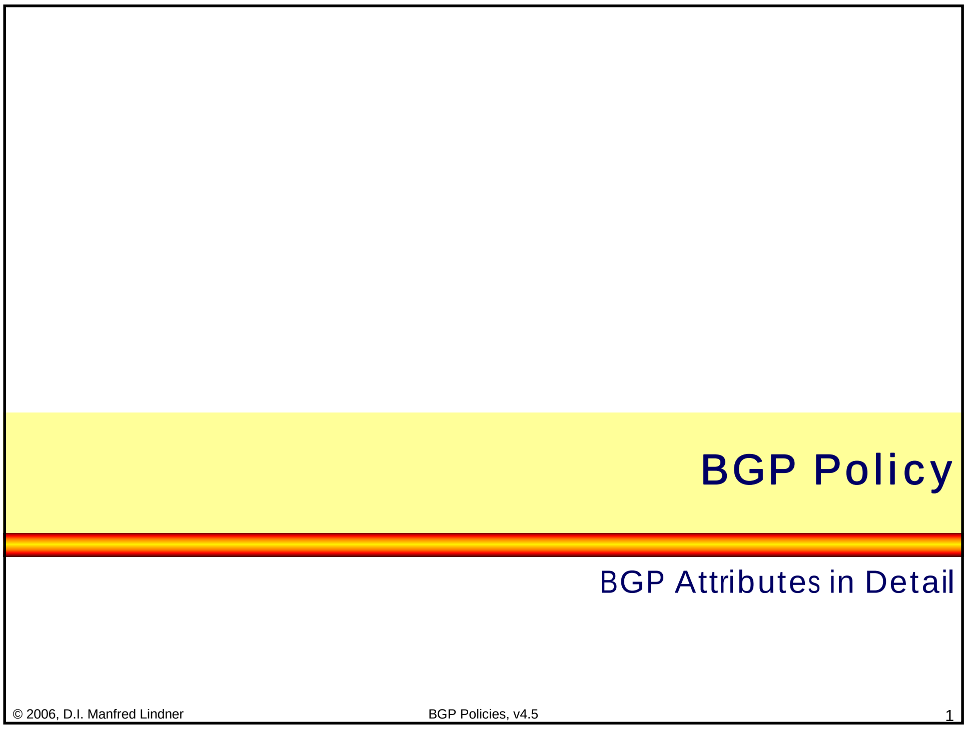
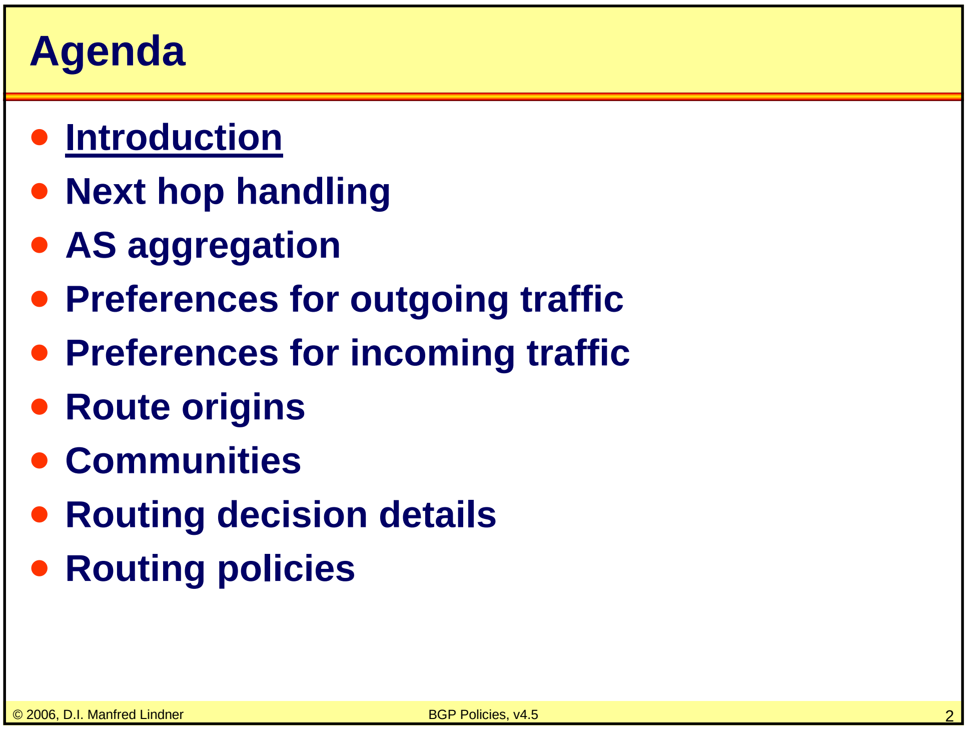


BGP Policy

BGP Attributes in Detail

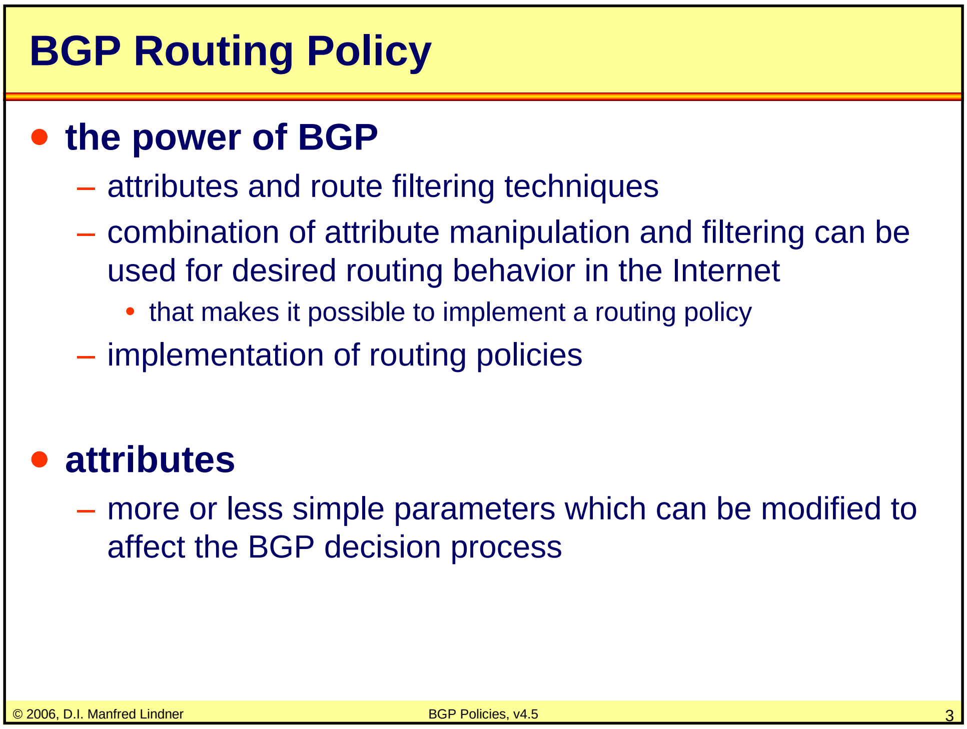
© 2006, D.I. Manfred Lindner BGP Policies, v4.5 1



Agenda

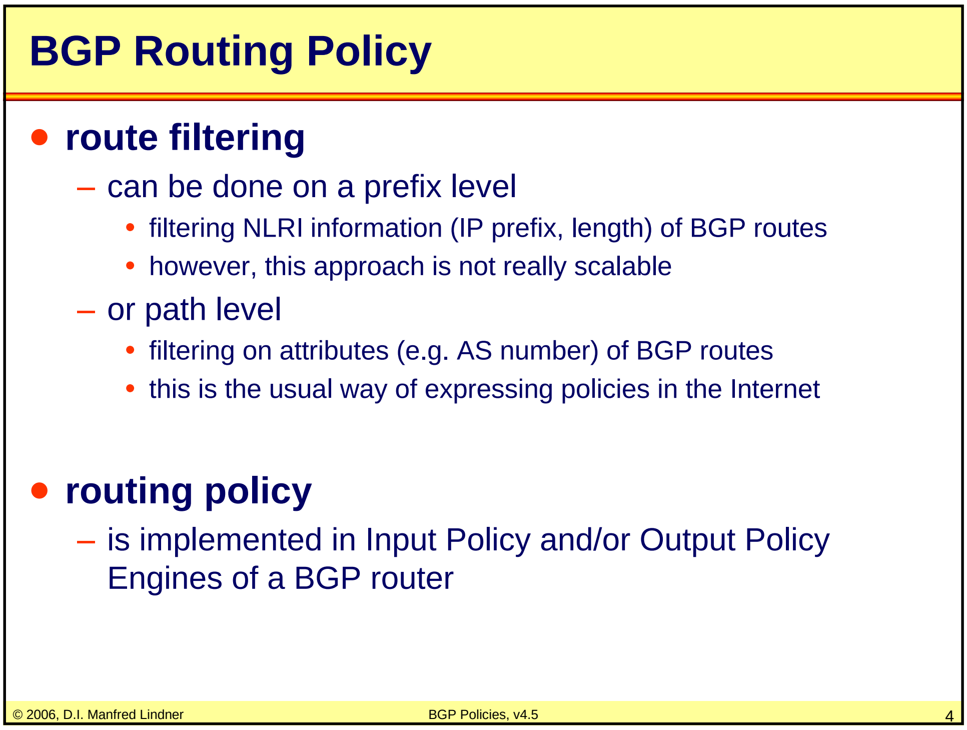
- Introduction
- Next hop handling
- AS aggregation
- Preferences for outgoing traffic
- Preferences for incoming traffic
- Route origins
- Communities
- Routing decision details
- Routing policies

© 2006, D.I. Manfred Lindner BGP Policies, v4.5 2



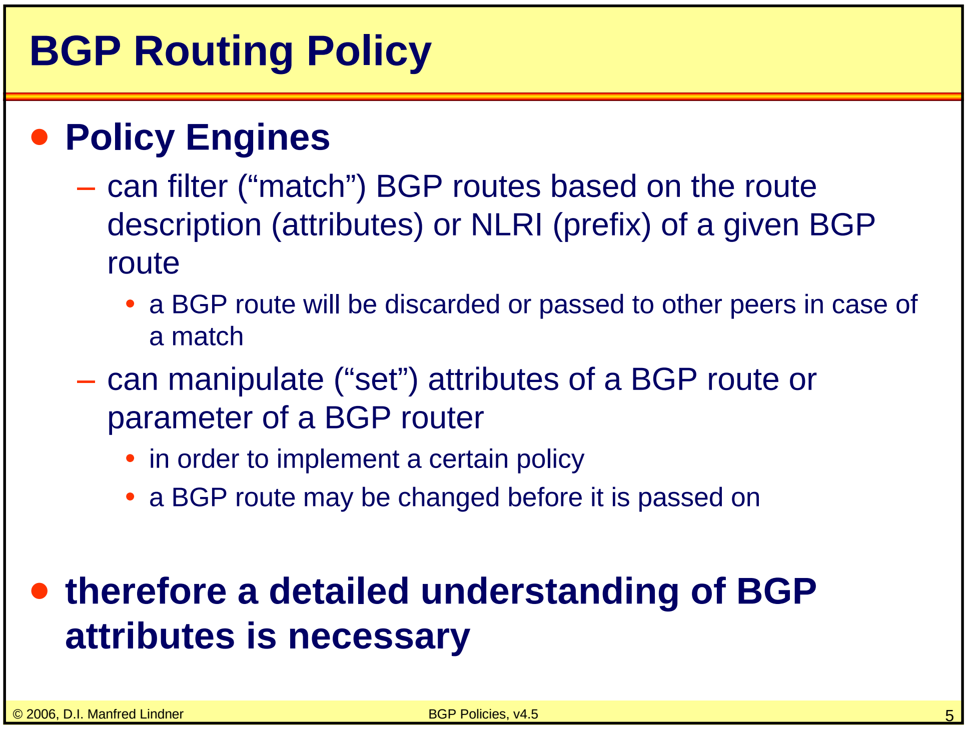
BGP Routing Policy

- **the power of BGP**
  - attributes and route filtering techniques
  - combination of attribute manipulation and filtering can be used for desired routing behavior in the Internet
    - that makes it possible to implement a routing policy
  - implementation of routing policies
- **attributes**
  - more or less simple parameters which can be modified to affect the BGP decision process



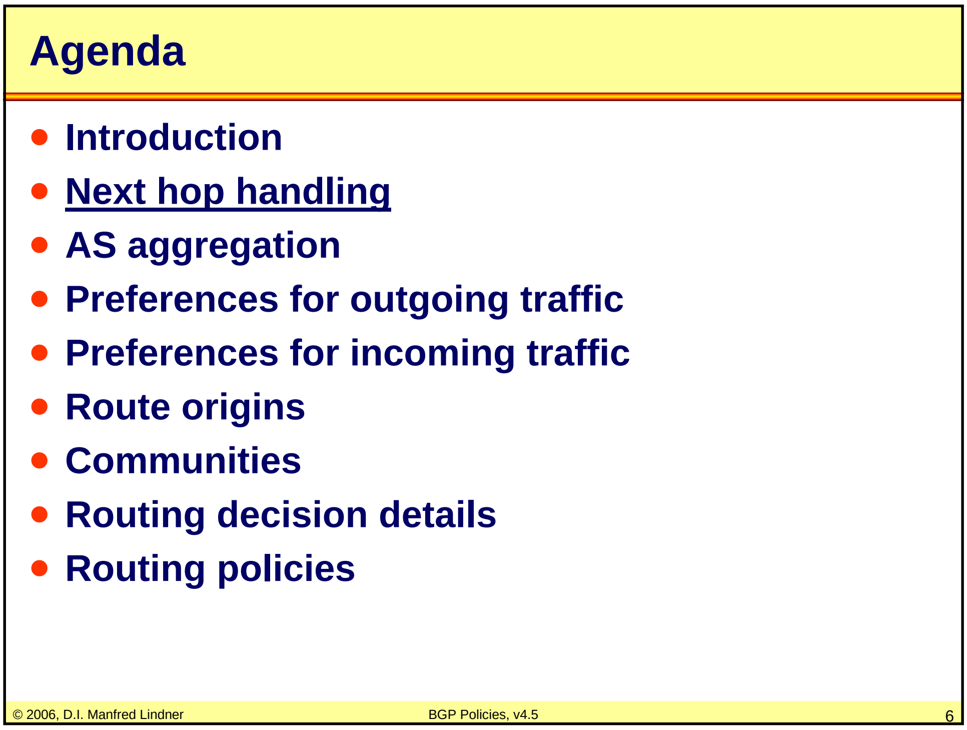
BGP Routing Policy

- **route filtering**
  - can be done on a prefix level
    - filtering NLRI information (IP prefix, length) of BGP routes
    - however, this approach is not really scalable
  - or path level
    - filtering on attributes (e.g. AS number) of BGP routes
    - this is the usual way of expressing policies in the Internet
- **routing policy**
  - is implemented in Input Policy and/or Output Policy Engines of a BGP router



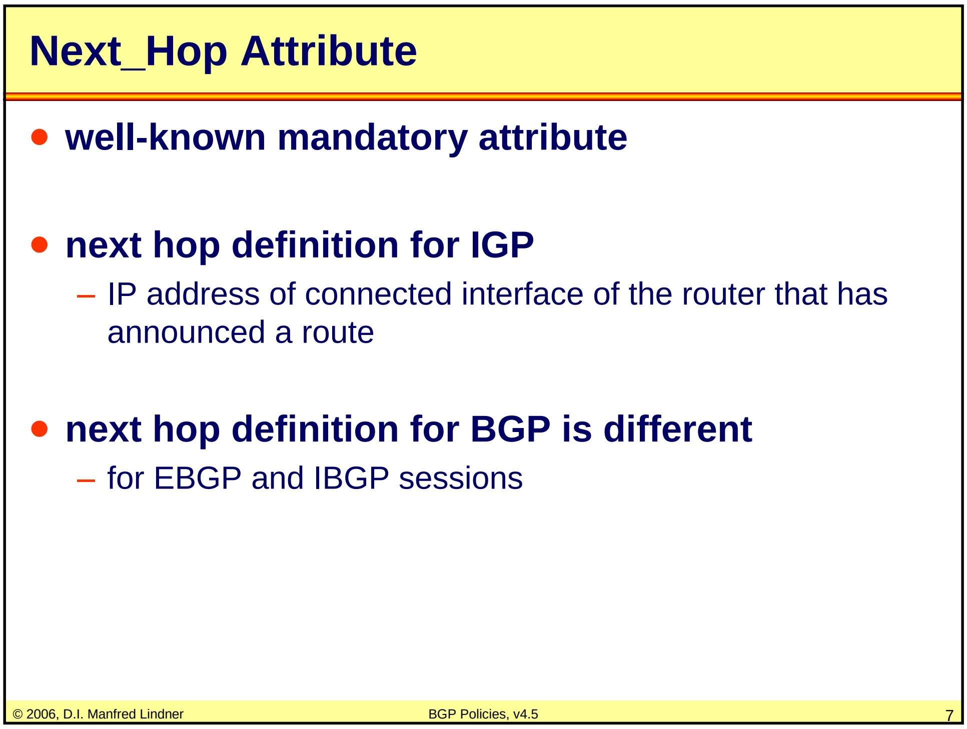
BGP Routing Policy

- **Policy Engines**
  - can filter (“match”) BGP routes based on the route description (attributes) or NLRI (prefix) of a given BGP route
    - a BGP route will be discarded or passed to other peers in case of a match
  - can manipulate (“set”) attributes of a BGP route or parameter of a BGP router
    - in order to implement a certain policy
    - a BGP route may be changed before it is passed on
- **therefore a detailed understanding of BGP attributes is necessary**



