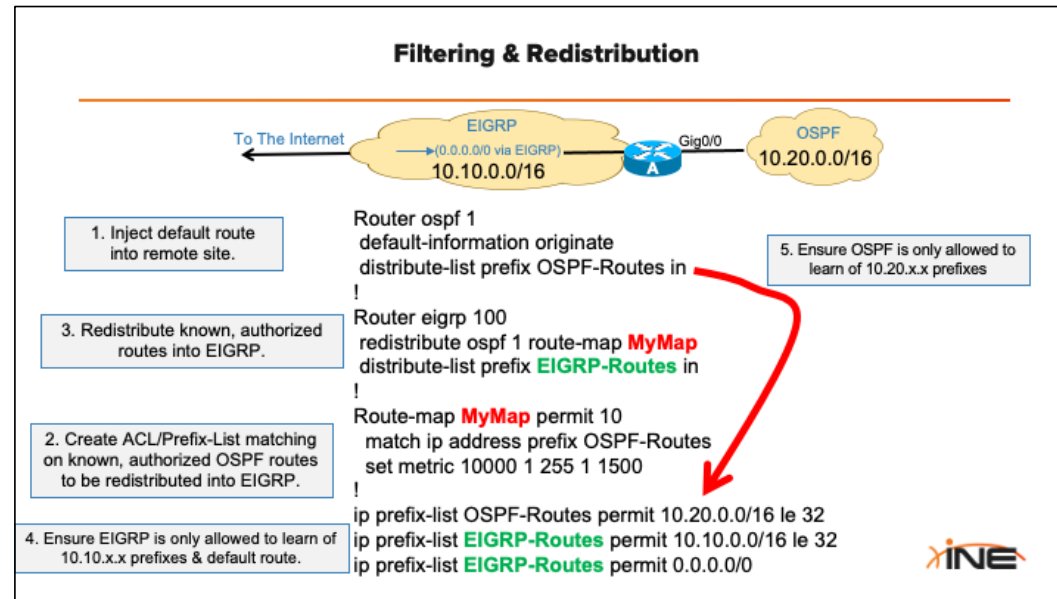
- + Avoid redistribution if possible and inject default routes instead.
- + Redistribute only what is needed, not what is possible
 - + Ask yourself, "What information does each routing domain NEED to know?"
- + Always pair your redistribute statements with Route-Maps
 - + Provides granular control over redistribution
 - + Provides for future scalability
 - + Easy to edit



Avoiding Redistribution-Induced Loops

- + Whenever possible, redistribute at only a single point
- + Always implement route filtering along with redistribution (i.e “Distribute-Lists”)
 - + Provides a double protection against private networks from being redistributed
 - + Ensures that only known networks are redistributed (in the event that your route-map is too permissive)
 - + Inbound filtering ensures that your own internal networks won't be learned from external sources



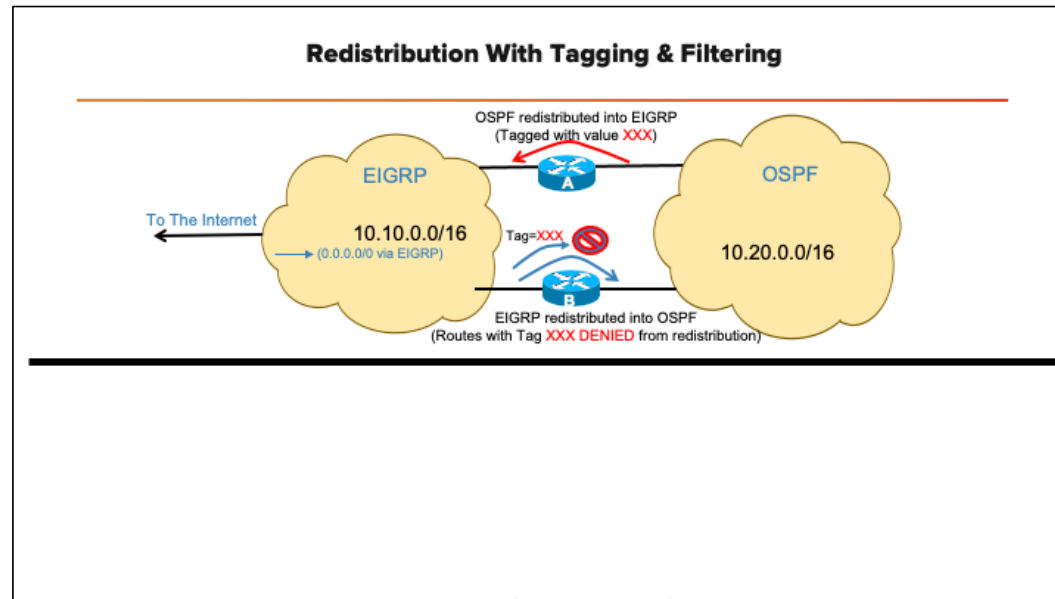


- Remember that with OSPF you can't control what OSPF LSAs are received and placed into your Link State Database. But an inbound Distribute-List CAN control if those LSAs are translated into routes for your Routing Table.

Bidirectional Redistribution Design


- + If bidirectional redistribution is required, take measures to prevent routes from being redistributed back into their originating domain.
 - + Route tagging
 - + Filters based on tags





- This is especially important to do when working with EIGRP, as EIGRP External routes have the worst (lowest preferred) Administrative Distance value of all routing protocols (outside of iBGP).

Redistribution Scalability

- + Whenever possible, redistribute from a less powerful routing protocol to a more powerful routing protocol
 - + The most accepted method of ranking protocols from most to least powerful is as follows: BGP, OSPF, IS-IS, EIGRP, IGRP, and RIP.

Redistribute in this direction.
- + When redistributing routes into EIGRP;
 - + Select a metric that is reasonable
(BW = 1,000,000, Dly = 1, Rly=255, Load= 1, MTU = 1500)
 - + Don't use "metric 1 1 1 1 1" as this will yield a very HIGH initial metric



- notice that we're setting the bandwidth here to "1" and the lower the bandwidth...the higher the computed distance value.
- An EIGRP route's distance will increase as it goes further and further from the redistribution point
- The route might quickly reach the point of "infinity" and no longer be considered usable
- You don't have to put much other thought into the metric UNLESS you have two-or-more redistribution points for the same routes.

