Native Multicast and Inter Domain Routing on TEN-155

JN-99-04-13

1. **Some FAQs**

   What is and why to use MBGP ?

   MBGP stands for Multiprotocol Extensions for BGP-4. The only difference from ordinary BGP-4 is different NLRI code in BGP messages, allowed by RFC2283. Routers, which communicate using this RFC can exchange routing information of several protocols - similar mechanism to IS-IS protocol, when configured to carry IP routing information. This functionality of BGP can be used to exchange reach ability information of multicast sources. In the current DVMRP MBONE tunnels are used to bypass non-multicast capable routers in the Internet. MBGP creates separate routing table, consisting of entries with NLRI=multicast, which can be quite different from routing table for unicast traffic. Separate routing tables enable to run multicast and unicast routing on one physical router and still have the possibility of different unicast and multicast topologies - some parts of the network can be used by unicast only, some by multicast only. Closer overview can be found at ftp://ftpeng.cisco.com/ipmulticast.html

   How to use the second routing table ?

   Older multicast routing protocols like DVMRP (multicast variant of RIP), MOSPF (multicast variant of OSPF) build separate routing table to look into for RPF (Reverse Path Forwarding) checks for data coming from different sources. RPF mechanism is the basic one to avoid looping of multicast packets. These two protocols have the same limitations as their unicast counterparts and can't be used for worldwide infrastructure. Currently more and more used PIM (Protocol Independent Multicast) in two variants PIM-SM (sparse mode) and PIM-DM (dense mode) does not build it's own routing table for RPF lookups. It uses whatever is available as routing table in the local machine - this "property" simplifies the protocol a lot, but limits it again to one administrative domain. The idea of MBGP usage for multicast routing is to build routing table for PIM using well-known and very powerful possibilities of BGP, used currently for the worldwide Internet unicast routing. The key point to understand - the router, when sending unicast data never looks into table created with NLRI=multicast. On the other hand, when doing RPF check for multicast packets, he can use both tables - if multicast and unicast topology is identical, we don't need second table at all. If this is not the case, we create second table and point the RPF look-ups to the appropriate routing entries - in cisco implementation using administrative distance for the MBGP routing entries.

   Why to use PIM-SM in the backbone ?

   The old MBONE uses currently DM operation, where data are flooded everywhere and each end user has to refuse explicitly to accept them (prune data). If the prune message is for any reason lost or not sent, whole network suffers from continuos, un-needed data flow. In the case of TEN-155 it means, that backbone is full of data and nobody can use multicast for some reasonable purposes. Dense mode operation was used for the easiness, with which all multicast data reach end users. In SM operation mode each users has to explicitly request needed data, e.g. if these messages are lost, the rest of the network is not influenced. But special measures have to be taken to assure the delivery of multicast data to all users.

How to interconnect SM and DM domains ?

When interconnecting sparse and dense mode domains we have the problem of "dense mode receiver" . No data are forwarded from SM to DM domain, routers in DM domain do not have any forwarding states for any groups and interface. Even if the receiver knows the group, he wants to join, the IGMP join message is discarded in nearest router as it does not have any state for the group. For cisco based dense mode domains there is workaround using "ip igmp helper-address" command, for mrouted domains there is only the possibility to start to send data to the group - data are flooded over the DM domain including border router, forwarding states are created and SM domain learns about receivers, willing that particular data. This problem in practice applies for SDR announcements only, all the widely used MBONE tools run RTCP, which always sends some control messages to the multicast group address e.g. you are always a sender, whenever you join any group using thess tools. As can be seen, the problem of DM non-pruners is completely and democratically enough solved.

How to interconnect several SM domains ?

PIM-SM supposes the existence of one administrative domain, with one point (RP Rendez-vous Point) where receivers will request data, which they want to see. This is not the case of real Internet, where everybody wants to control his own domain. Therefore, new protocol was proposed - MSDP, which enables to communicate the multicast sources among several RPs from different domains. MSDP runs over TCP similarly to eBGP - there is no need for full mesh of TCP connections, which wouldn't scale again.