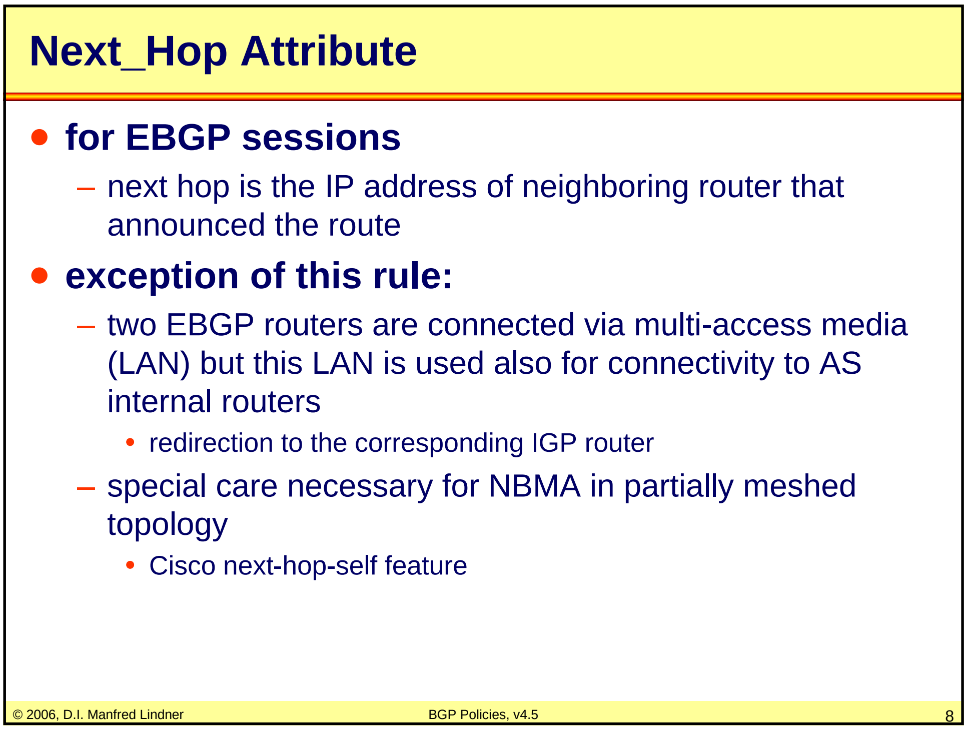
Agenda

- **Introduction**
- **Next hop handling**
- **AS aggregation**
- **Preferences for outgoing traffic**
- **Preferences for incoming traffic**
- **Route origins**
- **Communities**
- **Routing decision details**
- **Routing policies**



Next_Hop Attribute

- **well-known mandatory attribute**
- **next hop definition for IGP**
  - IP address of connected interface of the router that has announced a route
- **next hop definition for BGP is different**
  - for EBGP and IBGP sessions



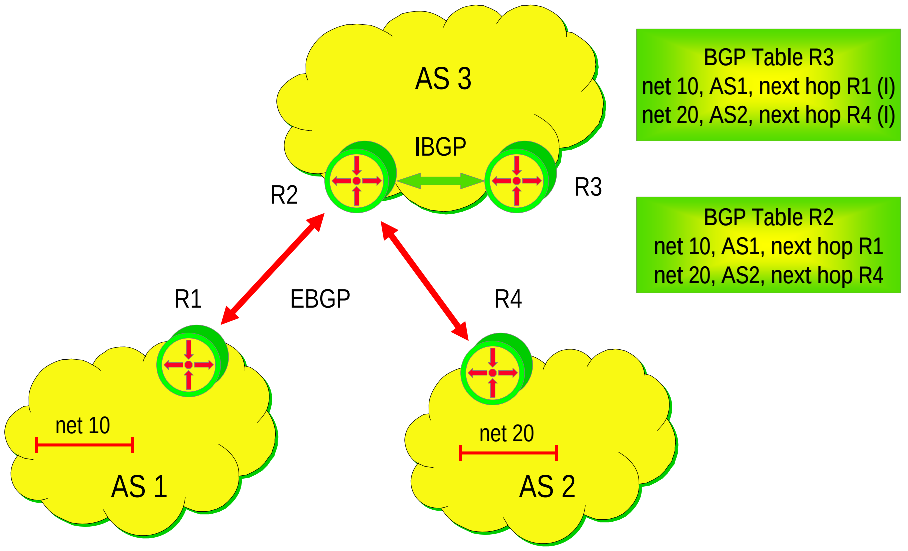
Next_Hop Attribute

- **for EBGP sessions**
  - next hop is the IP address of neighboring router that announced the route
- **exception of this rule:**
  - two EBGP routers are connected via multi-access media (LAN) but this LAN is used also for connectivity to AS internal routers
    - redirection to the corresponding IGP router
  - special care necessary for NBMA in partially meshed topology
    - Cisco next-hop-self feature

Next_Hop Attribute

- **for IBGP sessions**
  - 1.) for routes originated inside the AS next hop is the IP address of the neighbor that announced the route
  - 2.) for routes injected into the AS via EBGP next hop learned from EBGP is carried unaltered into IBGP
- **because of this IBGP behavior**
  - recursive IP lookup is necessary if next hop is not directly reachable
  - reachability of next hop must be advertised via some IGP or static routing
    - next hop must be reachable via normal IP routing table

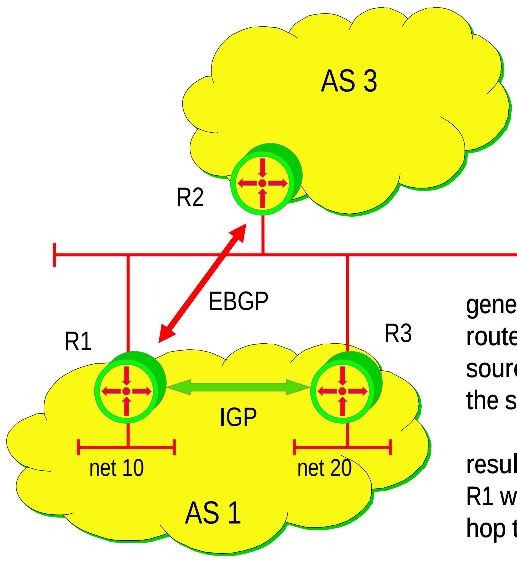
Next Hop Example 1



BGP Table R3
net 10, AS1, next hop R1 (I)
net 20, AS2, next hop R4 (I)

BGP Table R2
net 10, AS1, next hop R1
net 20, AS2, next hop R4

Next Hop Example 2



BGP Table R2
net 10, AS1, next hop R1
net 20, AS1, next hop R3

general IP rule on multi-access media: router should always advertise the actual source of a route in case the source is on the same media

result:
R1 will advertises via EBGP that the next hop to reach net 20 is R3

Agenda

- Introduction
- Next hop handling
- <u>AS aggregation</u>
- Preferences for outgoing traffic
- Preferences for incoming traffic
- Route origins
- Communities
- Routing decision details
- Routing policies

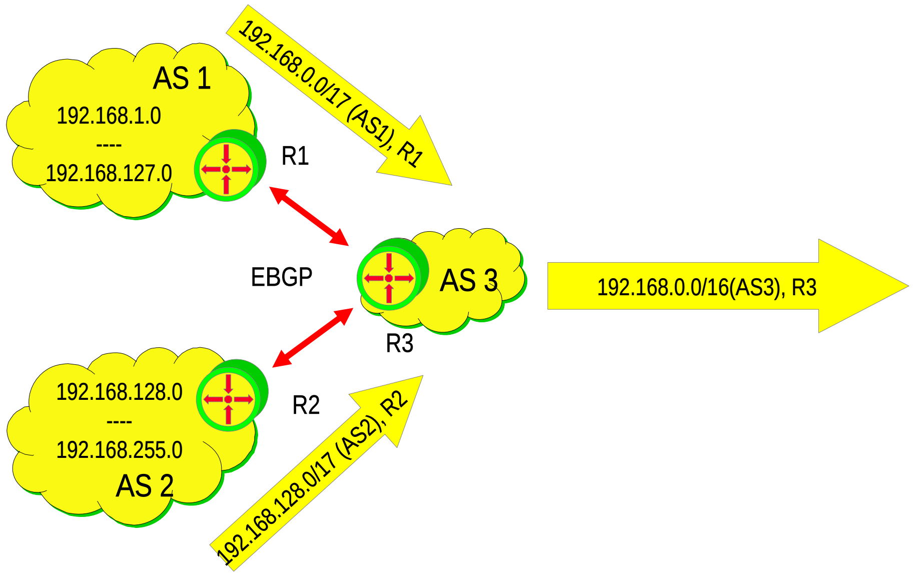
AS_Path Attribute

- **describes sequence of AS numbers (list) a route traversed to reach a destination**
  - – well-known mandatory attribute
  - – originator of a route adds its own AS number when sending the route to its external BGP peers
  - – each receiver adds its AS number to the beginning of the list before it passes the route to other external BGP peers
  - – passing a route to an internal BGP peer leaves AS_Path intact
- **used to ensure loop-free topology**
- **used to determine best route to a destination**
  - – shorter path is always preferred

AS_Path Aggregation

- **aggregation (summarization) of IP addresses**
  - – can lead to loss of path information and hence to routing loops or sub-optimal routing
  - – information about origination of a route will be lost
- **therefore the following attributes are introduced**
  - – Atomic_Aggregate attribute
  - – Aggregator attribute
  - – AS-Set
- **but be very careful doing aggregation for another party**
  - – try do avoid it
    - in most cases it is a design problem but not a principle problem

Aggregation Example 1



AS 1

192.168.1.0

192.168.127.0

R1

192.168.0.0/17 (AS1), R1

EBGP

AS 3

R3

192.168.0.0/16(AS3), R3

192.168.128.0

192.168.255.0

AS 2

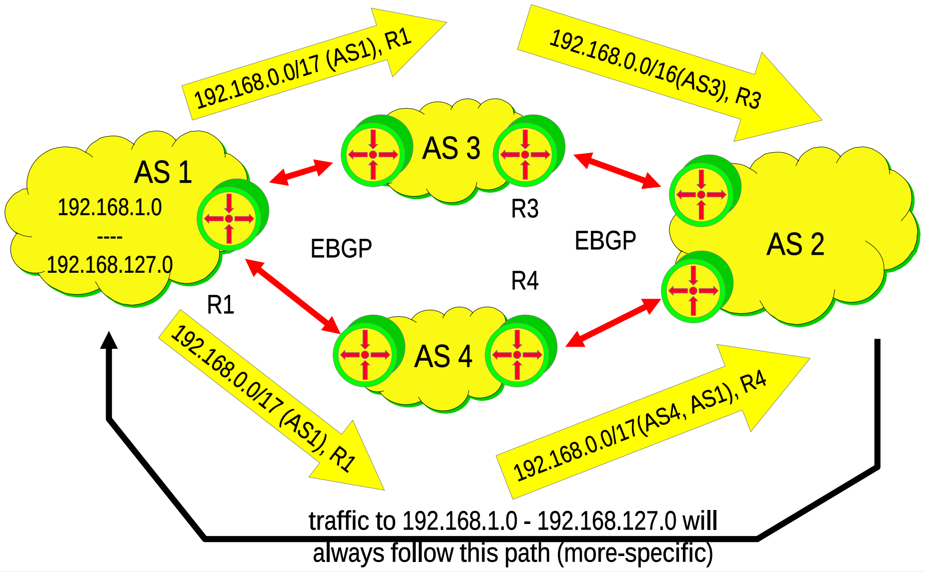
R2

192.168.128.0/17 (AS2), R2

How specific is a route?

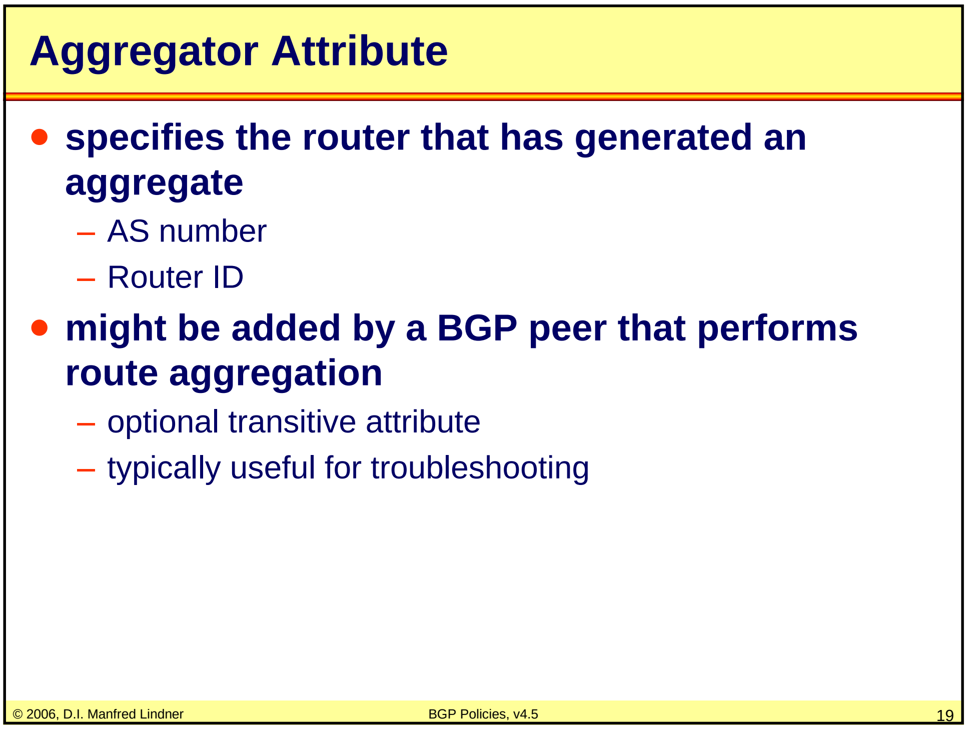
- **more specific = smaller set of destinations**
  - longer prefix
- **less specific = larger set of destinations**
  - shorter prefix
- **general IP routing rule:**
  - when overlapping routes are present in the routing table the more specific route shall take precedence
  - routing rule of longest match prefix
  - also used for BGP

Aggregation Example 2



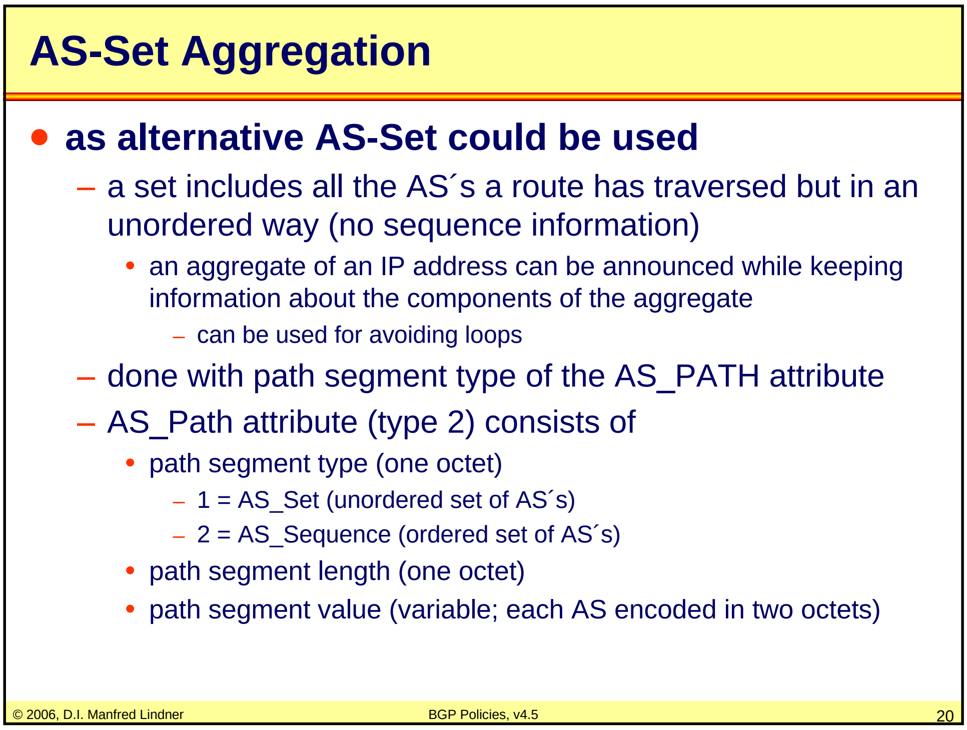
Atomic_Aggregate Attribute

- **if route aggregation done by an BGP router**
  - would cause a loss of information
    - e.g. a certain AS number will not longer be seen in the path
  - then this BGP router must attach the Atomic_Aggregate attribute to this route description
    - well-known discretionary attribute
  - that specifies that some AS´s may be missing from the AS_Path attribute
    - but does not specify which router was the aggregator
      - however can be done optionally by Aggregator attribute
    - also does not specify what AS numbers are missing
- **exception of this rule**
  - aggregate is described by AS-Set parameter



