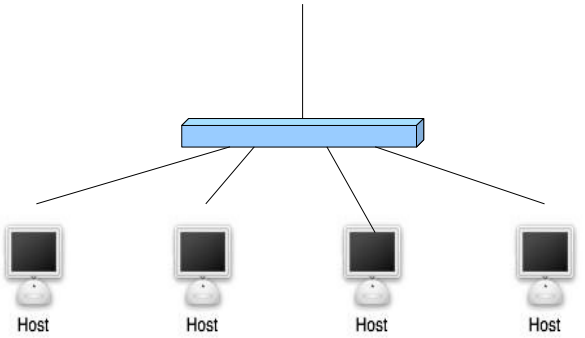
Appunti sullo studio dell'InterWorking

Poco più di 15 slides per un'overview su IP interworking tra autonomous system, BGP e Route Reflector

by Ciro Carbone

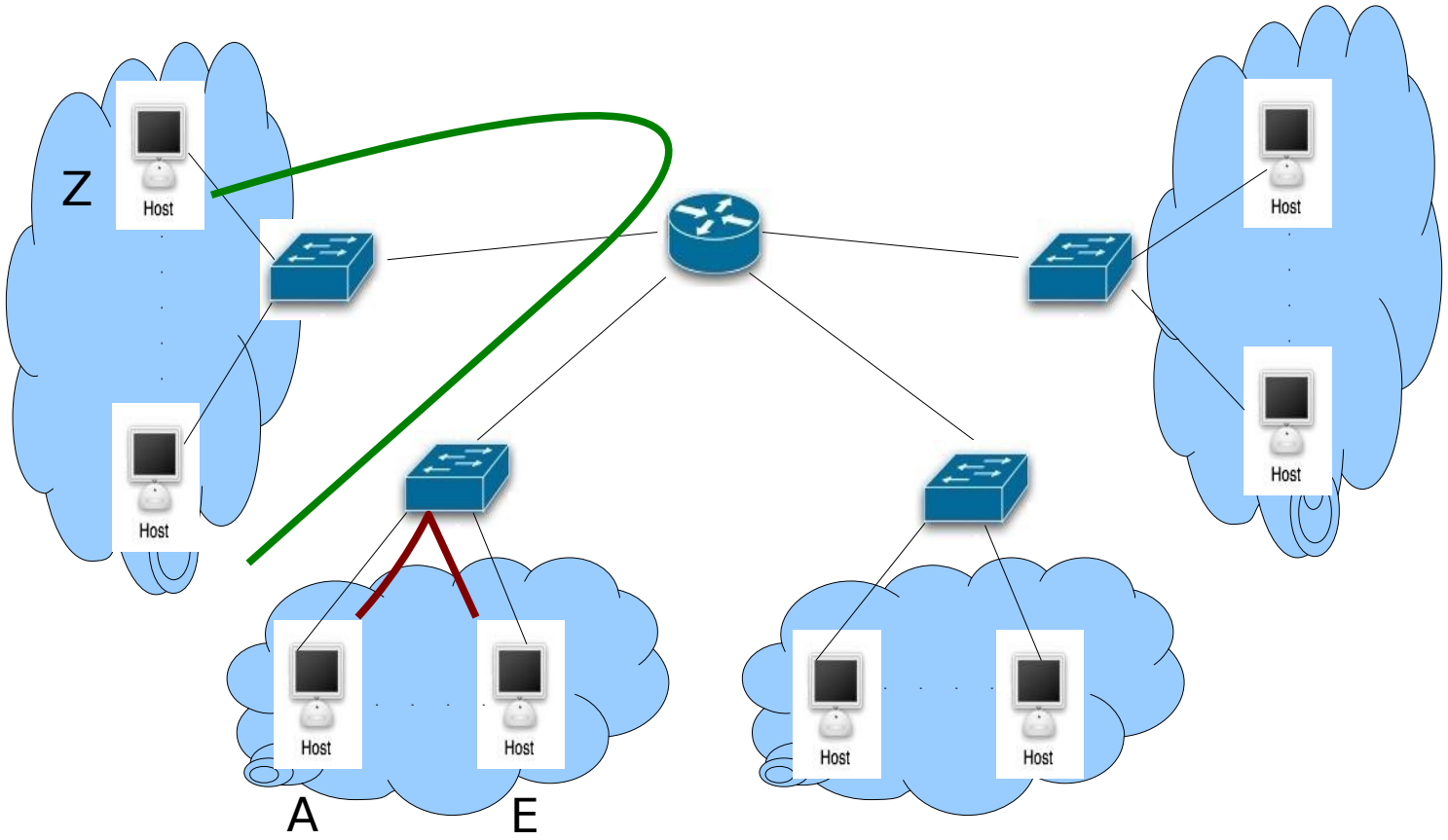
1.rivedere il cablaggio dei RR

in una piccola rete domestica collegata ad internet, basterebbe collegare i vari hosts ad un HUB o uno Switch ed avere un via di gateway verso il mondo esterno. Tuttavia, oggi, esistono soluzioni integrate embedded con funzioni COMBO che consentono di acquistare un unico dispositivo chiamato ADSL Gateway Router. Questi possono avere la possibilità anche di interfaccie radio come il WiFi.