Do we need MBGP on TEN-155 ?

In principle, unicast and multicast topology of TEN-155 backbone will be always the same - we will have the same paths between NRNs for both kind of traffic. This applies even for European commercial peers. The problem arrives when trying to connect US multicast sources - TEN-155 does not have full Internet routing table, e.g. we need to import these routes for multicast RPF checks from TEN-155 multicast connection to the US and we do not want this routes to be known for unicast traffic. As mentioned above, this is the case, where MBGP comes into play - we will create another routing table for US multicast routes, which will never be used for any unicast traffic for sure.

Furthermore, the application of MBGP over whole TEN-155 and NRNs networks will enable easy multi homing of all parties as easily as it is done now with unicast BGP peerings.

2. **Migrating TEN-155 MBONE**

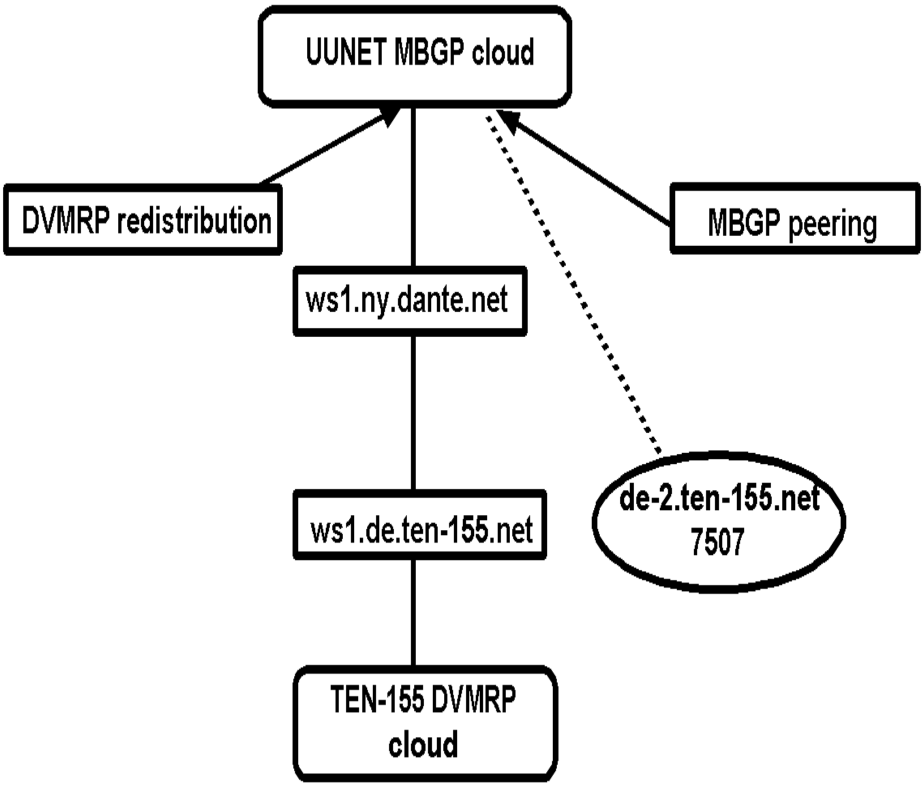
Step -1

Currently TEN-155 has multicast connectivity to the outside world (e.g. non TEN-155 world) through DANTE transatlantic line to UUNET. UUNET runs MBGP/PIM cloud, accepts specified set of our DVMRP routes and send us a default route over an ordinary DVMRP tunnel to mrouted workstation in NY PoP.

In step 1 we want to install separate router (DE-2) in DE PoP and connect it to TEN-155 through ATM and Ethernet interfaces. The outside world will be connected either via GRE tunnel running MBGP/PIM to UUNET or the possibility of straight ATM VC (e.g. native multicast configuration) from UUNET to this router will be investigated.