Aggregator Attribute

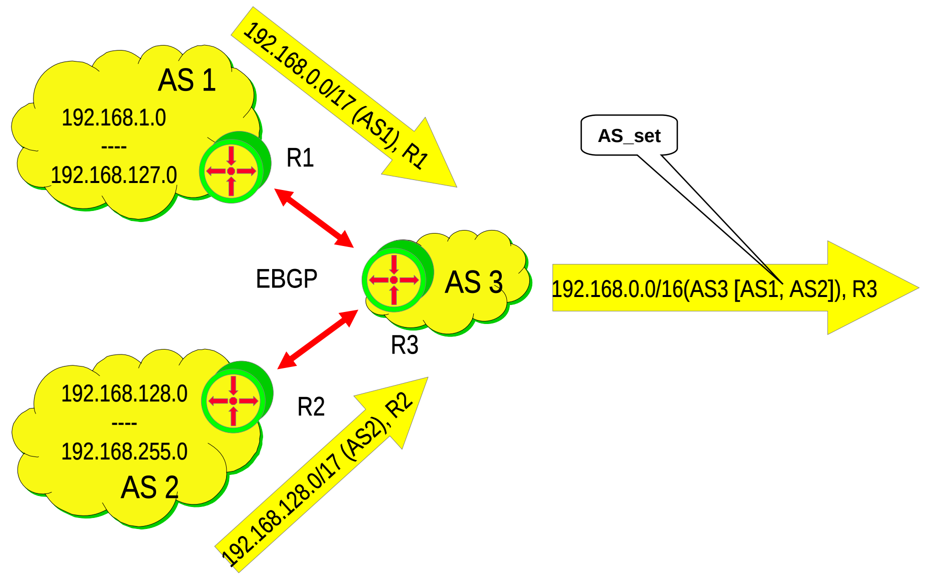
- **specifies the router that has generated an aggregate**
  - AS number
  - Router ID
- **might be added by a BGP peer that performs route aggregation**
  - optional transitive attribute
  - typically useful for troubleshooting



AS-Set Aggregation

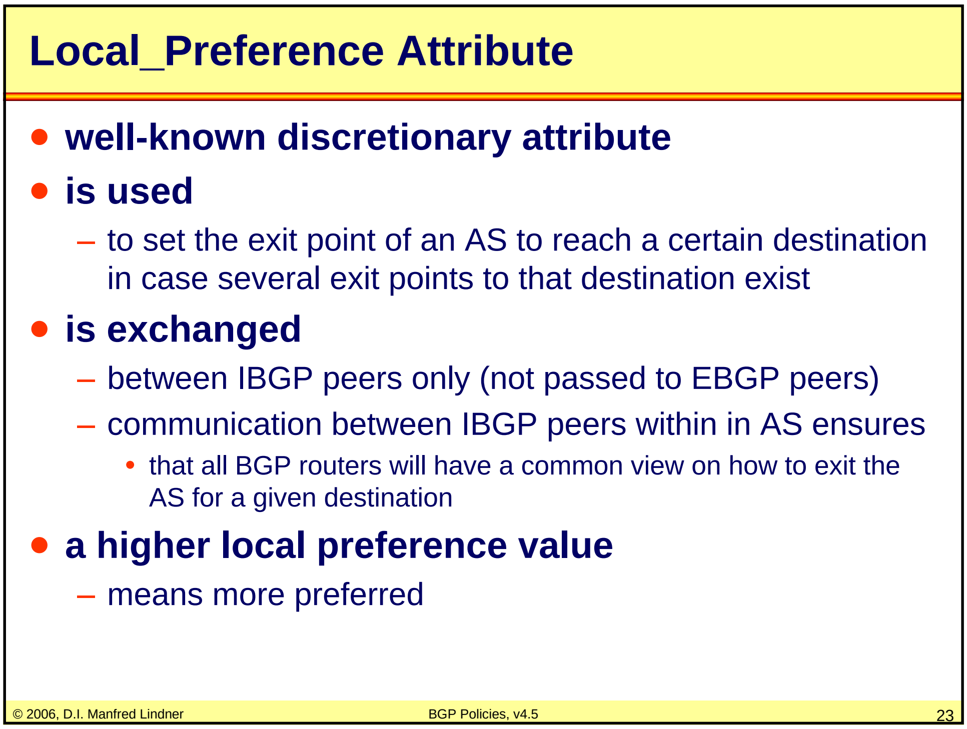
- **as alternative AS-Set could be used**
  - a set includes all the AS´s a route has traversed but in an unordered way (no sequence information)
    - an aggregate of an IP address can be announced while keeping information about the components of the aggregate
      - can be used for avoiding loops
  - done with path segment type of the AS_PATH attribute
  - AS_Path attribute (type 2) consists of
    - path segment type (one octet)
      - 1 = AS_Set (unordered set of AS´s)
      - 2 = AS_Sequence (ordered set of AS´s)
    - path segment length (one octet)
    - path segment value (variable; each AS encoded in two octets)

AS-Set Example



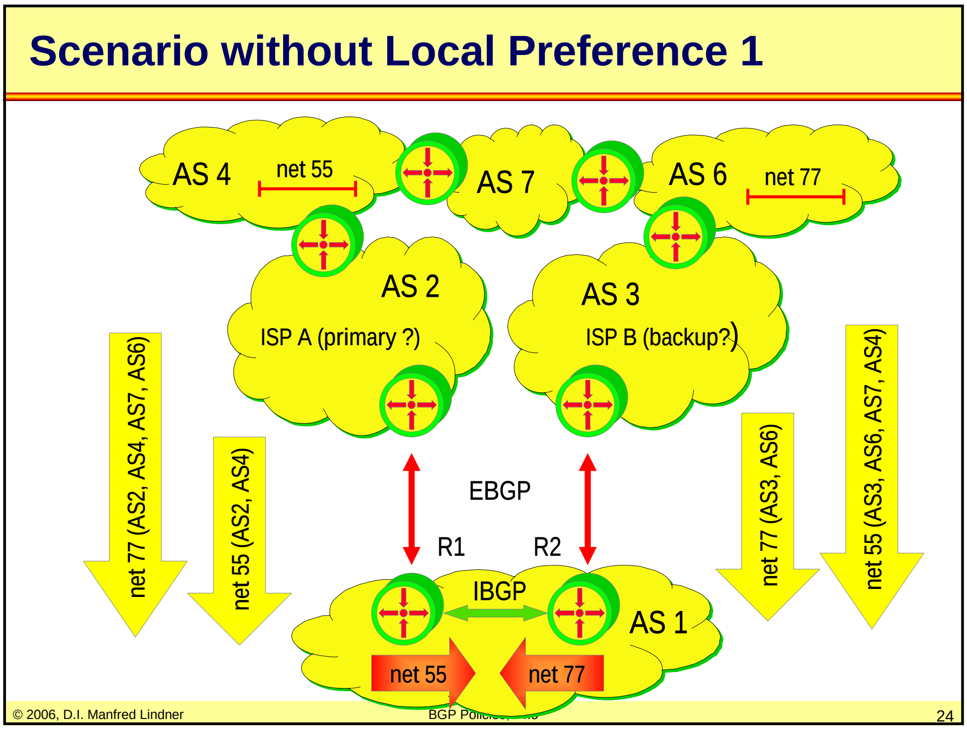
Agenda

- Introduction
- Next hop handling
- AS aggregation
- <u>Preferences for outgoing traffic</u>
- Preferences for incoming traffic
- Route origins
- Communities
- Routing decision details
- Routing policies



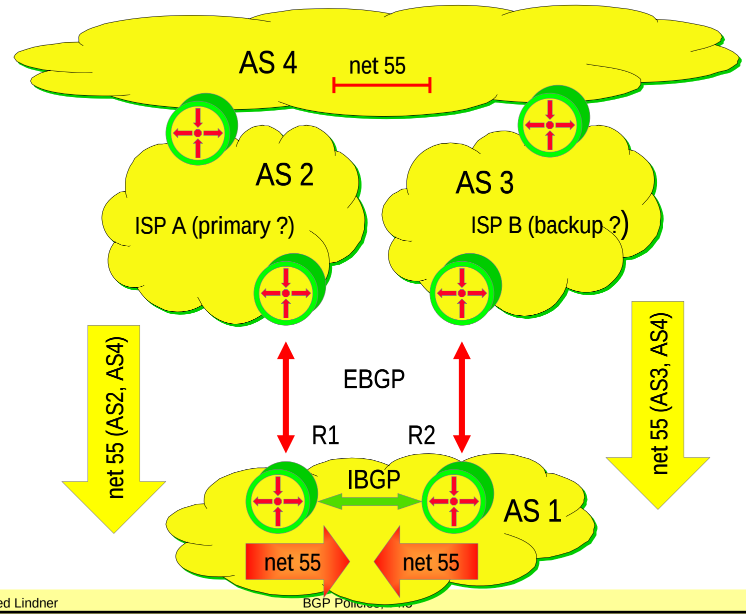
Local_Preference Attribute

- **well-known discretionary attribute**
- **is used**
  - to set the exit point of an AS to reach a certain destination in case several exit points to that destination exist
- **is exchanged**
  - between IBGP peers only (not passed to EBGP peers)
  - communication between IBGP peers within in AS ensures
    - that all BGP routers will have a common view on how to exit the AS for a given destination
- **a higher local preference value**
  - means more preferred

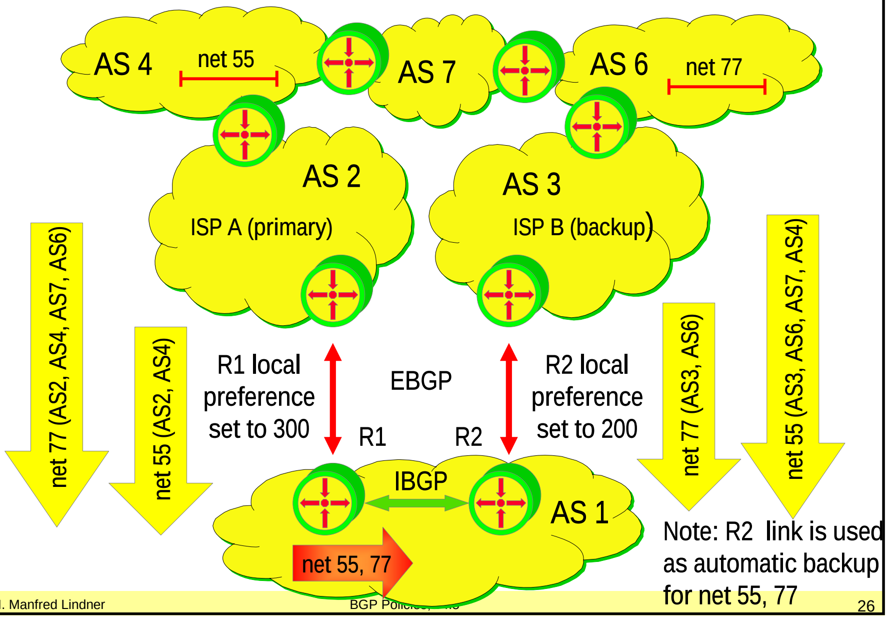


Scenario without Local Preference 1

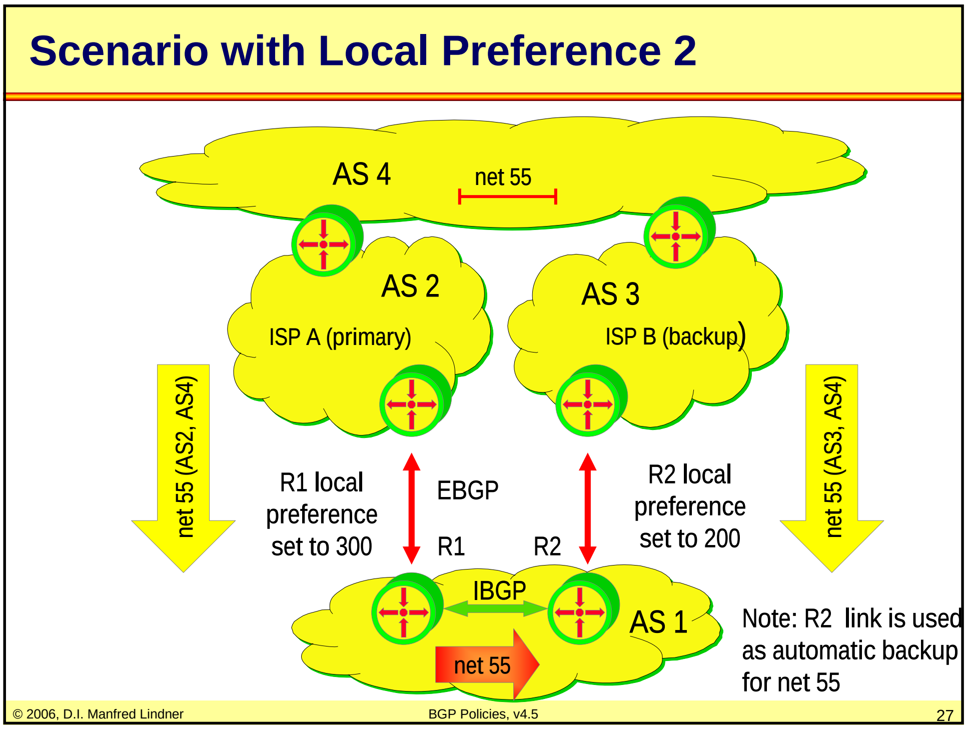
Scenario without Local Preference 2



Scenario with Local Preference 1



Note: R2 link is used as automatic backup for net 55, 77



Scenario with Local Preference 2

AS 4

net 55

AS 2

ISP A (primary)

AS 3

ISP B (backup)

net 55 (AS2, AS4)

net 55 (AS3, AS4)