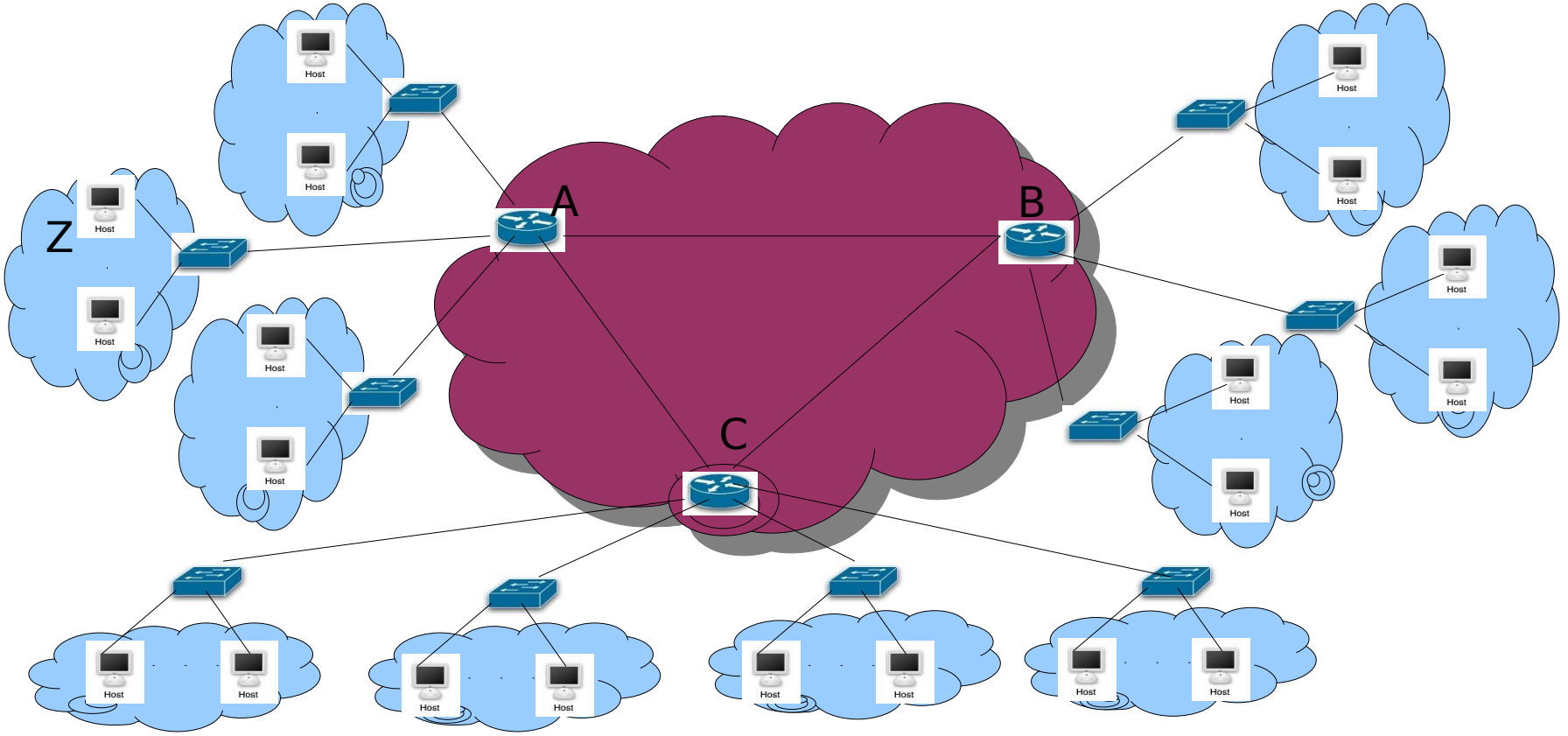
questi dispositivi hanno funzionalità di routers, lavorano quindi sul livello 3 (livello IP) leggendo gli indirizzi IP assegnati agli hosts e se un un host vuole colloquiare con un dispositivo con indirizzo ip diffetrente da quello del proprio collision Domain, il router provvede ad instradare i pacchetti verso il gateway, ovvero verso la linea telefonica.

Se dovessi essere in possesso di molti hosts, potrei suddividere la mia rete LAN in più collision domain, aiutandomi con gli Switch. Lo Switch lavora a livello 2(nel caso di una rete LAN, quindi, uno switch commuta le frames ethernet e non “guarda” gli indirizzi IP ma solo i MAC address). Lo Switch è l'evoluzione moderna degli HUB o del cablaggio a bus che si usavano una volta. Nella figura sotto avremo 4 collision domain (nelle nuvole) raggruppati da switch edun Router che funge da instradatore di pacchetti tra i vari collision domain.



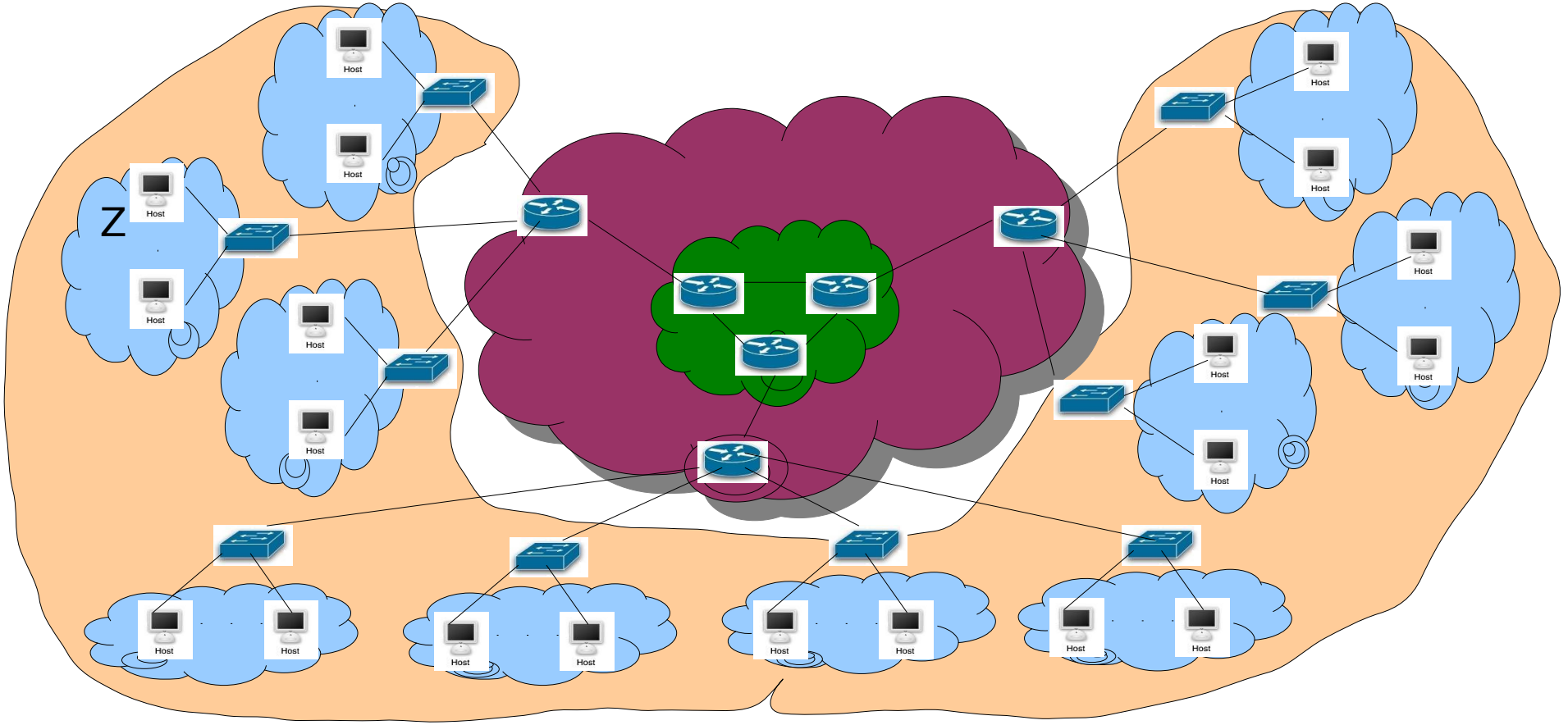
Nella rete sopra, gli switch provvedono a commutare gli hosts “gurdando” MAC frames ethernet (A che vuole parlare con E) e il Router provvede a commutare gli hosts dei C.D. qualora gli hosts vogliano parlarsi tra CD differenti (A vuole parlare con Z), “guardando” il pacchetto IP a livello più alto degli Switch.

