Network Working Group
Internet Draft
Intended status: Informational
Expires: December 12, 2012

Maria Napierala
AT&T
Luyuan Fang
Dennis Cai
Cisco Systems

June 12, 2012

IP-VPN Data Center Problem Statement and Requirements
draft-fang-vpn4dc-problem-statement-01.txt

Abstract

Network Service Providers commonly use BGP/MPLS VPNs [RFC 4364] as the control plane for virtual networks. This technology has proven to scale to a large number of VPNs and attachment points, and it is well suited for Data Center connectivity, especially when supporting all IP applications.

The Data Center environment presents new challenges and imposes additional requirements to IP VPN technologies, including multi-tenancy support, high scalability, VM mobility, security, and orchestration. This document describes the problems and defines the new requirements.

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress".

This Internet-Draft will expire on December 12, 2012.

Table of Contents

1. Introduction 3
2. Terminology 4
3. IP-VPN in Data Center Network 4
3.1. Data Center Connectivity Scenarios 5
4. Data Center Virtualization Requirements 6
5. Decoupling of Virtualized Networking from Physical Infrastructure 6
6. Encapsulation/Decapsulation Device for Virtual Network Payloads 7
7. Decoupling of Layer 3 Virtualization from Layer 2 Topology 8
8. Requirements for Optimal Forwarding of Data Center Traffic 9
9. Virtual Network Provisioning Requirements 9
10. Application of BGP/MPLS VPN Technology to Data Center Network 10
10.1. Data Center Transport Network 12
10.2. BGP Requirements in a Data Center Environment 12
11. Virtual Machine Migration Requirement 14
12. IP-VPN Data Center Use Case: Virtualization of Mobile Network 15
13. Security Considerations 17
14. IANA Considerations 17
15. Normative References 17
16. Informative References 17
17. Authors' Addresses 17
18. Acknowledgements 18

Requirements Language

Although this document is not a protocol specification, the key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC 2119].

1. Introduction

Data Centers are increasingly being consolidated and outsourced in an effort, both to improve the deployment time of applications as well as reduce operational costs. This coincides with an increasing demand for compute, storage, and network resources from applications. In order to scale compute, storage, and network resources, physical resources are being abstracted from their logical representation. This is referred as server, storage, and network virtualization. Virtualization can be implemented in various layers of computer systems or networks.

The compute loads of many different customers are executed over a common infrastructure. Compute nodes are often executed as Virtual Machines in an "Infrastructure as a Service" (IaaS) Data Center. The set of virtual machines corresponding to a particular customer should be constrained to a private network.

New network requirements are presented due to the consolidation and virtualization of Data Center resources, public, private, or hybrid. Large scale server virtualization (i.e., IaaS) requires scalable and robust Layer 3 network support. It also requires scalable local and global load balancing. This creates several new problems for network connectivity, namely elasticity, location independence (referred to also as Virtual Machine mobility), and extremely large number of virtual resources.

In the Data Center networks, the VMs of a specific customer or application are often configured to belong to the same IP subnet. Many solutions proposed for large Data Center networks rely on the assumption that the layer-2 inter-server connectivity is required, especially to support VM mobility within a virtual IP subnet. Given that VM mobility consists in moving VMs anywhere within (and even across) Data Centers, the virtual subnet locality associated with small scale deployments cannot be preserved. A Data Center solution should not prevent grouping of virtual resources into IP subnets but the virtual subnets have no benefits of locality across a large data-center.

While some applications may expect to find other peers in a particular user defined IP subnet, this does not imply the need to provide a Layer 2 service that preserves MAC addresses. A network virtualization solution should be able to provide IP unicast connectivity between hosts in the same and different subnets without any assumptions regarding the underlying media layer. A solution should also be able to provide a multicast service that implements IP subnet broadcast as well as IP multicast.

One of the main goals in designing a Data Center network is to minimize the cost and complexity of its core/"fabric" network. The cost and complexity of Data Center network is a function of the number of virtualized resources, that is, the number of "closed user-groups". Data Centers use VPNs to isolate compute resources associated with a specific "closed user-group". Some use VLANs as a VPN technology, others use Layer 3 based solutions often with proprietary control planes. Service Providers are interested in interoperability and in openly documented protocols rather than in proprietary solutions.