Reason for this step: verify the stability of multicast code on TEN-155 HW set-up, verify functionality of Netflow and Interdomain routing on one SW/HW platform, test interconnection of different multicast domains (PIM-SM, PIM-DM, DVMRP), measure the processor load caused by multicast itself (later on, when "production" multicast data will be forwarded through this box).

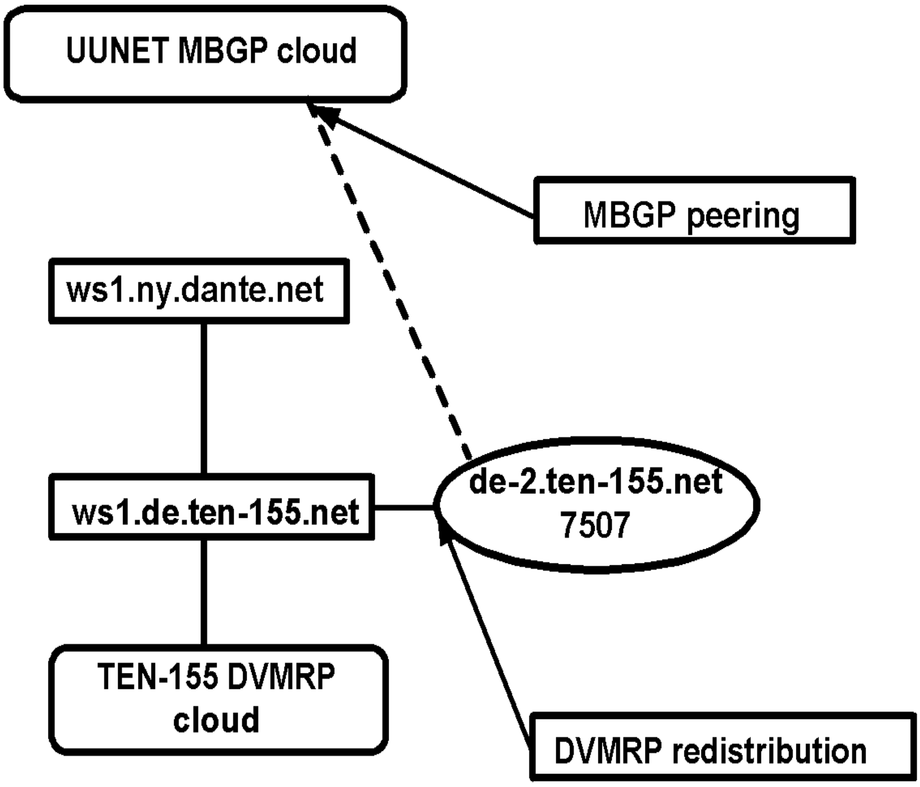
Realization: router was already installed, we run MBGP with DFN, we have a cisco case open concerning PIM-SM/MSDP



Step-2

We connect DE-2 to ws1.de.ten-155.net via DVMRP tunnel and do, what UUNET does for us now - accepting specified routes from TEN-155 (suggestion would be not to do DVMRP to MBGP redistribution but doing aggregation using "summary-only" BGP network command, initiating summary generation by subnet available in the DVMRP routing table - in my view, people generating prefixes longer than /24 into BGP should be prosecuted) and injecting default to DVMRP cloud. The DVMRP tunnel from ws1.ny. dante.net to UUNET will be cut-off simultaneously. All these steps require discipline from all connected NRNs - there is danger of causing routing loops if somebody announces TEN-155 DVMRP routes to the other world via another connection - this does not exclude multihoming of NRNs but requires some care.