Se la rete dovesse essere troppo grande, cioè il numero degli hosts dovesse essere davvero elevato ed anche i CD molti, immaginate quale possa essere il carico di lavoro del router centrale. In tal caso sarebbe più opportuno far condividere il lavoro di routing a più router a non ad uno solo.



Con ben tre router, posso ridistribuire meglio il carico che un solo router avrebbe potuto non permettersi. I 3 Router dovranno essere cablati tra loro per consentire l'interoperabilità a 360° sul routing di tutte le rotte (magliatura completa o “fully-meshed”). Ciascun Router ha una tabellina di routing per instradare le rotte correttamente. Finchè queste tabelline fossero dei soli tre router, si potrebbero anche compilare manualmente ma con lo svantaggio che se nella rete dovesse cambiare qualcosa, tutte e tre le tabelle devono essere ricompilate manualmente. Inoltre se la rete è ancor più grande, il numero di router impattati potrebbe essere ben superiore a tre e la compilazione delle tabelle di routing diverrebbe dispendiosa, specie consideranto eventuali cambiamenti della morfologia topologica della rete stessa.
Si è perciò ovviato a questo inconveniente realizzando dei protocolli particolari cosidetti di “INTERWORKING”. Questi protocolli danno la possibilità ai vari router di “parlarsi” tra loro indipendentemente da tutto ciò che
accade sulla rete e indipendentemente dal traffico normale generato tra gli hosts. Questi protocolli di interworking, permettendo ai router di palare tra loro non solo per scambiare ed instradare il normale traffico dagli hosts, ma di auto-istruirsi sulle loro rotte, sul “sentire” quale altre funzioni svolge il router collegato alla sua porta e così facendo ciascun router, “parlando” con gli altri suoi router, si crea una mappa abbastanza precisa dell'intera rete in cui esso né è un tassello. Così facendo, grazie all'interworking, i router saranno in grado autonomamente di apprendere la rete in cui essi stessi sono cablati ed auto-costruirsi ciascuno la propria tabella di routing. Un insieme di router con interworking prende, perciò, il nome di: Autonomous System.

Approfittando della figura sopra è possibile aprire un piccola parentesi e accennare ad un altro importante concetto sul design delle grandi reti: la distribuzione dei layers. Infatti, osservando la figura sotto, è facile comprendere la distinzione tra i tre layer fondamentali in cui una grande rete dovrebbe essere sempre composta: layer di accesso, layer di distribuzione e layer di core o di backbone.



Access layer

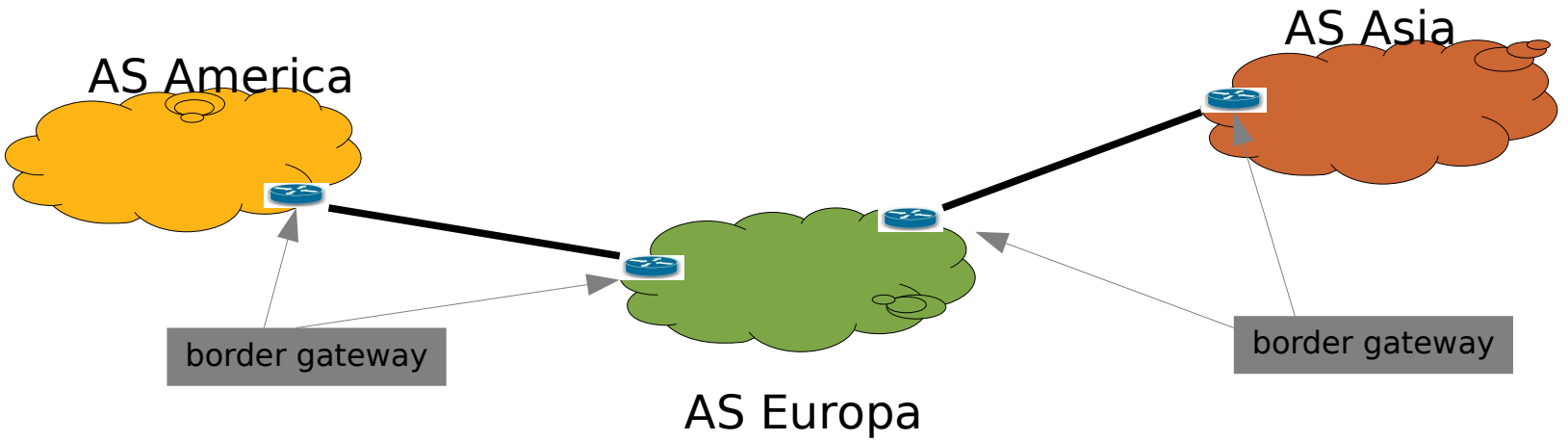
Distribution layer

Core layer

Questi protocolli di interworking si dividono in due categorie: gli IGP (Interior Gateway Protocol) e EGP (External Gateway Protocol).