R1 local preference set to 300

EBGP

R1

R2

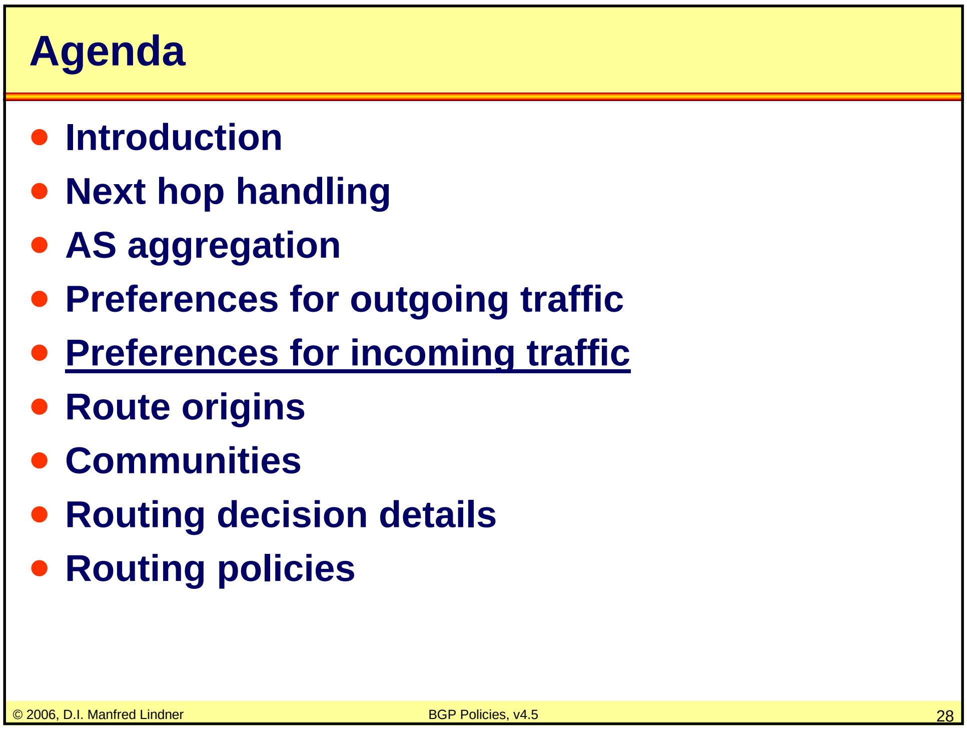
R2 local preference set to 200

IBGP

AS 1

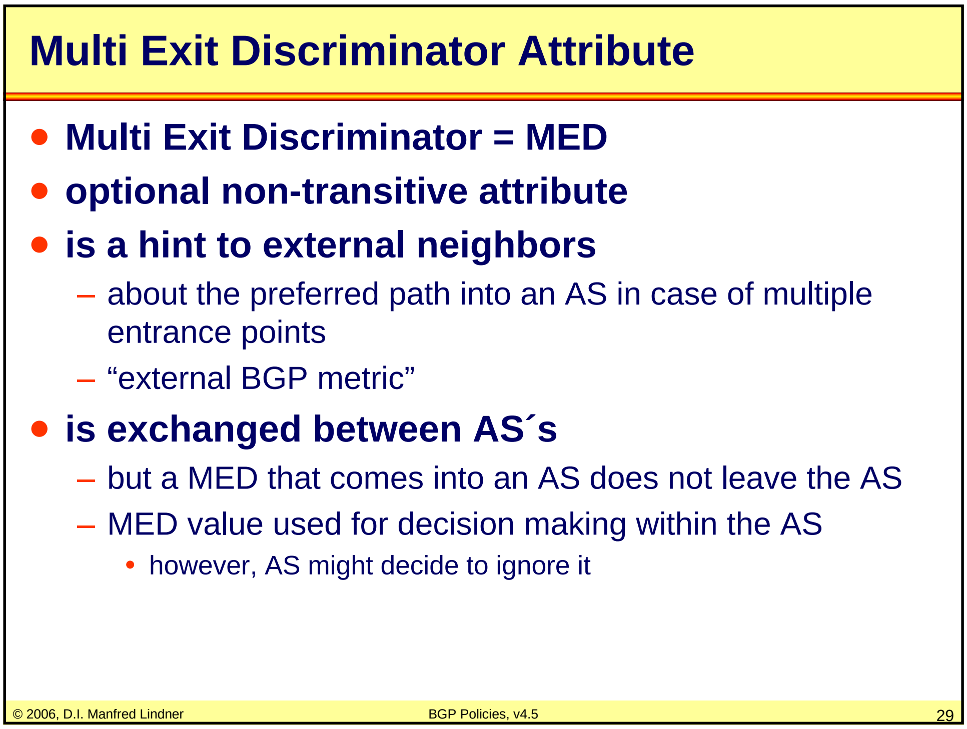
net 55

Note: R2 link is used as automatic backup for net 55



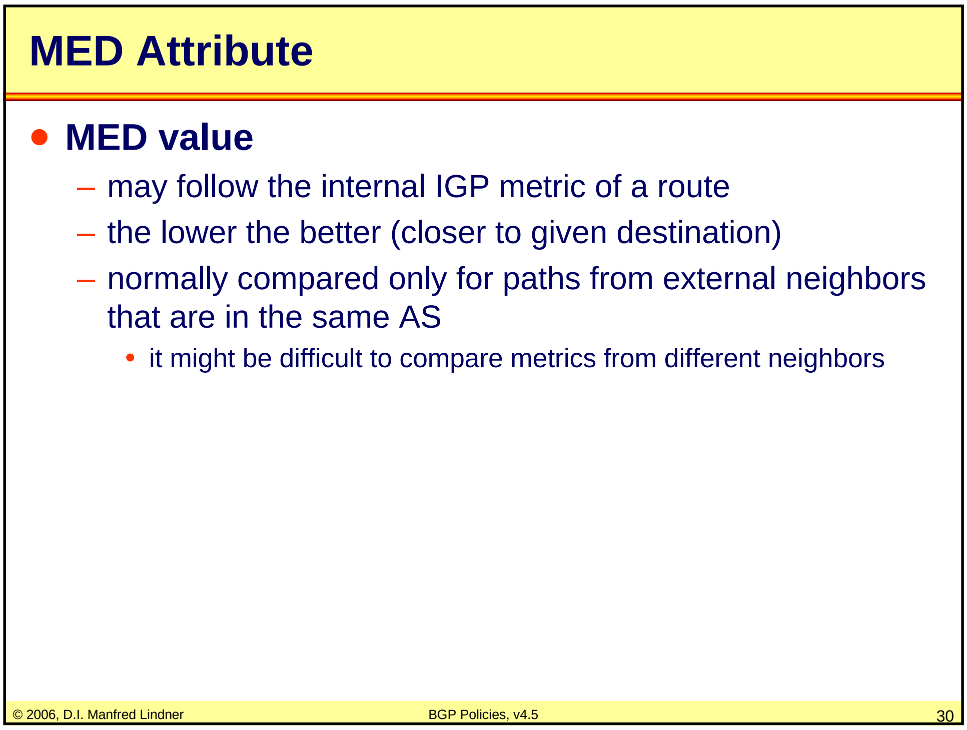
Agenda

- Introduction
- Next hop handling
- AS aggregation
- Preferences for outgoing traffic
- <u>Preferences for incoming traffic</u>
- Route origins
- Communities
- Routing decision details
- Routing policies



Multi Exit Discriminator Attribute

- **Multi Exit Discriminator = MED**
- **optional non-transitive attribute**
- **is a hint to external neighbors**
  - about the preferred path into an AS in case of multiple entrance points
  - “external BGP metric”
- **is exchanged between AS´s**
  - but a MED that comes into an AS does not leave the AS
  - MED value used for decision making within the AS
    - however, AS might decide to ignore it

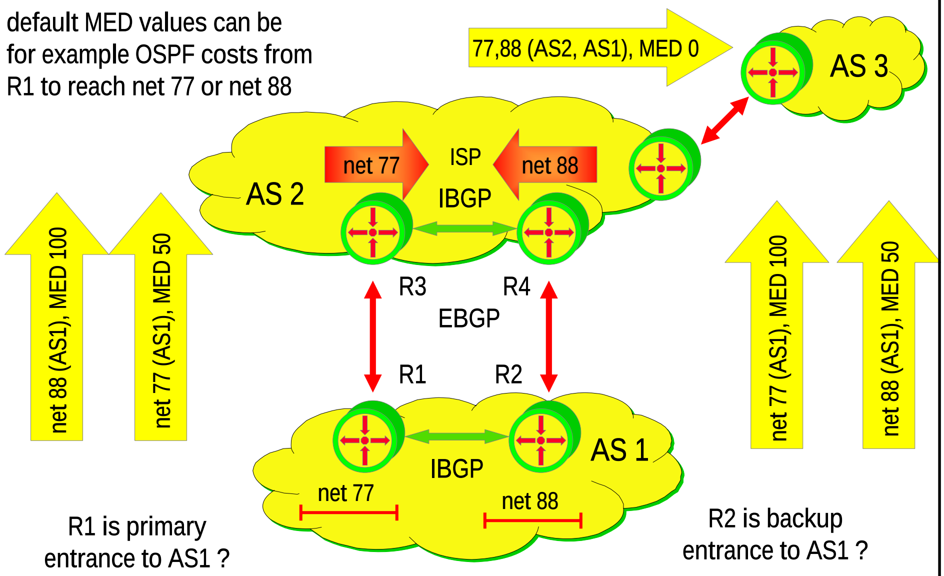


MED Attribute

- **MED value**
  - may follow the internal IGP metric of a route
  - the lower the better (closer to given destination)
  - normally compared only for paths from external neighbors that are in the same AS
    - it might be difficult to compare metrics from different neighbors

Scenario MED Default Behavior

default MED values can be for example OSPF costs from R1 to reach net 77 or net 88



77,88 (AS2, AS1), MED 0

AS 3

AS 2

net 77

ISP

net 88

IBGP

net 88 (AS1), MED 100

net 77 (AS1), MED 50

net 77 (AS1), MED 100

net 88 (AS1), MED 50

R3

R4

EBGP

R1

R2

AS 1

IBGP

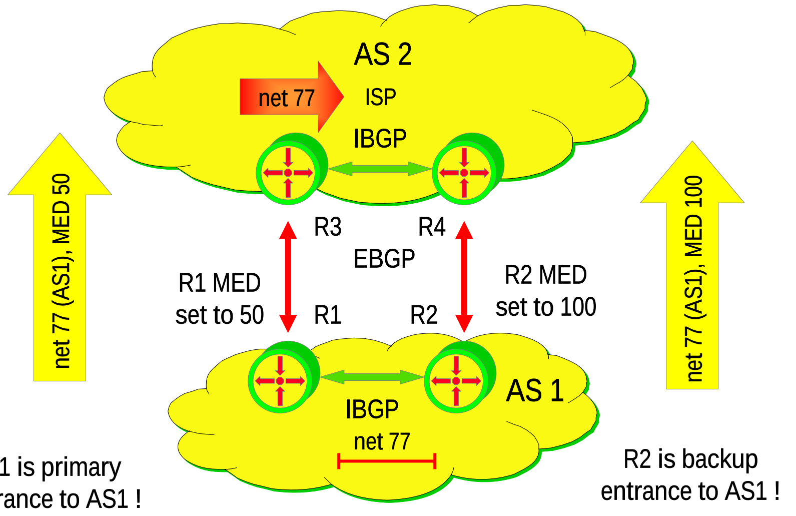
net 77

net 88

R1 is primary entrance to AS1 ?

R2 is backup entrance to AS1 ?

Scenario MED with Manual Set



AS 2

net 77

ISP

IBGP

net 77 (AS1), MED 50

net 77 (AS1), MED 100

R3

R4

EBGP

R1 MED set to 50

R1

R2

R2 MED set to 100

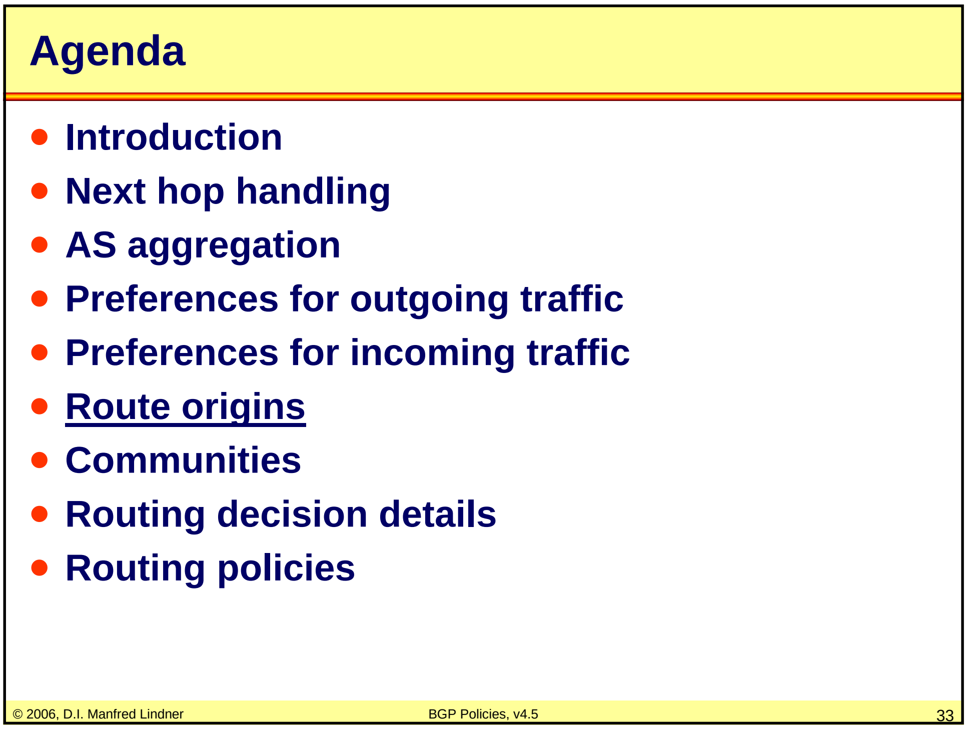
AS 1

IBGP

net 77

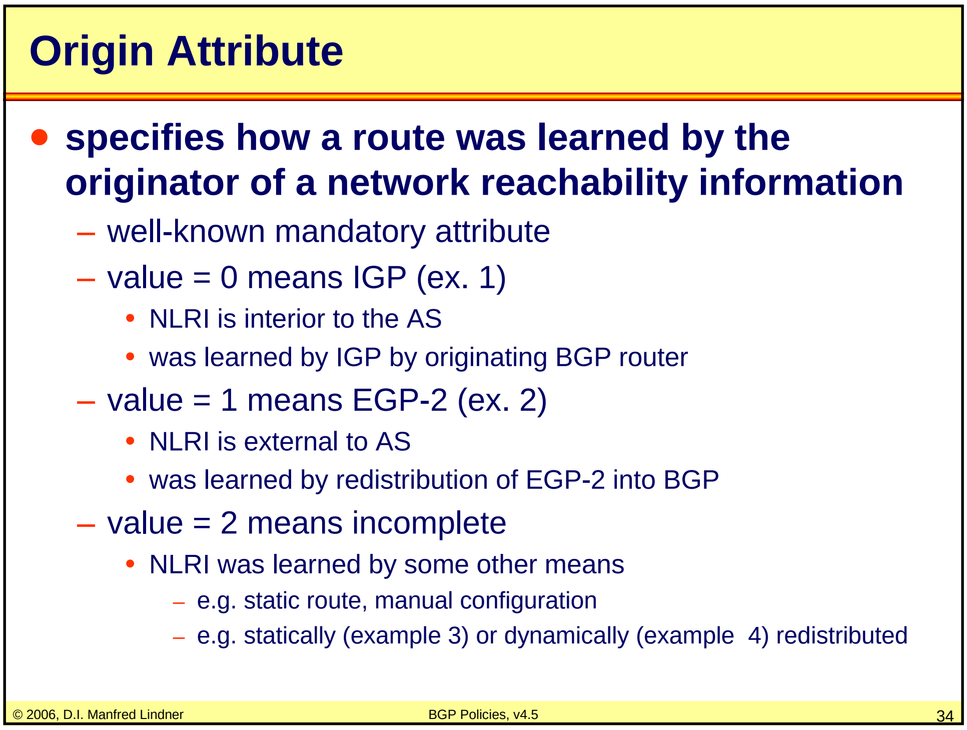
R1 is primary entrance to AS1 !

R2 is backup entrance to AS1 !