2. Terminology

AS	Autonomous Systems
DC	Data Center
DCI	Data Center Interconnect
EPC	Evolved Packet Core
End-System	A device where Guest OS and Host OS/Hypervisor reside
IaaS	Infrastructure as a Service
LTE	Long Term Evolution
PCEF	Policy Charging and Enforcement Function
RT	Route Target
ToR	Top-of-Rack switch
VM	Virtual Machine
Hypervisor	Virtual Machine Manager
SDN	Software Defined Network
VPN	Virtual Private Network

3. IP-VPN in Data Center Network

In this document, we define the problem statement and requirements for Data Center connectivity based on the assumption that applications require IP connectivity but no Layer 2 direct adjacencies. Applications do not send or receive Ethernet frames directly. They are restricted to IP services due to several reasons such as privileges, address discovery, portability, APIs, etc. IP service can be unicast, VPN broadcast, or multicast.

An IP-VPN DC solution is meant to address IP-only Data Center, defined by a Data Center where VMs, applications, and appliances require only IP connectivity and the underlying DC core infrastructure is IP only. Non-IP applications are addressed by other solutions and are not in scope of this document.

It is also assumed that both IPv4 and IPv6 unicast communication is to be supported. Furthermore, the multicast transmission, i.e., allowing IP applications to send packets to a group of IP addresses should also be supported. The most typical multicast applications

are service, network, device discovery applications and content

distribution. While there are simpler and more effective ways to provide discovery services or reliable content delivery, a Data Center solution should support multicast transmission to applications. A Data Center solution should cover the case where the Data Center transport network does not support IP multicast transmission service.

The Data Center multicast service should also support a delivery of traffic to all endpoints of a given VPN even if those endpoints have not sent any control messages indicating the need to receive that traffic. In other words, the multicast service should be capable of delivering the IP broadcast traffic in a virtual topology.

3.1. Data Center Connectivity Scenarios

There are three different cases of Data Center (DC) network connectivity:

1. Intra-DC connectivity: Private network connectivity between compute resources within a public (or private) Data Center.

2. Inter-DC connectivity: Private network connectivity between different Data Centers, either public or private.

3. Client-to-DC connectivity: Connectivity between client and a private or public Data Center. The later includes interconnection between a service provider and a public Data Center (which may belong to the same or different service provider).

Private network connectivity within the Data Center requires network virtualization solution. In this document we define Layer 3 VPN requirements to Data Center network virtualization. The Layer 3 VPN technology (i.e., MPLS/BGP VPN) also applies to the interconnection of different data-centers.

When private networks interconnect with public Data Centers, the VPN provider must interconnect with the public Data Center provider. In this case we are in the presence of inter-provider VPNs. The Inter-AS MPLS/BGP VPN Options A, B, or C [RFC 4364] provide network-to-network interconnection service and they constitute the basis of SP network to public Data Center network connectivity. There might incremental improvements to the existing inter-AS solutions, pertaining to scalability and security, for example.

Service Providers can leverage their existing Layer 3 VPN services and provide private VPN access from client's branch sites to client's own private Data Center or to SP's own Data Center. The service provider-based VPN access can provide additional value compared with public internet access, such as security, QoS, OAM, and troubleshooting.

4. Data Center Virtualization Requirements

Private network connection service in a Data Center must provide traffic isolation between different virtual instances that share a common physical infrastructure. A collection of compute resources dedicated to a process or application is referred to as a "closed user-group". Each "closed user-group" is a VPN in the terminology used by IP VPNs.

Any DC solution needs to assure network isolation among tenants or applications sharing the same Data Center physical resources. A DC solution should allow a VM or application end-point to belong to multiple closed user-groups/VPNs. A closed user-group should be able to communicate with other closed-user groups according to specified routing policies. A customer or tenant should be able to define multiple closed user-groups.

Typically VPNs that belong to different tenants do not communicate with each other directly but they should be allowed to access common appliances such as storage, database services, security services, etc. It is also common for tenants to deploy a VPN per "application tier" (e.g. a VPN for web front-ends and a different VPN for the logic tier). In that scenario most of the traffic crosses VPN boundaries. That is also the case when "network attached storage" (NAS) is used or when databases are deployed as-a-service.

Another reason for the Data Center network virtualization is the need to support VM move. Since the IP addresses used for communication within or between applications may be anywhere across the data-center, using a virtual topology is an effective way to solve this problem.

5. Decoupling of Virtualized Networking from Physical Infrastructure

The Data Center switching infrastructure (access, aggregation, and core switches) should not maintain any information that pertains to the virtual networks. Decoupling of virtualized networking from the physical infrastructure has the following advantages: 1) provides

better scalability; 2) simplifies the design and operation; 3) reduces the cost of a Data Center network. It has been proven (in Internet and in large BGP IP VPN deployments) that moving complexity associated with virtual entities to network edge while keeping network core simple has very good scaling properties.