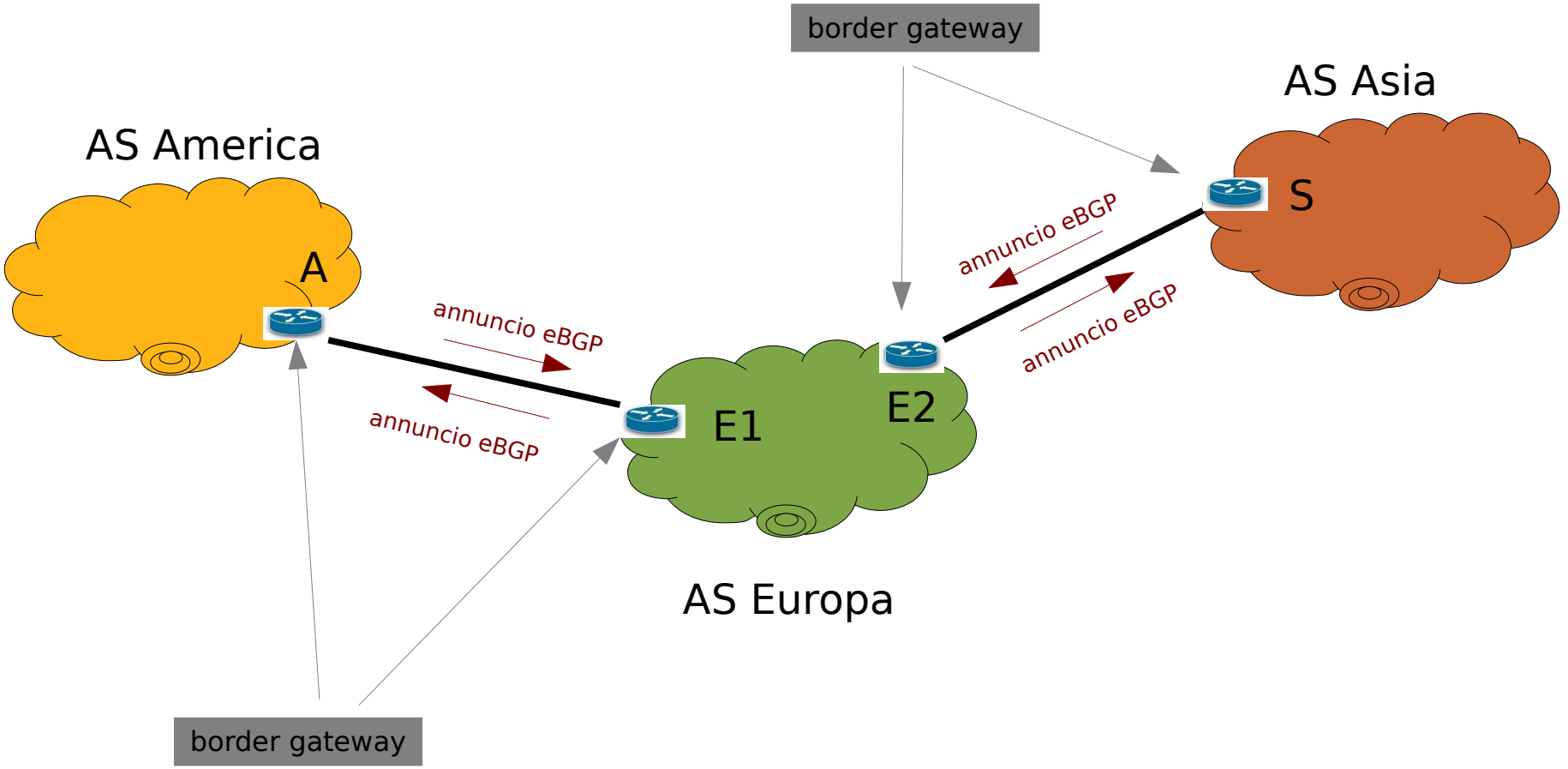
Gli IGP, tra cui si annoverano i principali OSPF (Open Shotest Path First) e IS-IS (Intermediate System to Intermediate System), rendono possibile ad un gruppo di router interconnessi la creazione di un Autonomous System (AS), creandosi le tabelle di intradamento interno secondo algoritmi variabili tra l'uno o l'altro protocollo IGP.

In seguito, con la enorme crescita delle reti in Internet, si decise di adottare le conosciute tecniche di interworking IGP anche tra reti indipendenti e separate e quindi tra AS differenti cosicché connettere automaticamente anche grossi AS tra loro. Tale necessità avveniva per risparmiare sui collegamenti in lunga distanza. Infatti, supponiamo dell'esistenza di tre grossi AS, uno in America, uno in Europa ed un ultimo in Asia. Originariamente supponiamo che l'AS di America voglia essere collegato con l'AS d'Europa cosicché i due AS possano “parlarsi”. Supponiamo, inoltre, che l'AS Europa avesse esigenza di collegarsi con l'AS dell'Asia.



I router che per ciascun AS si affacciano al mondo esterno e offrono la possibilità di interfacciamento col mondo esterno e, quindi, con altri AS, prendono nome di border-gateway, ovvero gateway di bordo o di confine, considerata la loro posizione sul bordo o confine del proprio AS.

Affinché più AS possano automaticamente conoscere i loro collegamenti esterni e costruire le routing tables per i propri border gateway, parimente ai protocolli IGP, sono stati implementati protocolli di interworking tra AS di livello più alto, chiamati EGP (External Gateway protocol). Tra i protocolli EGP più famosi c'è il BGP (Border Gateway Protocol) ed il RIP (Routing Interworking Protocol). Ad oggi, il protocollo EGP più utilizzato è il BGP, giunto alla sua versione 4.
Il BGP si basa su annunci (*announces*) inoltrati verso l'esterno del proprio AS e contenenti l'identificazione del proprio AS. Ad esempio se l'AS America deve essere connesso all'AS Europa è, innanzitutto, necessario che venga piazzato un router border gateway di confine che si affacci verso l'AS Europa e, similmente, ciò deve avvenire anche per l'AS Europa. Poi è indispensabile che il router border gateway di America e di Europa vengano configurati come router BGP, affinchè parlino la stessa lingua. Stessa cosa avverrà tra i router border-gateway tra l'AS Europa e l'AS Asia.
Sulle due tratte di collegamento dei tre AS, viaggeranno, quindi, messaggi EGP di tipo BGP e più precisamente di tipo eBGP (external Border Gateway Protocol).
La necessità di specificare con una “e” il protocollo BGP che già rientra nei protocolli di tipo esterno tra AS EGP verrà ripresa più avanti.
Ogni annuncio da parte di un router border-gateway verso l'altro router border-gateway presuppone l'instaurazione di una sessione di dialogo BGP, chiamata propriamente “peer-session” (sessione tra pari). Questa nomenclatura tiene a sottolineare che la comunicazione di un router BGP deve avvenire con un suo pari router di tipo BGP.