Agenda

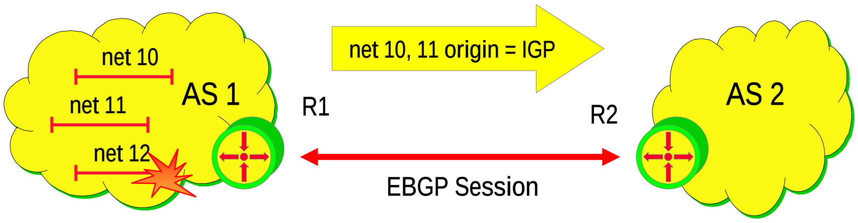
- Introduction
- Next hop handling
- AS aggregation
- Preferences for outgoing traffic
- Preferences for incoming traffic
- Route origins
- Communities
- Routing decision details
- Routing policies



Origin Attribute

- specifies how a route was learned by the originator of a network reachability information
  - well-known mandatory attribute
  - value = 0 means IGP (ex. 1)
    - NLRI is interior to the AS
    - was learned by IGP by originating BGP router
  - value = 1 means EGP-2 (ex. 2)
    - NLRI is external to AS
    - was learned by redistribution of EGP-2 into BGP
  - value = 2 means incomplete
    - NLRI was learned by some other means
      - e.g. static route, manual configuration
      - e.g. statically (example 3) or dynamically (example 4) redistributed

Origin Example 1

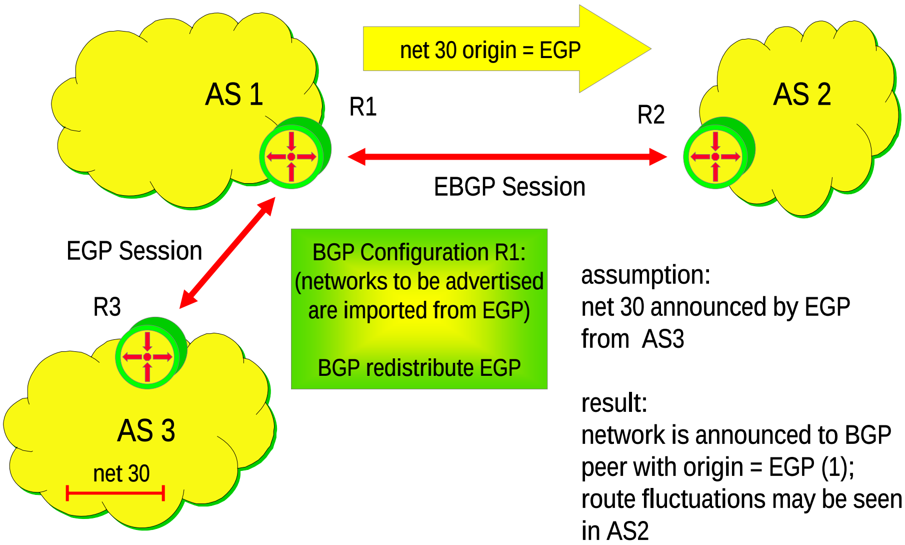


BGP Configuration R1:
(networks to be advertised
are manually specified)
network 10
network 11
network 12

assumption:
net 10 and net 11 reachable via IGP within AS1
net 12 not reachable via IGP within AS1
(does not exist in IP routing table of R1)

result:
only networks reachable via IGP are announced to BGP peer with origin = IGP (0);
route fluctuations will be seen in AS2

Origin Example 2

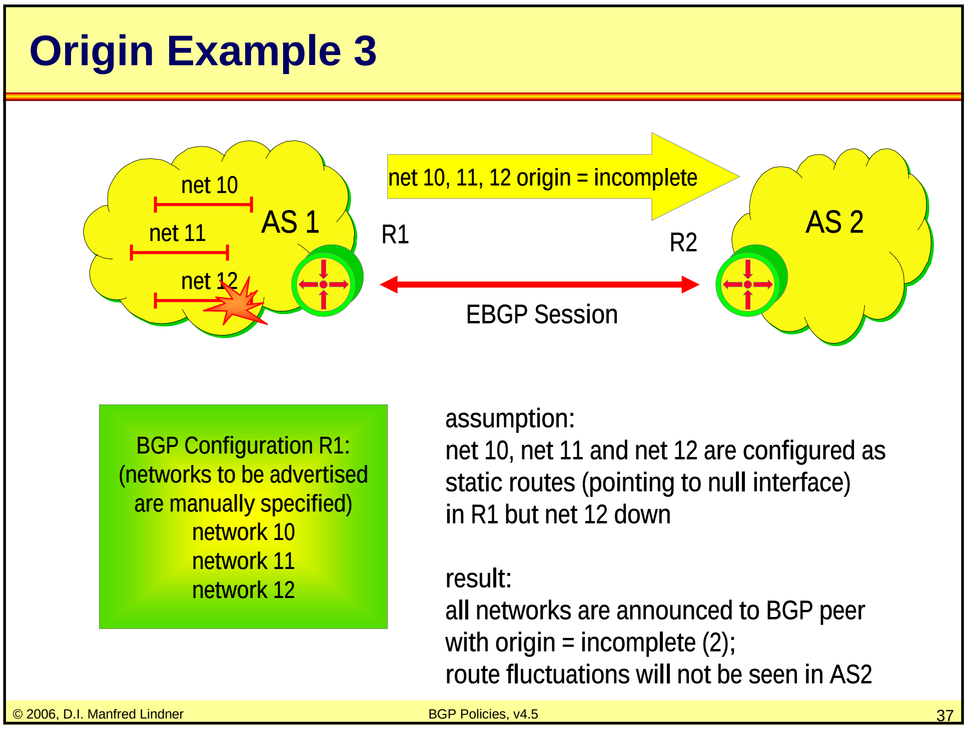


BGP Configuration R1:
(networks to be advertised
are imported from EGP)

BGP redistribute EGP

assumption:
net 30 announced by EGP
from AS3

result:
network is announced to BGP peer with origin = EGP (1);
route fluctuations may be seen in AS2



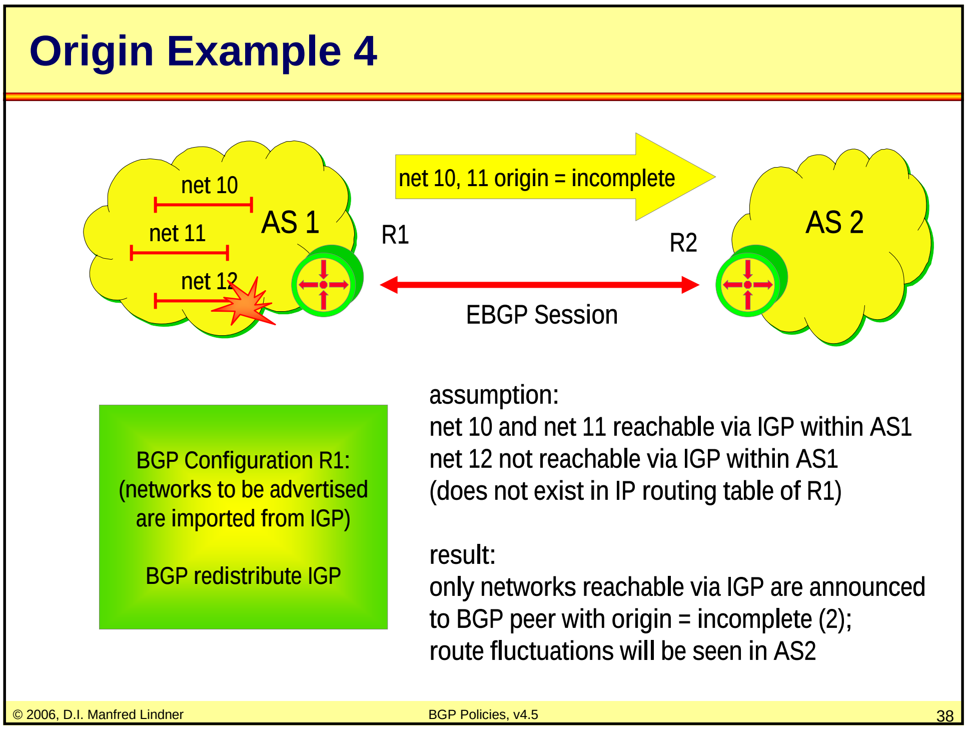
Origin Example 3

BGP Configuration R1:
(networks to be advertised are manually specified)
network 10
network 11
network 12

assumption:
net 10, net 11 and net 12 are configured as static routes (pointing to null interface) in R1 but net 12 down

result:
all networks are announced to BGP peer with origin = incomplete (2);
route fluctuations will not be seen in AS2

© 2006, D.I. Manfred Lindner BGP Policies, v4.5 37



Origin Example 4

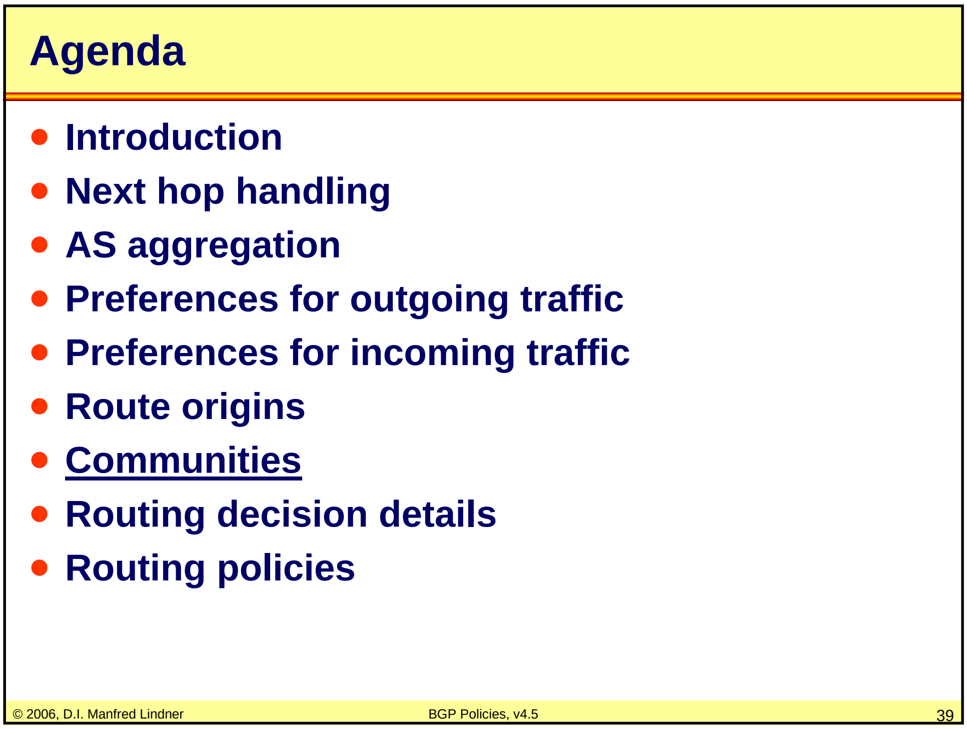
BGP Configuration R1:
(networks to be advertised are imported from IGP)

BGP redistribute IGP

assumption:
net 10 and net 11 reachable via IGP within AS1
net 12 not reachable via IGP within AS1
(does not exist in IP routing table of R1)

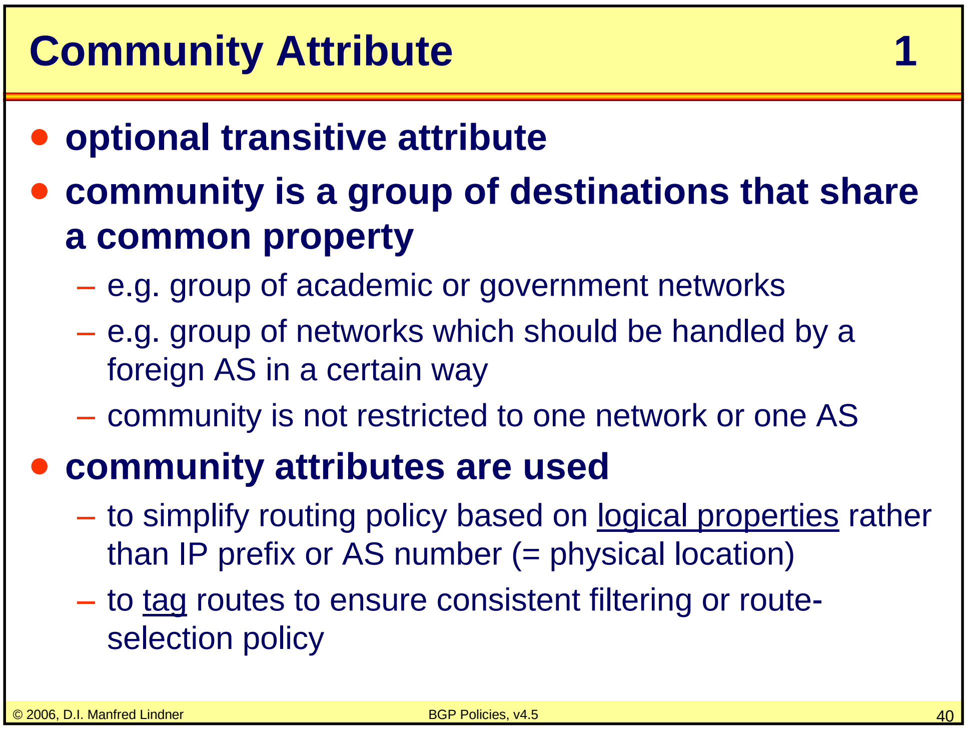
result:
only networks reachable via IGP are announced to BGP peer with origin = incomplete (2);
route fluctuations will be seen in AS2

© 2006, D.I. Manfred Lindner BGP Policies, v4.5 38



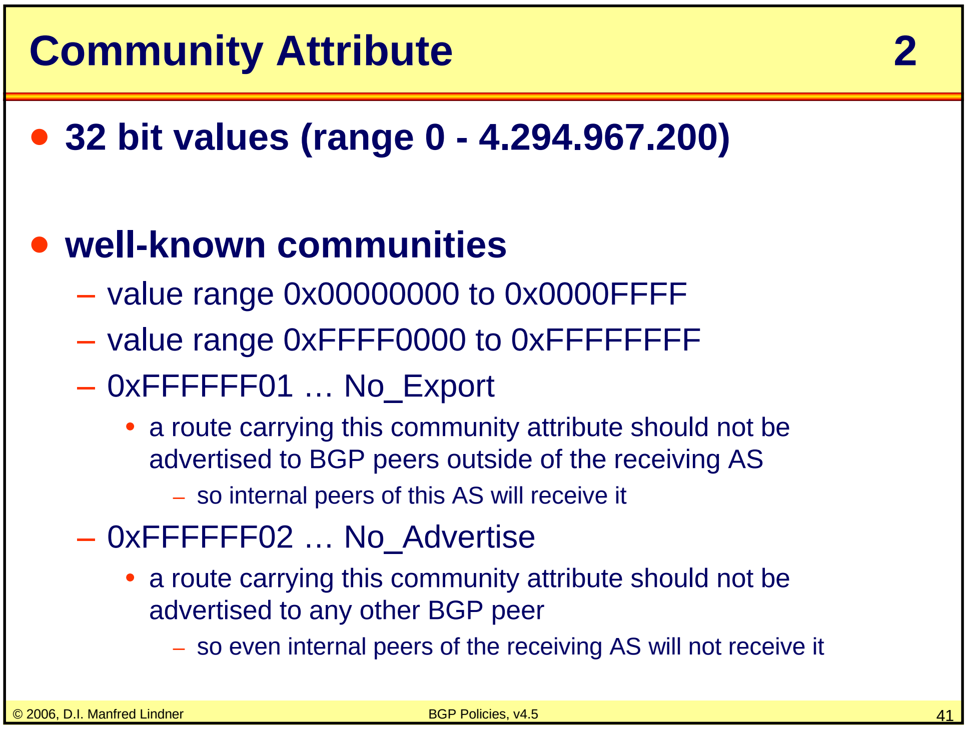
Agenda

- Introduction
- Next hop handling
- AS aggregation
- Preferences for outgoing traffic
- Preferences for incoming traffic
- Route origins
- Communities
- Routing decision details
- Routing policies



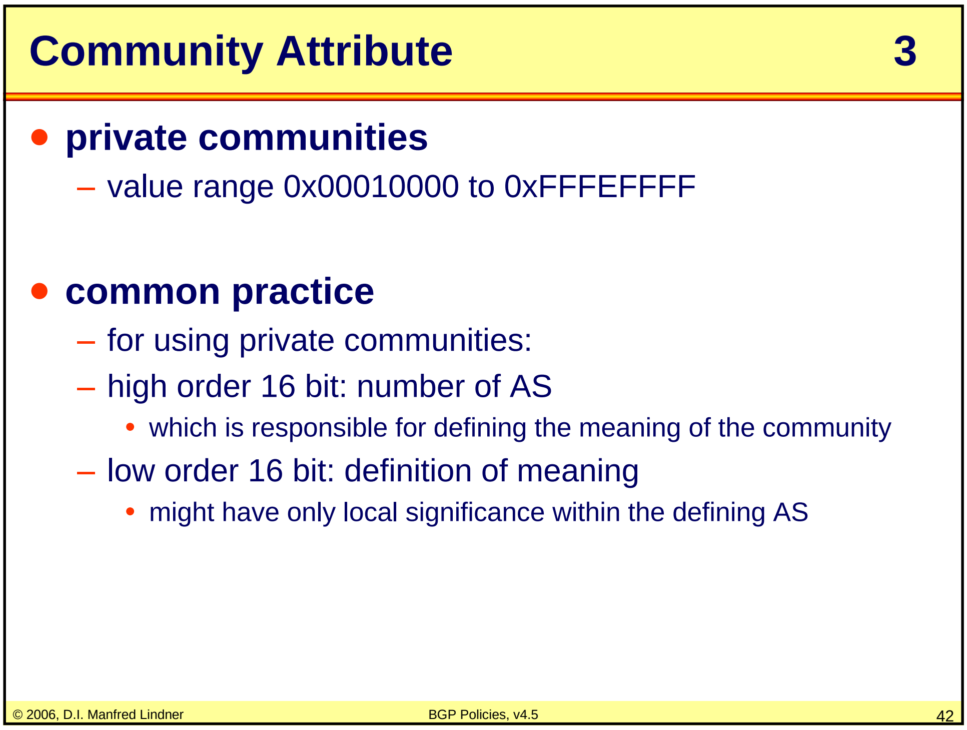
Community Attribute 1

- optional transitive attribute
- community is a group of destinations that share a common property
  - e.g. group of academic or government networks
  - e.g. group of networks which should be handled by a foreign AS in a certain way
  - community is not restricted to one network or one AS
- community attributes are used
  - to simplify routing policy based on logical properties rather than IP prefix or AS number (= physical location)
  - to tag routes to ensure consistent filtering or route-selection policy