Realization: done 6.5.99, except of UUNET connection - when trying to order MBGP, we discovered UUNET can`t provide it. DFN migrated to MBGP



Step-3

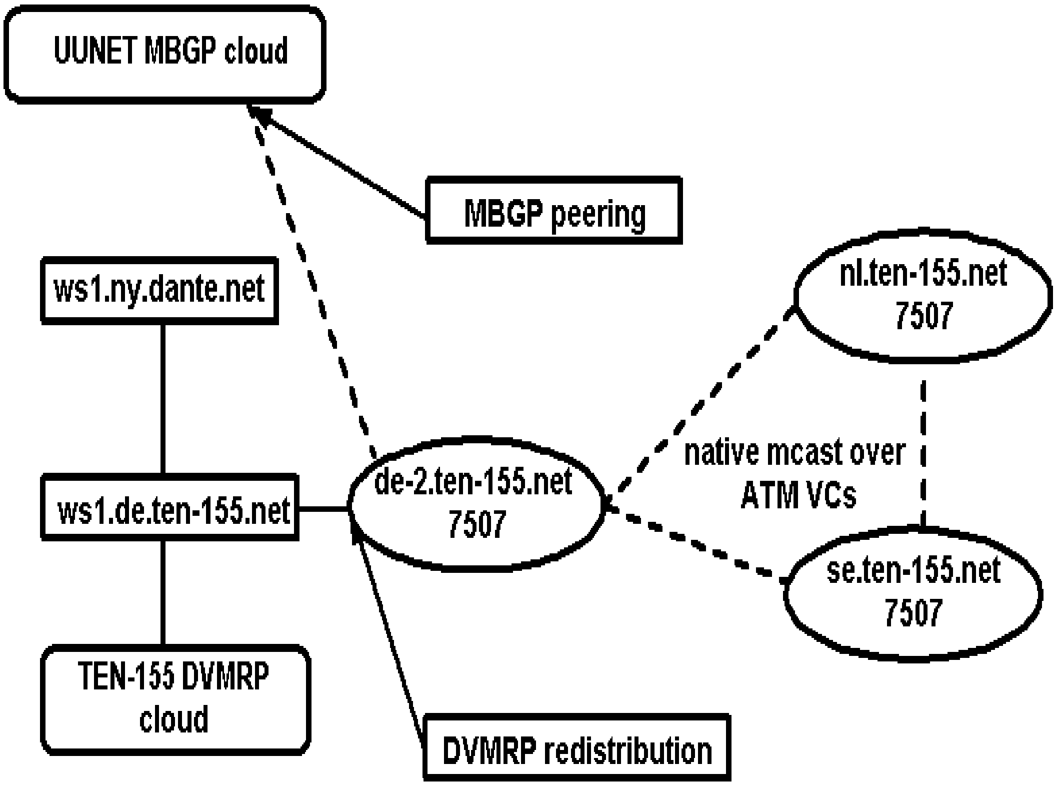
We connect separate ATM VCs from DE-2 to SE.ten-155.net and NL.ten-155.net, enabling MBGP and PIM-SM between (hopefully IP NOC will not jump up now, as they were roughly warned beforehand several times ...) all these three routers and starting to have native multicast PIM-SM backbone. Nordunet, Surfnet, Grnet can be connected, all of them using MBGP/PIM-SM and most probably natively, again cutting off their DVMRP tunnels to TEN-155 workstations simultaneously. Redistribution of their routes to DVMRP cloud will be done in DE-2 - at this stage we will see the load caused by all off this. We could/should set-up the unicast peering with de.ten-155.net at this stage or even earlier to see load and learning to point PIM to the right routing table.

Reason for separate ATM VCs: Possibility to run native backbone. Check if switching of multicast traffic between several ATM point-to-point interfaces works properly. This is not meant as building separate infrastructure for multicast, the final goal still is to have it fully on production unicast network - DE.ten-155.net is the most loaded and "strategic" router to risk troubles here.

Expected realization: Multicast on SE.TEN-155.NET 18.5.1999, including connection to Nordunet

Multicast on NL.TEN-155.NET 25.5.1999. Due to simultaneous work in AT connection to Surfnet will be configured later on

More detailed config



Step-4

The worst we have behind us (hopefully), so we can go ahead and broaden the production routers utilization to the rest of TEN-155 routers and trying to decrease the DVMRP domain as much as possible.