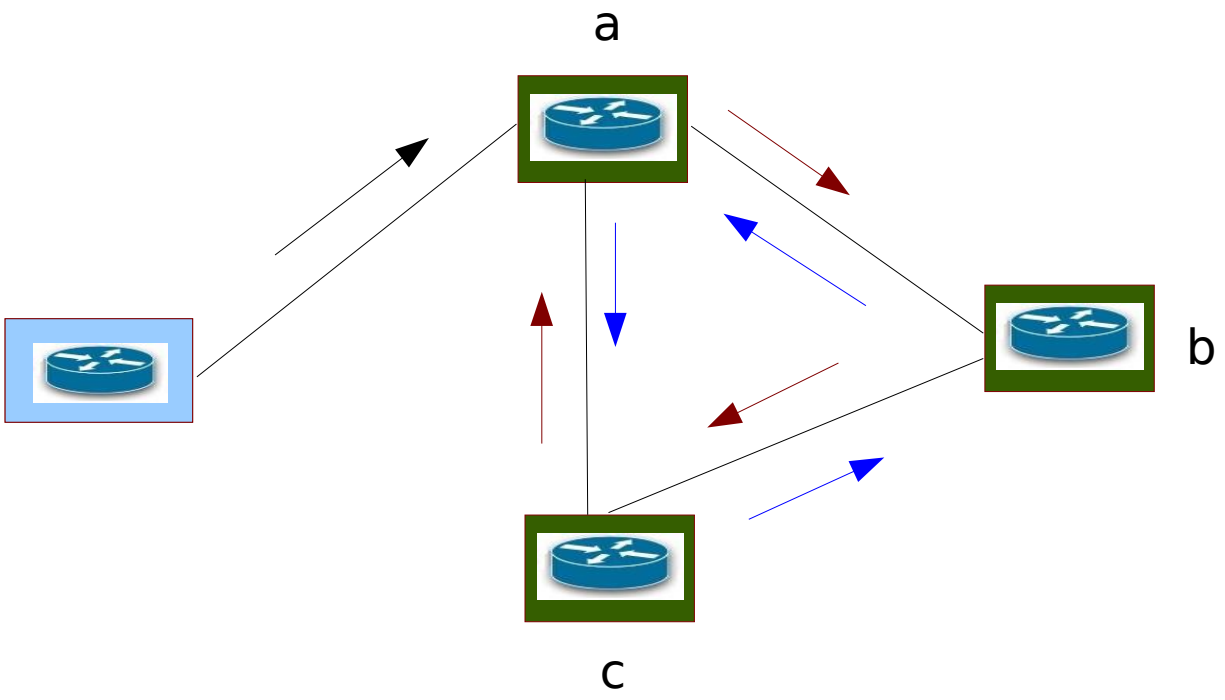
Gli annunci eBGP sono annunci di livello 4, cioè incapsulati in pacchetti TCP e contengono l'IP address della propria rete unita alla subnet mask. Ad esempio, se l'AS America ha indirizzi appartenenti alla classe IP 10.2.0.0/16 (ovvero rete 10.2.0.0 con subnet mask 255.255.0.0), il border gateway “A” dell'AS America genererà un BGP announce, comunicando al border gateway dell'AS Europa i suoi identificativi.

Questo annuncio consentirà al border-gateway “E1”, dell'AS Europa, di capire che esiste tramite esso, la possibilità di collegare Europa ad America, qualora gli hosts di Europa vogliano comunicare con America. Questa informazione, il border gateway “E1” se la memorizzerà nella sua tabella di routing. Nello specifico, se un host di Europa vuole “parlare” con un host di America e quindi il suo pacchetto IP ha come destination Address un indirizzo compreso nel range 10.2.0.0/16, l'AS Europa saprà che dovrà dirottare questo traffico verso il suo border-gateway “E1”, che, a sua volta, inoltrerà al border gateway “A” dell' AS America.
Ugualmente, anche il border gateway “E1” annuncirà la presenza dell'AS ad “A” e, quest'ultimo, si scriverà i dati dell'AS Europa nella sua tabella, rendendo reciproco l'interscambio di probabili comunicazioni tra hosts dei due AS. In conclusione, tra il router A ed il router E1 dovremo attendere due “peer-sessions” per ottenere la completa e reciproca interconnessione tra i due AS.
Lo stesso identico processo avverrà anche tra i border-gateway “E2” ed “S” per consentire la mutua connettività tra l'AS Europa e l'AS Asia.

Detto ciò, rimane da dipanare un problema per completare il discorso sulla funzionalità dell'interworking tra AS: come fanno i router interni di un AS a capire come instradare del traffico verso un host di un altro AS e, quindi, dirottare il traffico verso un determinato border-gateway? Ad esempio, volendo essere più precisi, potremo chiederci nel dettaglio, come possano, i router interni dell'AS Europa, capire che è necessario instradare traffico verso il border-gateway “E1” se un host del proprio AS Europa voglia dialogare con un host dell'AS America. Infatti, se un host dell'AS Europa vuole “dialogare” con l'host 10.2.134.25/24, appartenente ad una subnet 10.2.134.0/24 dell'AS America (avente classe di rete 10.2.0.0/16), come possono i router interni di Europa, capire di instradare questo traffico verso il router E1 e non verso il router E2?