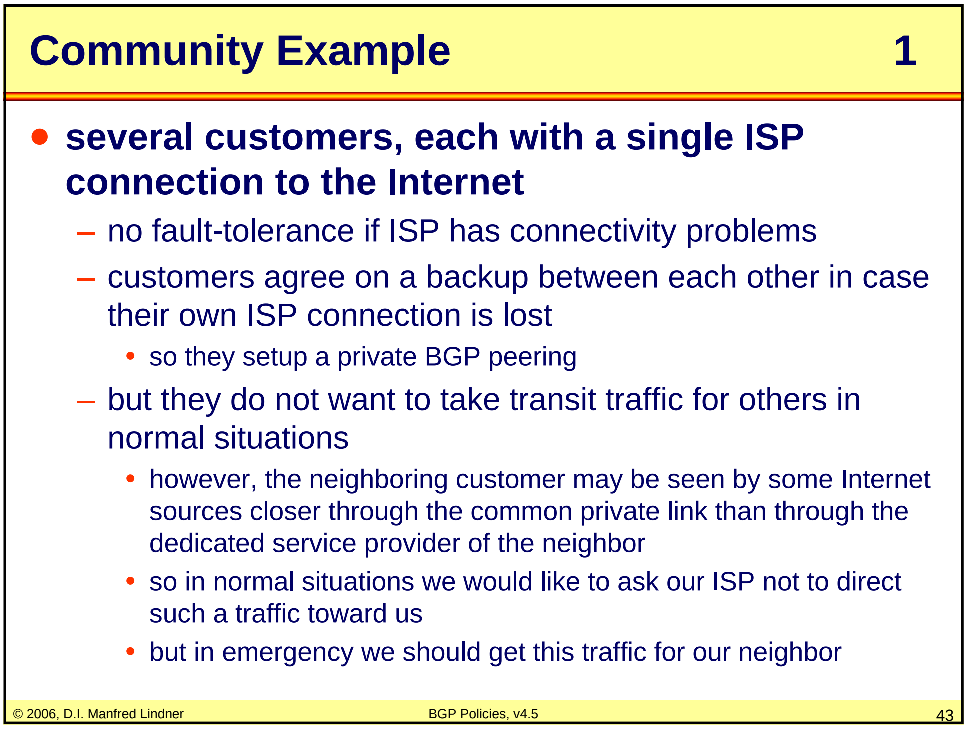
Community Attribute 2

- **32 bit values (range 0 - 4.294.967.200)**
- **well-known communities**
  - value range 0x00000000 to 0x0000FFFF
  - value range 0xFFFF0000 to 0xFFFFFFFF
  - 0xFFFFFF01 … No_Export
    - a route carrying this community attribute should not be advertised to BGP peers outside of the receiving AS
      - so internal peers of this AS will receive it
  - 0xFFFFFF02 … No_Advertise
    - a route carrying this community attribute should not be advertised to any other BGP peer
      - so even internal peers of the receiving AS will not receive it



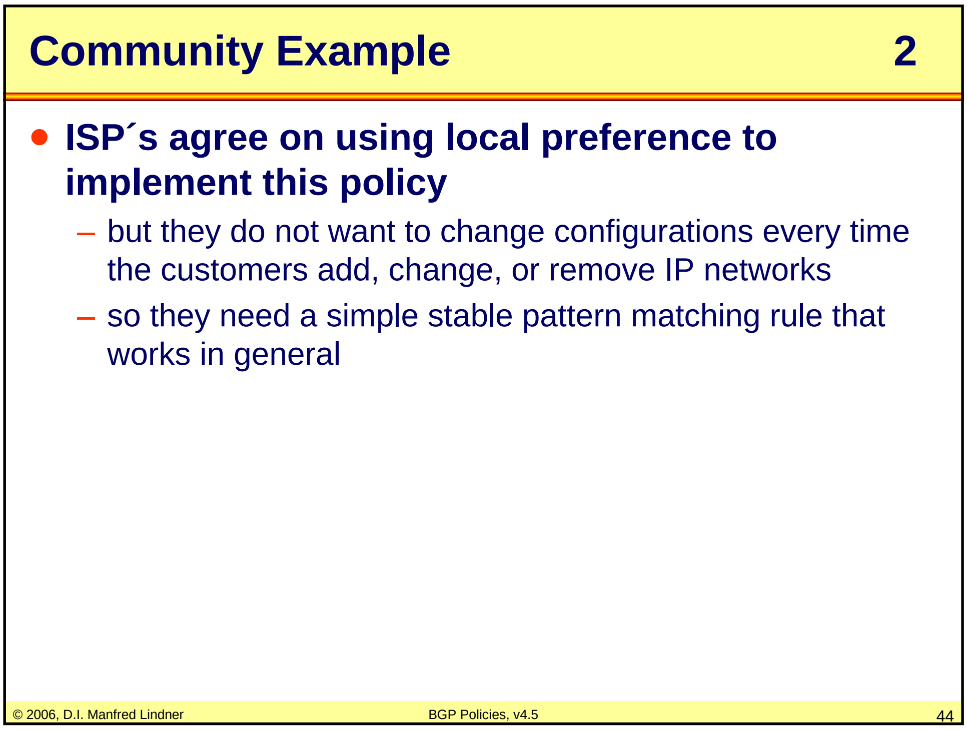
Community Attribute 3

- **private communities**
  - value range 0x00010000 to 0xFFFEFFFF
- **common practice**
  - for using private communities:
  - high order 16 bit: number of AS
    - which is responsible for defining the meaning of the community
  - low order 16 bit: definition of meaning
    - might have only local significance within the defining AS



Community Example 1

- **several customers, each with a single ISP connection to the Internet**
  - no fault-tolerance if ISP has connectivity problems
  - customers agree on a backup between each other in case their own ISP connection is lost
    - so they setup a private BGP peering
  - but they do not want to take transit traffic for others in normal situations
    - however, the neighboring customer may be seen by some Internet sources closer through the common private link than through the dedicated service provider of the neighbor
    - so in normal situations we would like to ask our ISP not to direct such a traffic toward us
    - but in emergency we should get this traffic for our neighbor



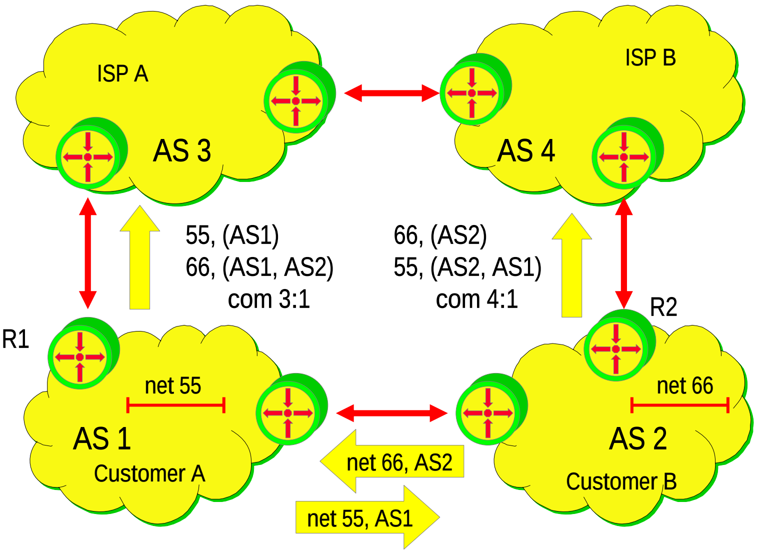
Community Example 2

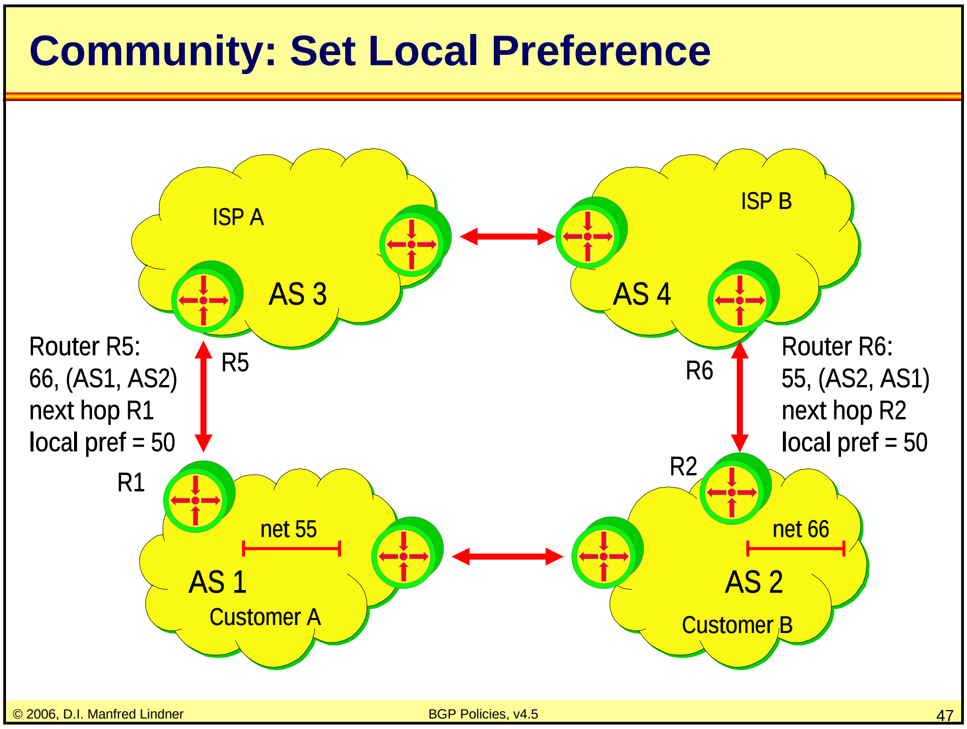
- **ISP´s agree on using local preference to implement this policy**
  - but they do not want to change configurations every time the customers add, change, or remove IP networks
  - so they need a simple stable pattern matching rule that works in general

Community Example 3

- **ISP´s define special community to identify routes which should be given lower local preference**
  - private community: <AS number of ISP>:1
    - 3:1
    - 4:1
  - if routes with certain condition match this community value, local preference should be reduced from the ISP’s default 100 to 50
  - remember: routes with larger local preference are preferred
- **Lets see how this works**