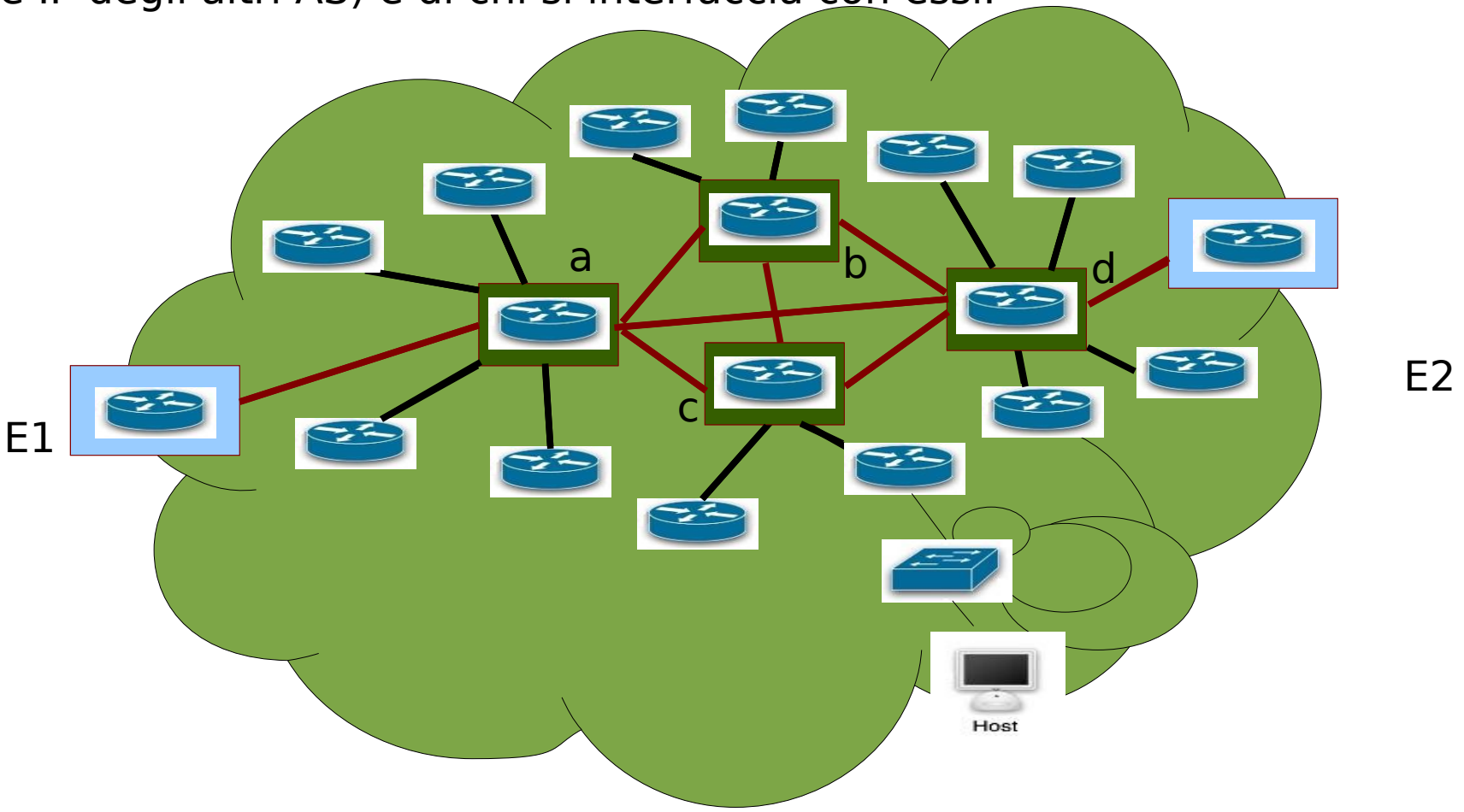
Per rispondere a questa domanda è necessario precisare che non sono solo i border gateway che possono “parlare” il BGP con i suoi pari, ma che, necessariamente, anche alcuni router interni ad un AS che disponga di border-gateway BGP, devono poter dialogare in BGP.
In particolare è stata creata un'estensione del BGP per comunicazione intena al proprio AS chiamata iBGP (internal Border Gateway Protocol) così che possa essere differenziata dalle comunicazioni BGP esterne eBGP verso i border-gateway di altri AS.
Vengono perciò scelti, all'interno di un AS, alcuni router che, oltre a svolgere le sue normali funzioni di “router dell'AS”, svolga anche funzioni di iBGP.
Ad esempio, prendiamo in esame l'AS Europa e supponiamo che utilizza internamente un IGP di tipo OSPF ed ha due border-gateway di tipo BGP chimati “E1” ed “E2” con i quali si assicura la connettività con i due AS adiacenti (neighbors).

Il protocollo OSPF sarà utilizzato da tutti i 16 router interni per costruirsi le proprie tabelle di routing, cioè per apprendere la propria rete e costruire, quindi l'AS.
Di questi sedici, quattro, oltre a svolgere la propria funzione all'interno dell'AS, sono stati scelti anche come iBGP router.
Il BGP usa annunci di tipo broadcast e, come in una rete LAN esiste il pericolo di routing-loop sui messaggi broadcast (risolto con lospanning-tree), così un problema simile si verifica anche tra i router iBGP. Se un router iBGP riceve un annuncio e lo propagasse sugli altri router iBGP a cui è connesso, questi ultimi lo potrebbero ri-propagare di nuovo all'origine, originandoil fenomeno di routing-loop information.



Nella figura sopra c'è un esempio di routing-loop information. Il border-gateway sulla sinistra propaga l'annuncio BGP al primo ed unico router iBGP a cui è collegato (freccia nera). Se il router iBGP “a” propagasse l'annuncio a “b” e “c”, “b” lo propgherà a “c”, il quale ha già ottenuto lo stesso messaggio da “a” e che, a sua volta, lo ri-propagherà ad “a”. E così anche “c” che aveva ricevuto il messaggio da “a” lo propagherà a “b” che lo ri-propagherà ad “a” e così senza fine, generando due routing-loop information che bloccheranno tutto il sistema. Proprio perchè il BGP non prevede un sistema di smorzamento immediato del routing-loop, i router iBGP non devono propagare gli annunci ricevuti. A sua volta, il fatto di non riflettere gli annunci ricevuti implica la necessità che tutti i router interni che parlino BGP siano totalmente magliati (fully-meshed connected), ovvero che ciascuno sia connesso con l'altro.

Quando arriva un'annuncio eBGP dall'esterno al border-gateway router “E1” o “E2”, questi lo inoltrano ai router iBGP che devono essere connessi ai border-gateway. Gli iBGP router, a sua volta, inoltrano il messaggio BGP sulle loro porte a cui sono connessi gli altri iBGP-router. In tal modo tutti i router deputati a “parlare” BGP all'interno dell'AS, avranno la conoscenza degli AS esterni (cioè dell'identificazione di classe IP degli altri AS) e di chi si interfaccia con essi.



border-gateway o eBGP router

iBGP router

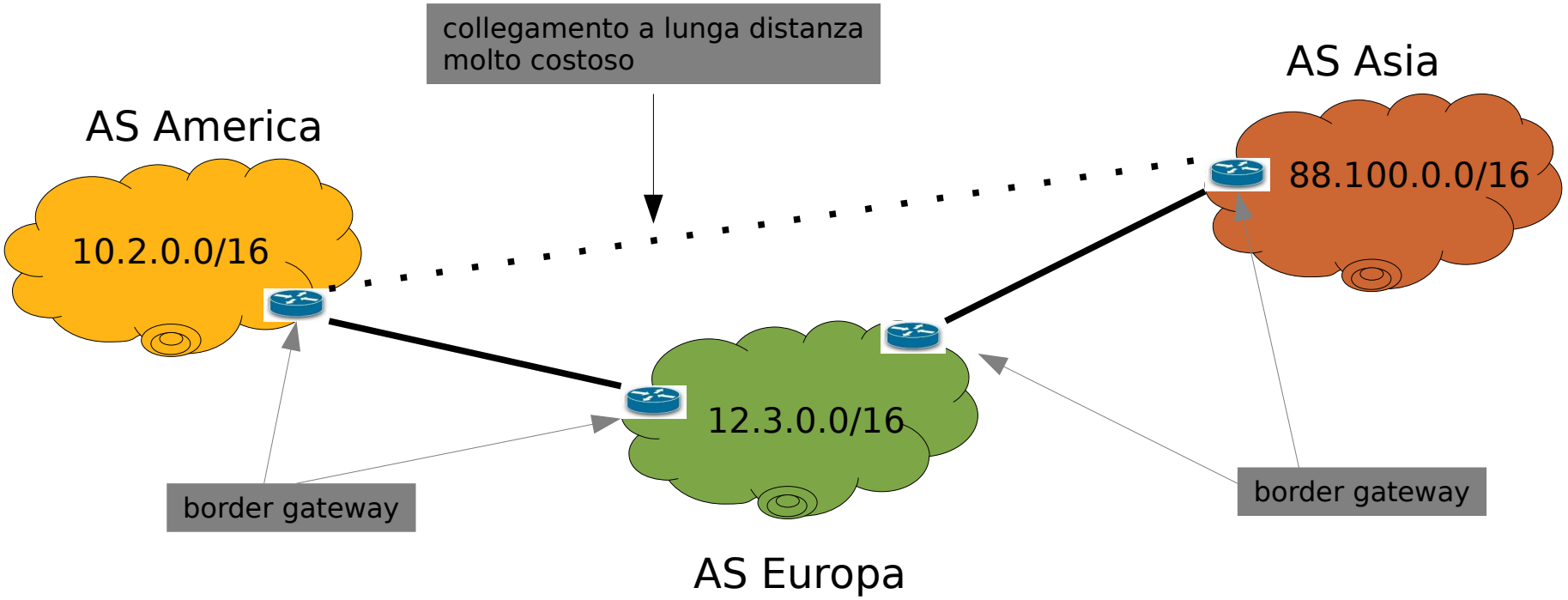
users router

Ad esempio se l'annuncio esterno eBGP, proveniente da AS America, è 10.2.0.0/16, questo verrà ricevuto da E1, il quale lo inoltrerà all'iBGP router “a” che lo rifletterà a tutti gli altri iBGP router,specificando, inoltre, che proviene da “E1”. Di conseguenza tutti gli iBGP router che sono connessi ad “a” (cioè “b”, “c” e “d”) ricevono l'annunc e si memorizzano nella loro tabella di routing l'associazione “10.2.0.0/16–ipaddress di “E1”.

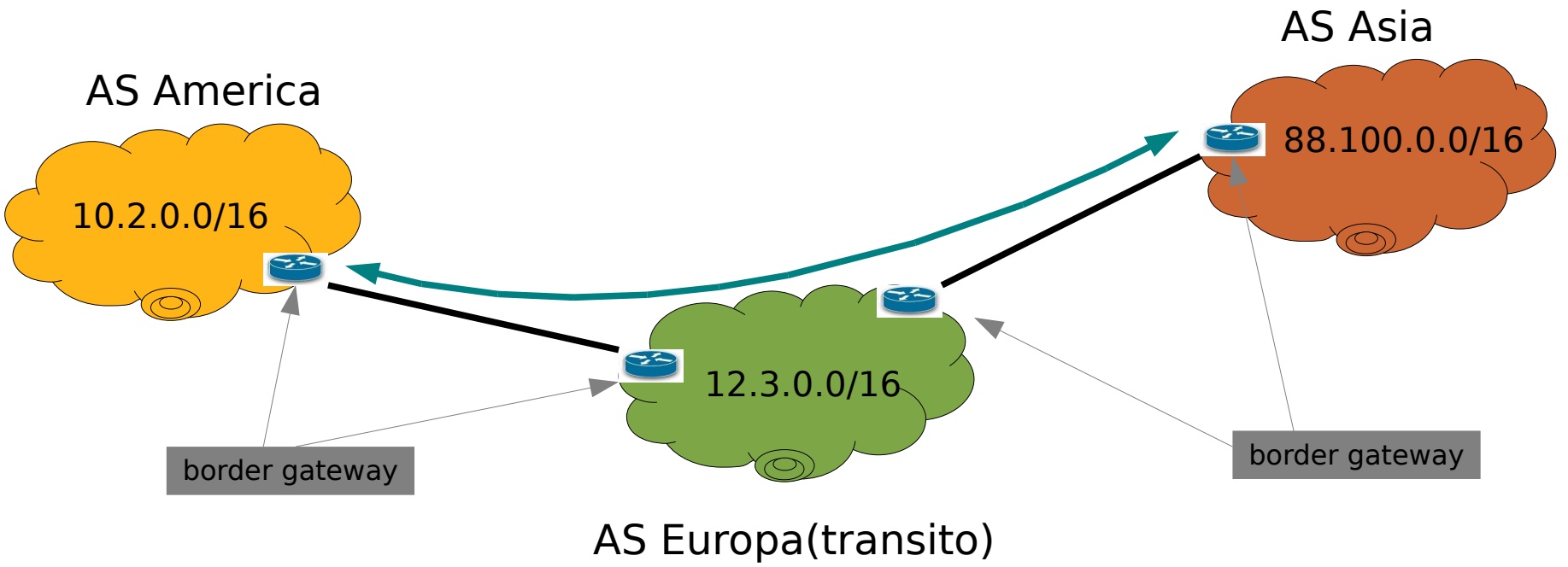
Quando l'host “Ciro” dell'AS Europa vuole raggiungere l'host dell'AS America avente ip address 10.2.134.25/24, l'iBGP router “c” avrà nella sua tabella la classe 10.2.0.0/16 appartenente al border gateway “E1” e quindi lo instraderà ad “a” che lo instraderà ad “E1”. Potrebbe instradare anche a “b” o a “d” (e di sicuro lo farà se il collegamento tra “c” ed “a” non dovesse risultare disponibile per qualsiasi motivo) ma questa funzione viene deputata al protocollo IGP utilizzato nell'AS.
Allo stesso modo se un utente dall'AS America volesse raggiungere l'host “Ciro” potrà farlo grazie agli iBGP router che intraderanno da “a” verso “c” e quest'ultimo verso il router d'utente che serve lo switch su cui è attestato “Ciro”. Dicasi similmente verso l'AS Asia interfacciato dal border gateway “E2”.

Questa struttura interna all'AS sì fatta, è detta “scalabile”. Per scalabile si intende un'architettura tale da poter prevedere la crescita o il ridimensionamento della stessa avendo minimo impatto col resto della rete già esistente. Difatti ad una rete simile è possibile aggiungere o eliminare router con effetti minimi di impatto sul resto della rete. E' importante prevedere la scalabilità, durante il design di una rete, altrimenti si corre il rischio di creare un qualcosa, senza fondamenta solide e che si dovrà abbatere per ricostruirla secondo i nuovi canoni di scalabilità.