Community: Route Tagging



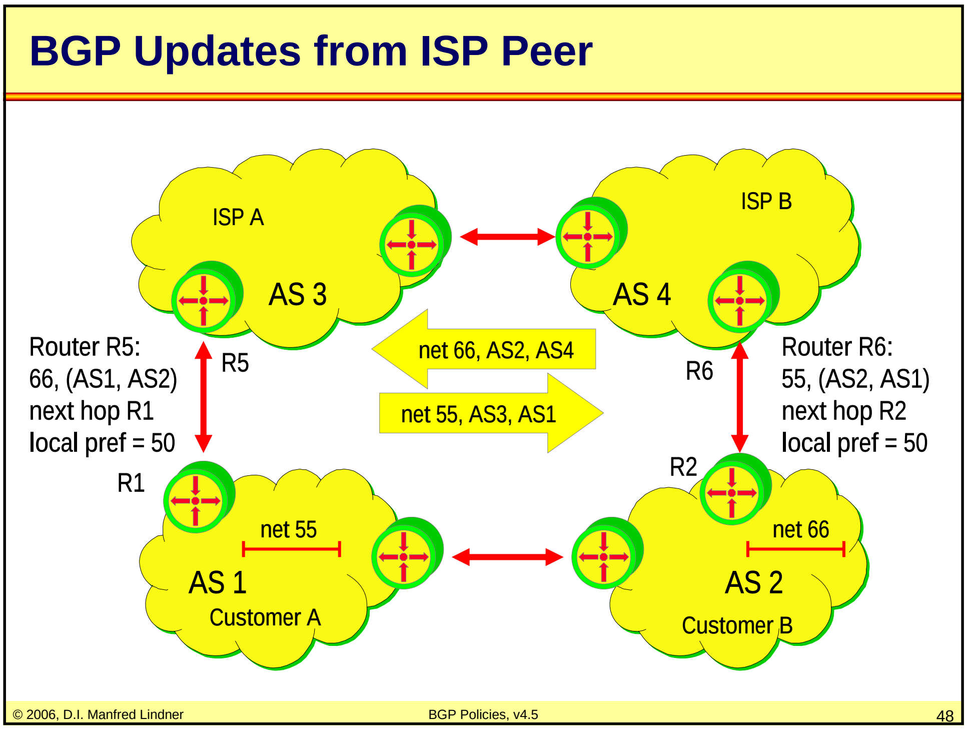


Community: Set Local Preference

Router R5:
66, (AS1, AS2)
next hop R1
local pref = 50

Router R6:
55, (AS2, AS1)
next hop R2
local pref = 50

© 2006, D.I. Manfred Lindner BGP Policies, v4.5 47

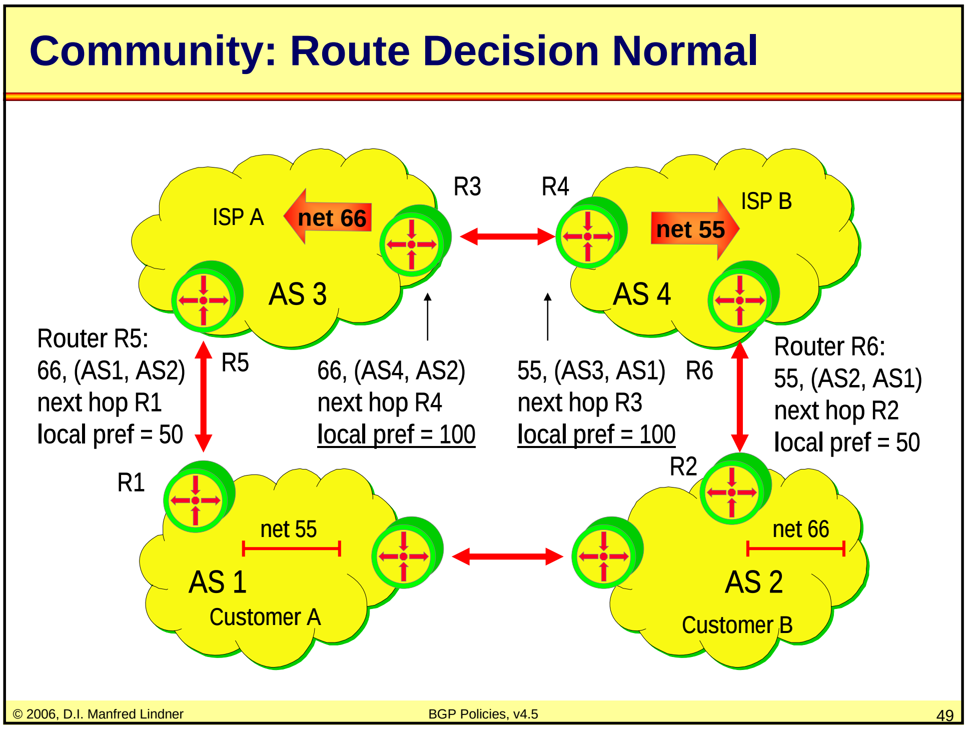


BGP Updates from ISP Peer

Router R5:
66, (AS1, AS2)
next hop R1
local pref = 50

Router R6:
55, (AS2, AS1)
next hop R2
local pref = 50

© 2006, D.I. Manfred Lindner BGP Policies, v4.5 48



Community: Route Decision Normal

ISP A

net 66

AS 3

R3

R4

net 55

ISP B

AS 4

Router R5:
66, (AS1, AS2)
next hop R1
local pref = 50

R5

66, (AS4, AS2)
next hop R4
local pref = 100

55, (AS3, AS1)
next hop R3
local pref = 100

R6

Router R6:
55, (AS2, AS1)
next hop R2
local pref = 50

R1

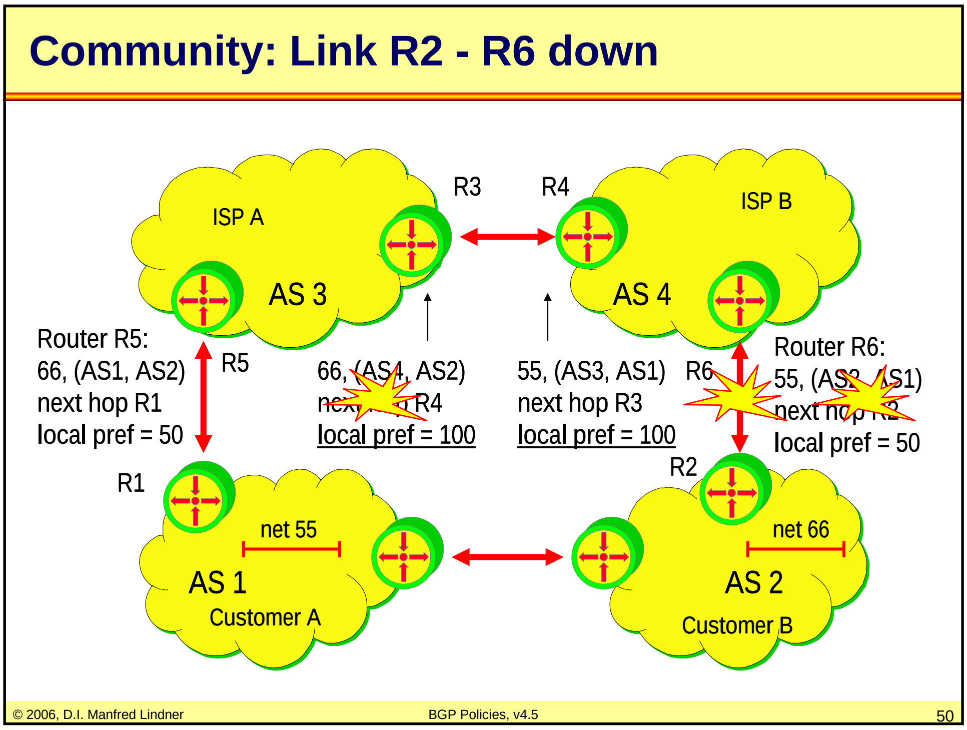
net 55

AS 1
Customer A

R2

net 66

AS 2
Customer B



Community: Link R2 - R6 down

ISP A

AS 3

R3

R4

ISP B

AS 4

Router R5:
66, (AS1, AS2)
next hop R1
local pref = 50

R5

66, (AS4, AS2)
next hop R4
local pref = 100

55, (AS3, AS1)
next hop R3
local pref = 100

R6

Router R6:
55, (AS2, AS1)
next hop R2
local pref = 50

R1

net 55

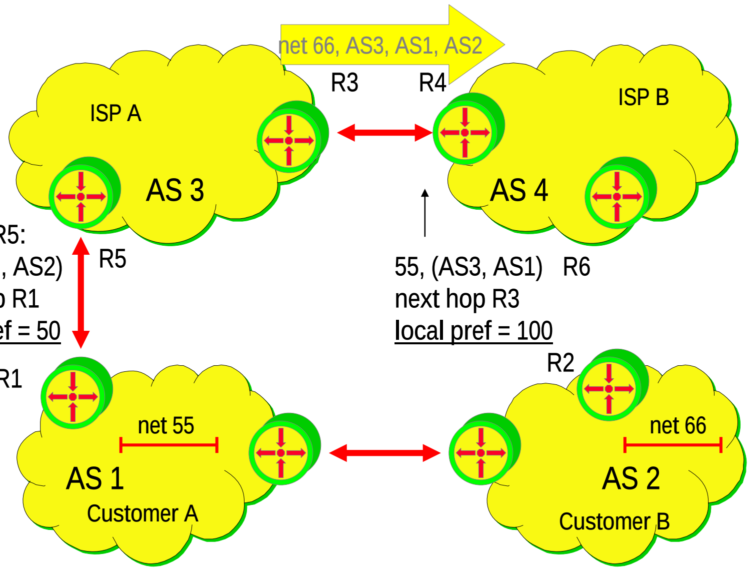
AS 1
Customer A

R2

net 66

AS 2
Customer B

Community: Route Decision Backup

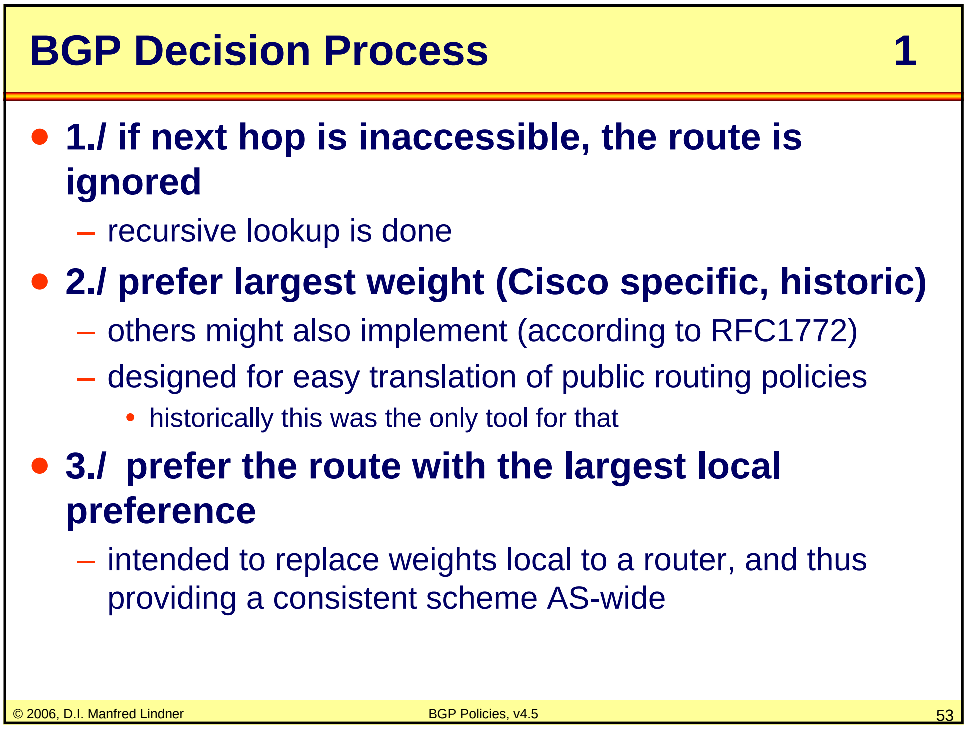


Router R5:
66, (AS1, AS2)
next hop R1
local pref = 50

55, (AS3, AS1)
next hop R3
local pref = 100

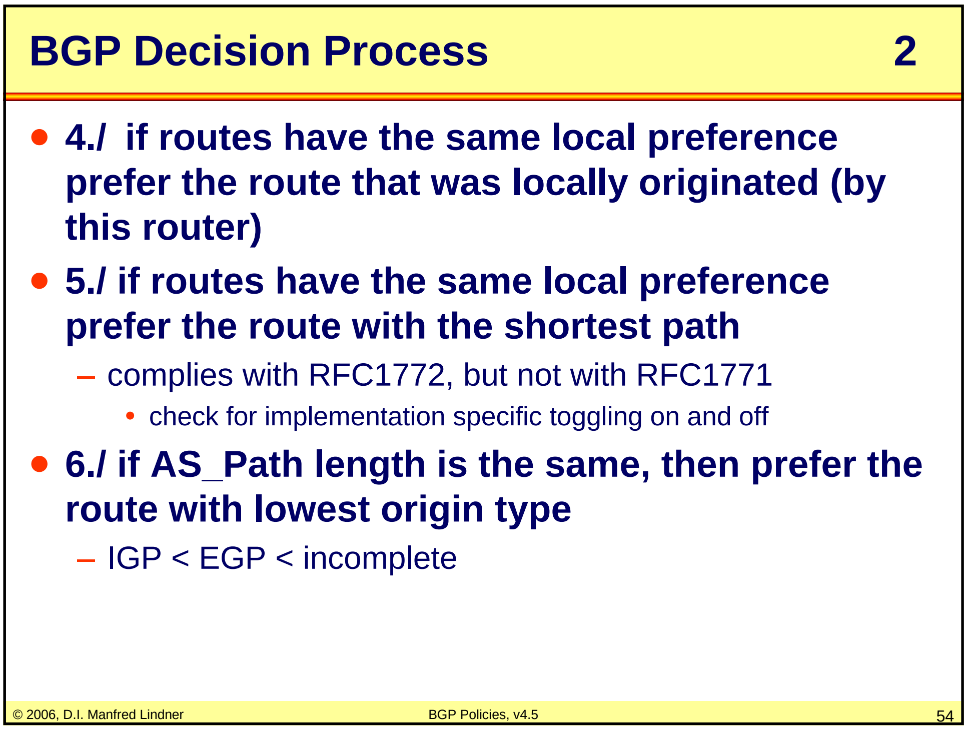
Agenda

- Introduction
- Next hop handling
- AS aggregation
- Preferences for outgoing traffic
- Preferences for incoming traffic
- Route origins
- Communities
- Routing decision details
- Routing policies



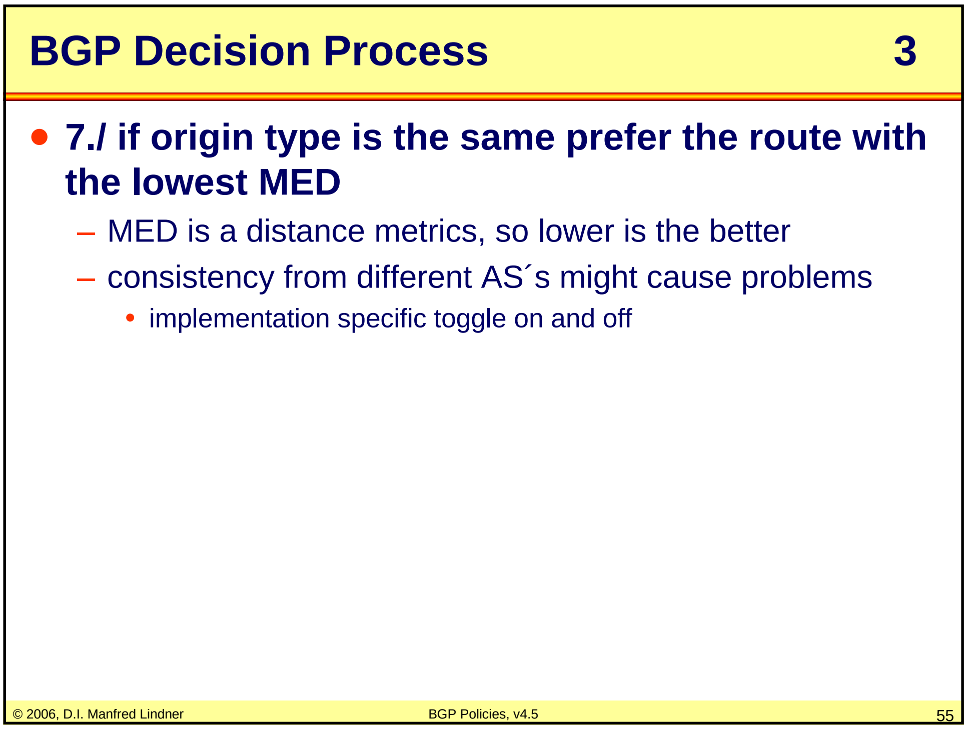
BGP Decision Process 1

- **1./ if next hop is inaccessible, the route is ignored**
  - recursive lookup is done
- **2./ prefer largest weight (Cisco specific, historic)**
  - others might also implement (according to RFC1772)
  - designed for easy translation of public routing policies
    - historically this was the only tool for that
- **3./ prefer the route with the largest local preference**
  - intended to replace weights local to a router, and thus providing a consistent scheme AS-wide



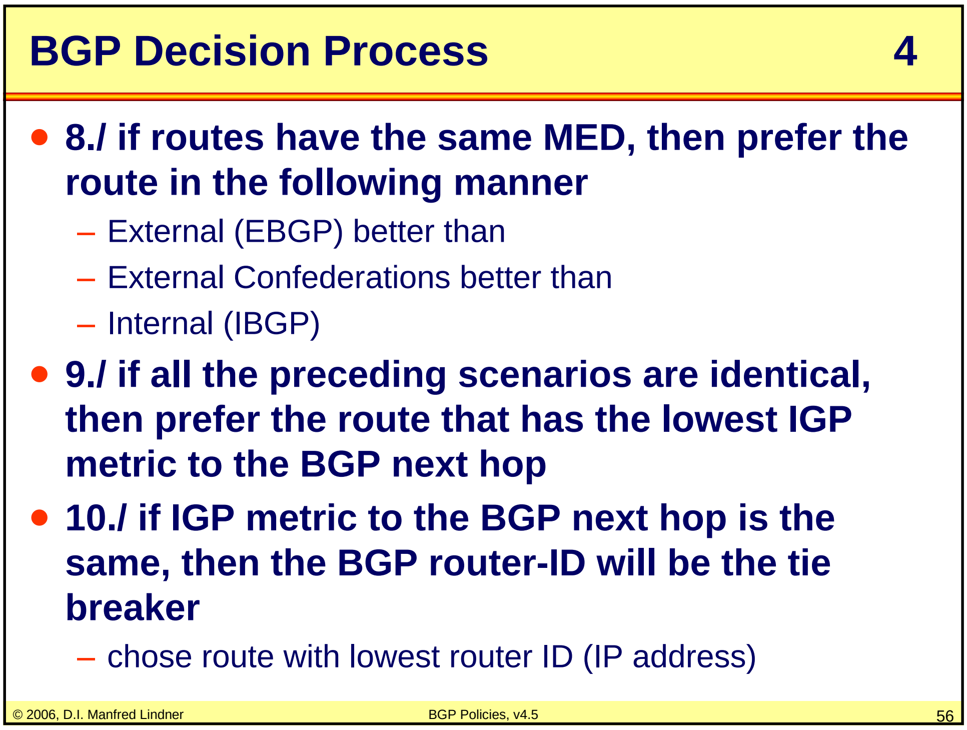
BGP Decision Process 2

- **4./ if routes have the same local preference prefer the route that was locally originated (by this router)**
- **5./ if routes have the same local preference prefer the route with the shortest path**
  - complies with RFC1772, but not with RFC1771
    - check for implementation specific toggling on and off
- **6./ if AS_Path length is the same, then prefer the route with lowest origin type**
  - IGP < EGP < incomplete



BGP Decision Process 3

- **7./ if origin type is the same prefer the route with the lowest MED**
  - – MED is a distance metrics, so lower is the better
  - – consistency from different AS´s might cause problems
    - implementation specific toggle on and off



BGP Decision Process 4

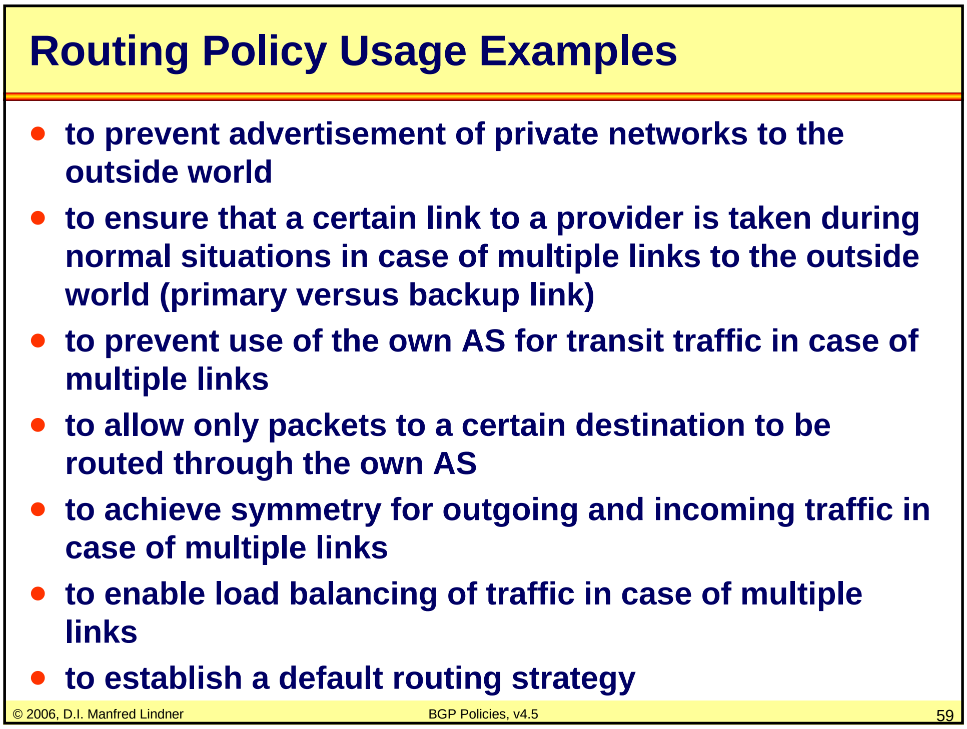
- **8./ if routes have the same MED, then prefer the route in the following manner**
  - – External (EBGP) better than
  - – External Confederations better than
  - – Internal (IBGP)
- **9./ if all the preceding scenarios are identical, then prefer the route that has the lowest IGP metric to the BGP next hop**
- **10./ if IGP metric to the BGP next hop is the same, then the BGP router-ID will be the tie breaker**
  - – chose route with lowest router ID (IP address)

Agenda

- Introduction
- Next hop handling
- AS aggregation
- Preferences for outgoing traffic
- Preferences for incoming traffic
- Route origins
- Communities
- Routing decision details
- <u>Routing policies</u>