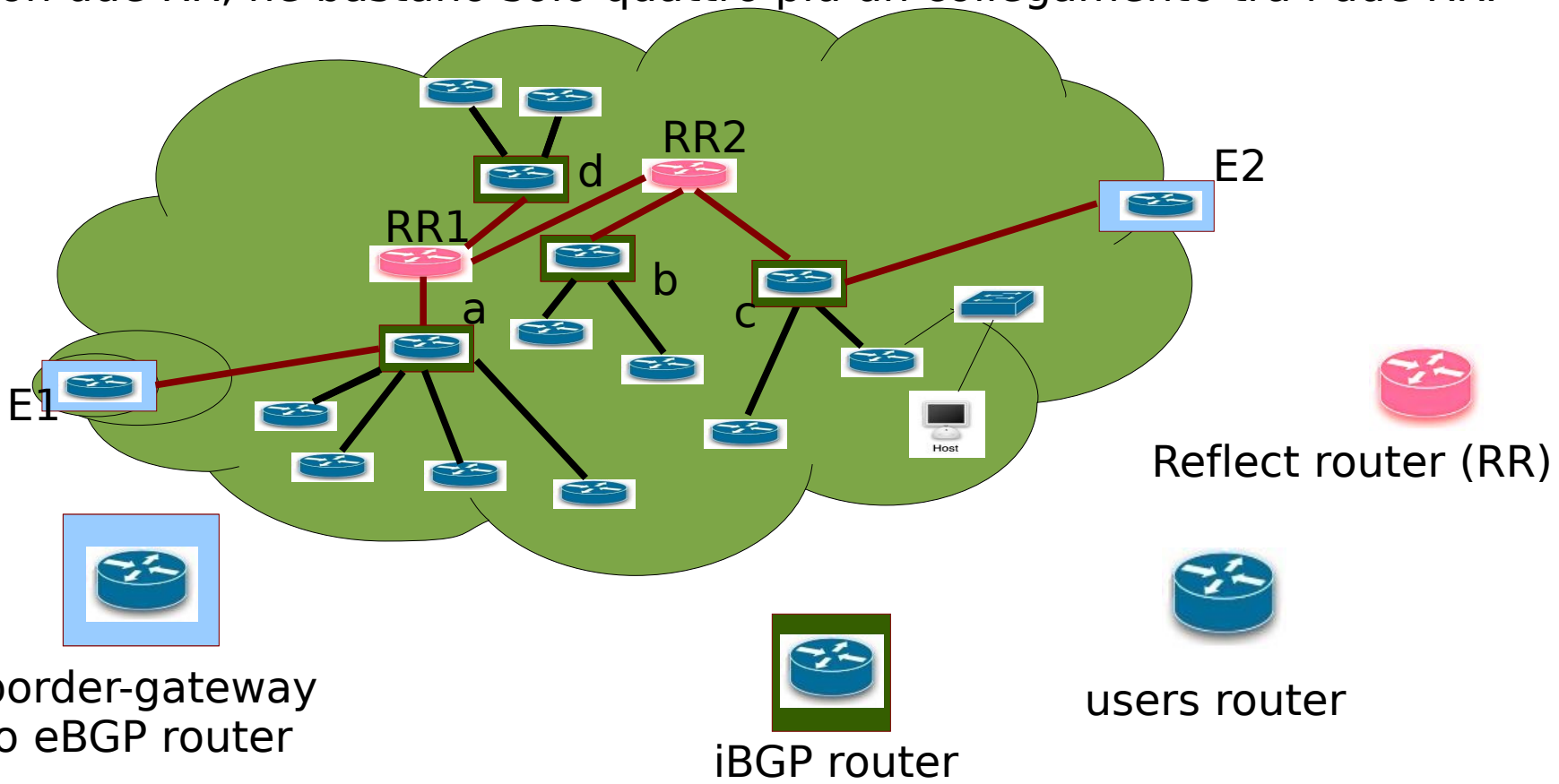
La presenza del protocollo iBGP interno ad un AS, consentirà anche la possibilità di transito di un AS verso altri AS. Infatti nel nostro esempio Europa ed America hanno connettività, così come anche Europa ed Asia. Supponiamo che, dopo un un po', anche America si imbatte nella necessità di dover parlare con Asia e quindi dovrebbe acquistare un costosissimo collegamento tra America ed Asia. I progettisti dell'AS America sono, però, al corrente che l' AS Europa ha già un collegamento con l'AS Asia e che il proprio AS America ha un collegamento con l'AS Europa. La soluzione meno costosa e più intelligente sarebbe di sfruttare l'AS Europa come tramite tra la connessione AS America <----> AS Asia. Naturalmente con un accordo commerciale America pagherà, su base traffico o con un forfait, il proprietario dell' AS Europa che si trasforma in una struttura di trasporto intermedia tra America ed Asia.



Grazie al protocollo BGP, trasformare Europa in un area di transito tra America ed Asia è possibile. Se un utente di America volesse “parlare” con un utente di Asia, genererà traffico verso un host appartenente alla rete 88.100.0.0/16. Con i router interni iBGP questo sarà possibile perché nelle loro tabelle di routing, durante l'annuncio di “E2” avranno appreso che 88.100.0.0/16 è un AS raggiungibile tramite “E2”. Il traffico proveniente dal border gateway “E1” e diretto verso Asia, sarà dirottato dai router iBGP di Europa in direzione di “E2”.



L'indiscussa capacità del BGP pone un solo limite: i router iBGP interni ad un AS devono essere totalmente magliati (fully-meshed) a causa della natura del protocollo che non prevede annunci indirizzati ad uno specifico router iBGP ma solo in un nativo “pseudo-formato broadcast”.
Nelle grandi reti dove è necessario disporre di molti iBGP router, magliarli tutti diviene un po' problematico e specialmente costoso. Per tale motivo è stato successivamente introdotto un metodo per evitare la magliatura completa degli iBGP ma assicurando, contemporaneamente, la raggiungibilità degli annunci a tutti gli iBGP di un AS: l'uso del Route-Reflector (RR = riflettore di rotte).
Il RR è un router iBGP presso cui sono collegati tutti i router iBGP del dominio o AS.
In tal modo, mentre prima era necessario stendere sei linee tra i quattro iBGP router, ora, con due RR, ne bastano solo quattro più un collegamento tra i due RR:



Il Route Reflector è così chiamato poiché la sua funzione principale è quella di “riflettere” gli annunci iBGP verso i router iBGP, evitando che questi ultimi debbano essere magliati completamente nell'AS di appartenneza. Gli stessi RR sono router iBGP e sono gli unici che stabiliscono una “peer-sessione” con i border-gateway.
I RR non hanno funzionalità all'interno dell'AS fine a se stessa, ovvero non partecipano all'instradamento interno all'AS e non sono quindi soggetti a protocolli di tipo IGP. Tutti i RR devono essere magliati tra di loro e ad essi saranno collegati a “Star” tutti gli iBGP dell'AS.
Con i RR si risparmia in collegamenti tra gli iBGP router e si guadagna, inoltre, una maggiore scalabilità dell'AS.