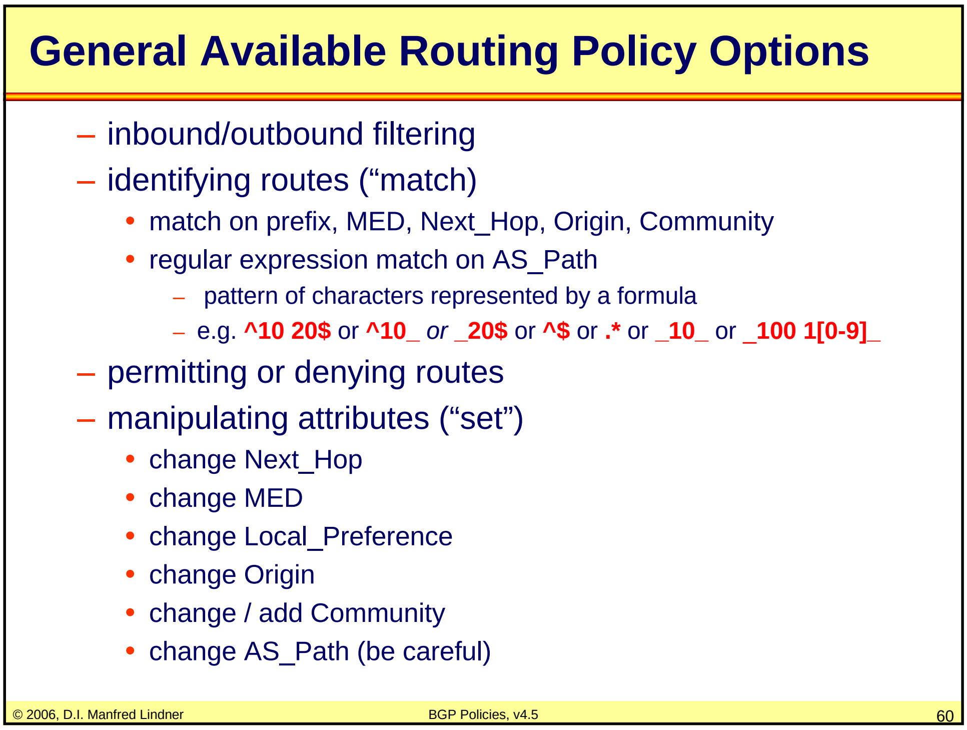
Routing Policy

- routing policies determine what routing information is exchanged with other AS´s
- can be implemented by filtering and manipulating BGP routes
- some attributes determine policy by their definition
  - AS_Path can be used to discard any route that passes a certain AS
  - MED can be used to distinguish between multiple exits of an AS to a neighbor AS
- NLRI (IP prefix, length) itself may be used for policy



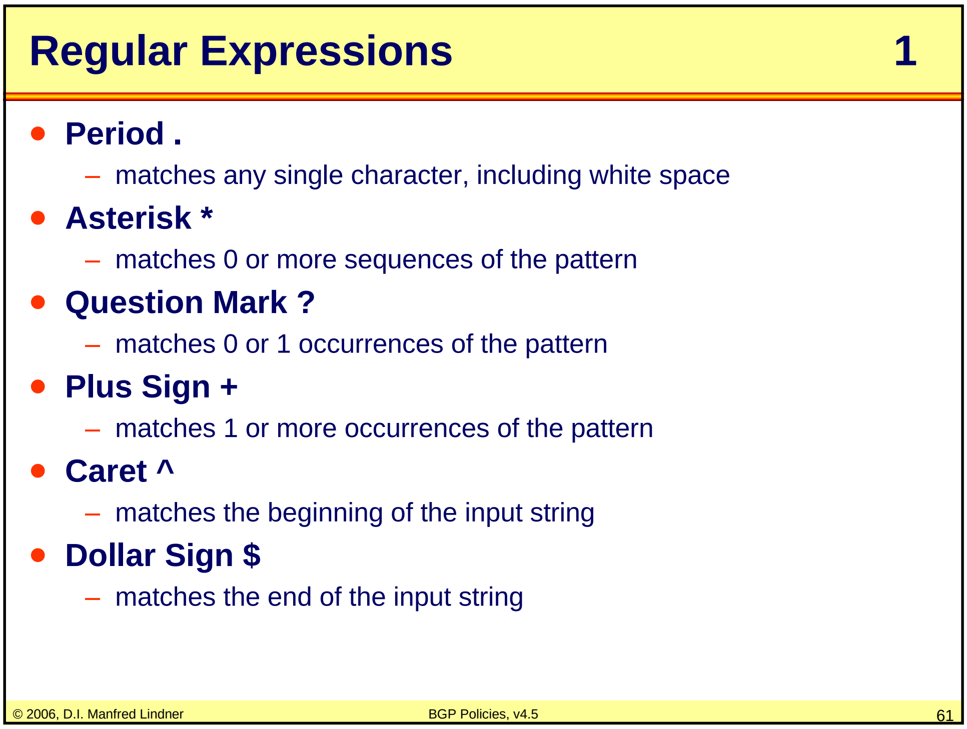
Routing Policy Usage Examples

- to prevent advertisement of private networks to the outside world
- to ensure that a certain link to a provider is taken during normal situations in case of multiple links to the outside world (primary versus backup link)
- to prevent use of the own AS for transit traffic in case of multiple links
- to allow only packets to a certain destination to be routed through the own AS
- to achieve symmetry for outgoing and incoming traffic in case of multiple links
- to enable load balancing of traffic in case of multiple links
- to establish a default routing strategy



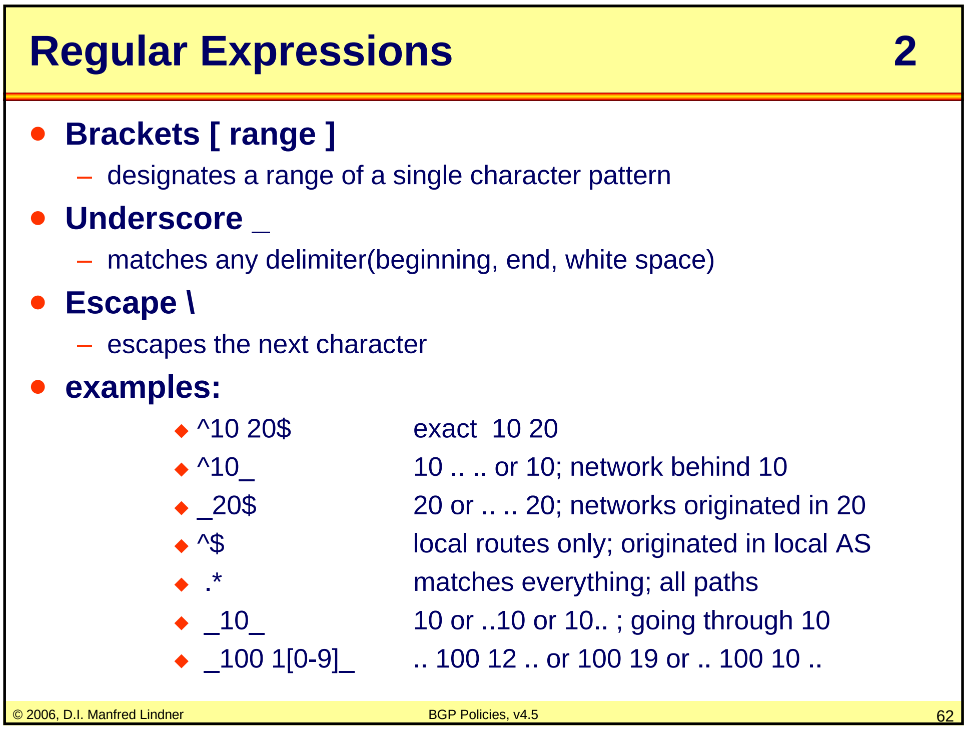
General Available Routing Policy Options

- inbound/outbound filtering
- identifying routes (“match)
  - match on prefix, MED, Next_Hop, Origin, Community
  - regular expression match on AS_Path
    - pattern of characters represented by a formula
    - e.g. **^10 20$** or **^10_** *or* **_20$** or **^$** or **.*** or **_10_** or **_100 1[0-9]_**
- permitting or denying routes
- manipulating attributes (“set”)
  - change Next_Hop
  - change MED
  - change Local_Preference
  - change Origin
  - change / add Community
  - change AS_Path (be careful)



Regular Expressions 1

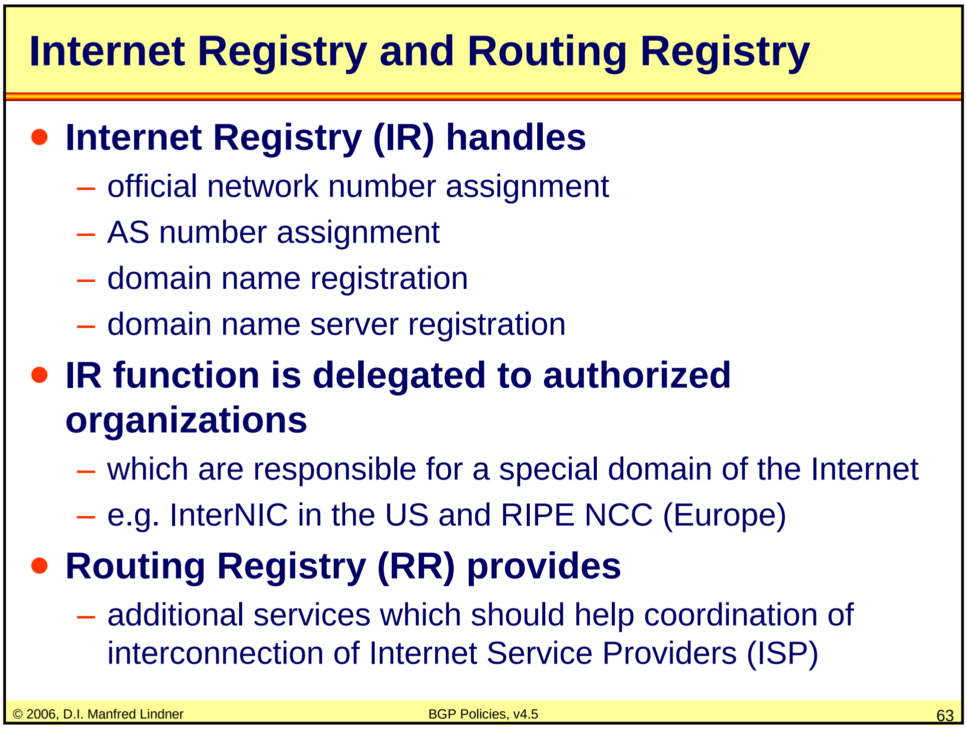
- **Period .**
  - matches any single character, including white space
- **Asterisk ***
  - matches 0 or more sequences of the pattern
- **Question Mark ?**
  - matches 0 or 1 occurrences of the pattern
- **Plus Sign +**
  - matches 1 or more occurrences of the pattern
- **Caret ^**
  - matches the beginning of the input string
- **Dollar Sign $**
  - matches the end of the input string



Regular Expressions 2

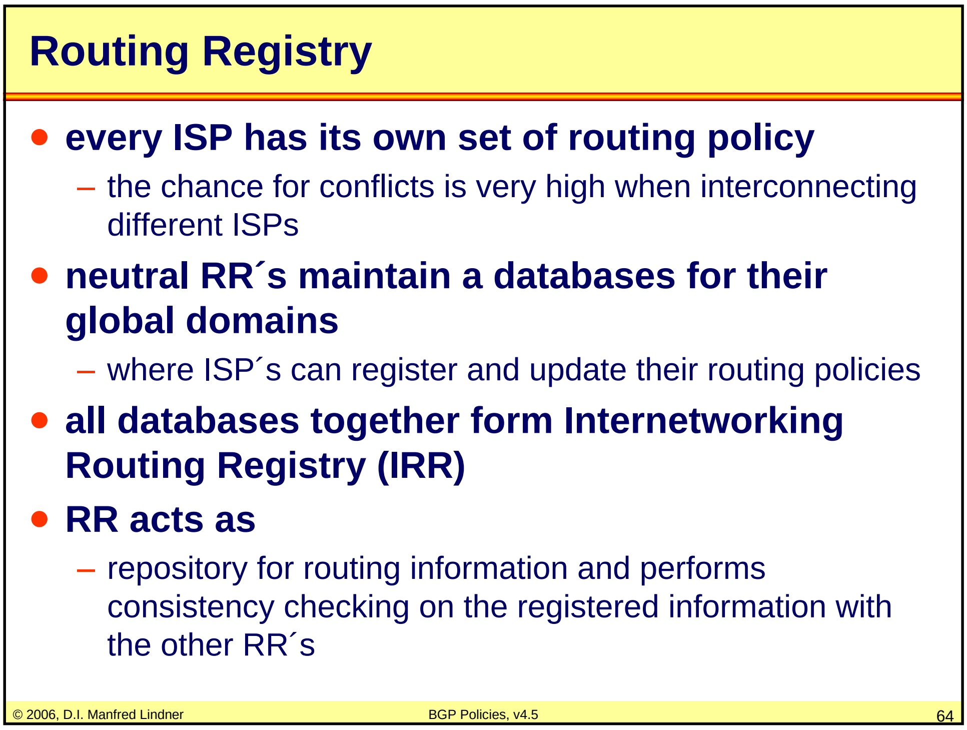
- **Brackets [range]**
  - designates a range of a single character pattern
- **Underscore _**
  - matches any delimiter(beginning, end, white space)
- **Escape \**
  - escapes the next character
- **examples:**

^10 20$	exact 10 20
^10_	10 .. .. or 10; network behind 10
_20$	20 or .. .. 20; networks originated in 20
^$	local routes only; originated in local AS
.*	matches everything; all paths
_10_	10 or ..10 or 10.. ; going through 10
_100 1[0-9]_	.. 100 12 .. or 100 19 or .. 100 10 ..



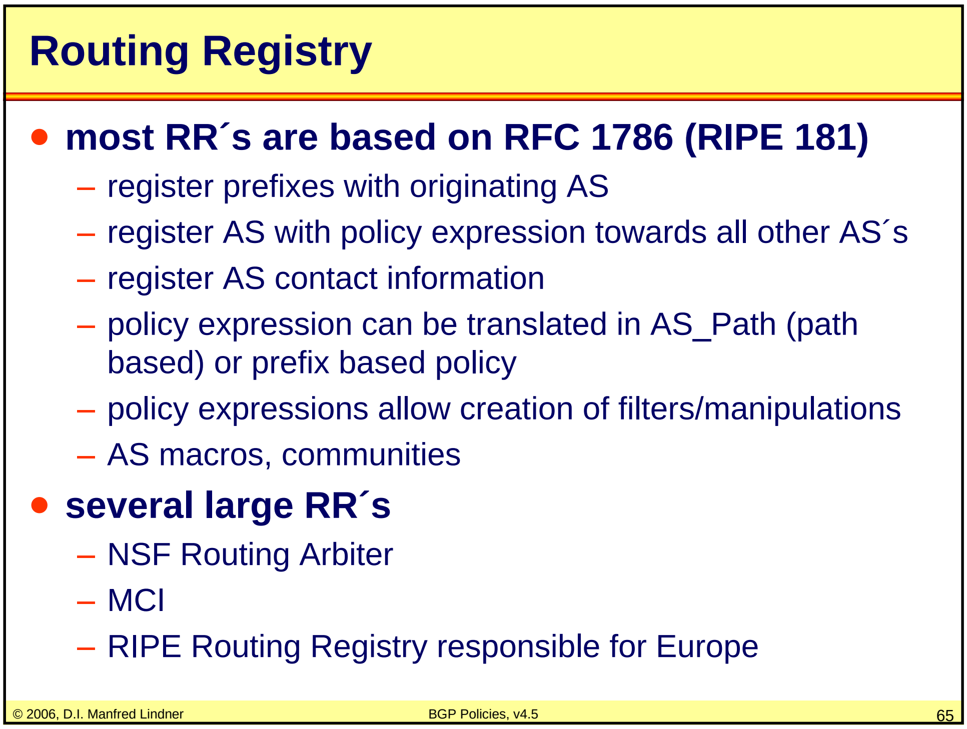
Internet Registry and Routing Registry

- Internet Registry (IR) handles
  - official network number assignment
  - AS number assignment
  - domain name registration
  - domain name server registration
- IR function is delegated to authorized organizations
  - which are responsible for a special domain of the Internet
  - e.g. InterNIC in the US and RIPE NCC (Europe)
- Routing Registry (RR) provides
  - additional services which should help coordination of interconnection of Internet Service Providers (ISP)



Routing Registry

- every ISP has its own set of routing policy
  - the chance for conflicts is very high when interconnecting different ISPs
- neutral RR´s maintain a databases for their global domains
  - where ISP´s can register and update their routing policies
- all databases together form Internetworking Routing Registry (IRR)
- RR acts as
  - repository for routing information and performs consistency checking on the registered information with the other RR´s



Routing Registry

- **most RR´s are based on RFC 1786 (RIPE 181)**
  - register prefixes with originating AS
  - register AS with policy expression towards all other AS´s
  - register AS contact information
  - policy expression can be translated in AS_Path (path based) or prefix based policy
  - policy expressions allow creation of filters/manipulations
  - AS macros, communities
- **several large RR´s**
  - NSF Routing Arbiter
  - MCI
  - RIPE Routing Registry responsible for Europe