There should be a total separation between the virtualized segments (virtual network interfaces that are associated with VMs) and the physical network (i.e., physical interfaces that are associated with the data-center switching infrastructure). This separation should include the separation of the virtual network IP address space from the physical network IP address space. The physical infrastructure addresses should be routable in the underlying Data Center transport network, while the virtual network addresses should be routable on the VPN network only. Not only should the virtual network data plane be fully decoupled from the physical network, but its control plane should be decoupled as well.
In order to decouple virtual and physical networks, the virtual networking should be treated as an "infrastructure" application. Only the solutions that meet those requirements would provide a truly scalable virtual networking.

MPLS labels provide the necessary information to implement VPNs. When crossing the Data Center infrastructure the virtual network payloads should be encapsulated in IP or GRE [RFC 4023], or native MPLS envelopes.

6. Encapsulation/Decapsulation Device for Virtual Network Payloads

In order to scale a virtualized Data Center infrastructure, the encapsulation (and decapsulation) of virtual network payloads should be implemented on a device as close to virtualized resources as possible. Since the hypervisors in the end-systems are the devices at the edge of a Data Center network they are the most optimal location for the VPN encap/decap functionality.
Data-plane device that implements the VPN encap/decap functionality acts as the first-hop router in the virtual topology.

The IP-VPN solution for Data Center should also support deployments where it is not possible or not desirable to implement VPN encapsulation in the hypervisor/Host OS. In such deployments encap/decap functionality may be implemented in an external physical switch such as aggregation switch or top-of-rack switch. The external device implementing VPN tunneling functionality should be a close as possible to the end-system itself. The same DC solution should support deployments with both, internal (in a hypervisor) and external (outside of a hypervisor) encap/decap

devices.

Whenever the VPN forwarding functionality (i.e., the data-plane device that encapsulates packets into, e.g., MPLS-over-GRE header) is implemented in an external device, the VPN service itself must be delivered to the virtual interfaces visible to the guest OS. However, the switching elements connecting the end-system to the encap/decap device should not be aware of the virtual topology. Instead, the VPN endpoint membership information might be, for example, communicated by the end-system using a signaling protocol. Furthermore, for an all-IP solution, the Layer 2 switching elements connecting the end-system to the encap/decap device should have no knowledge of the VM/application endpoints. In particular, the MAC addresses known to the guest OS should not appear on the wire.

7. Decoupling of Layer 3 Virtualization from Layer 2 Topology

The IP-VPN approach to Data Center network design dictates that the virtualized communication should be routed, not bridged. The Layer 3 virtualization solution should be decoupled from the Layer 2 topology. Thus, there should be no dependency on VLANs or Layer 2 broadcast.

In solutions that depend on Layer 2 broadcast domains, the VM-to-VM communication is established based on flooding and data plane MAC learning. Layer 2 MAC information has to be maintained on every switch where a given VLAN is present. Even if some solutions are able to eliminate data plane MAC learning and/or unicast flooding across Data Center core network, they still rely on VM MAC learning at the network edge and on maintaining the VM MAC addresses on every (edge) switch where the Layer 2 VPN is present.

The MAC addresses known to guest OS in end-system are not relevant to IP services and introduce unnecessary overhead. Hence, the MAC addresses associated with virtual machines should not be used in the virtual Layer 3 networks. Rather, only what is significant to IP communication, namely the IP addresses of the VMs and application endpoints should be maintained by the virtual networks. An IP-VPN solution should forwards VM traffic based on their IP addresses and not on their MAC addresses.

From a Layer 3 virtual network perspective, IP packets should reach the first-hop router in one-hop, regardless of whether the first-hop router is a hypervisor/Host OS or it is an external device. The VPN first-hop router should always perform an IP lookup on every packet it receives from a VM or an application. The first-hop router should encapsulate the packets and route them towards the destination end-system. Every IP packet should be forwarded along

the shortest path towards a destination host or appliance,

regardless of whether the packet's source and destination are in the same or different subnets.

8. Requirements for Optimal Forwarding of Data Center Traffic

The Data Center solutions that optimize for the maximum utilization of compute and storage resources require that those resources may be located anywhere in the data-center. The physical and logical spreading of appliances and computations implies a very significant increase in data-center infrastructure bandwidth consumption. Hence, it is important that DC solutions are efficient in terms of traffic forwarding and assure that packets traverse Data Center switching infrastructure only once. This is not possible in DC solutions where a virtual network boundary between bridging (Layer 2) and routing (Layer 3) exists anywhere within the Data Center transport network. If a VM can be placed in an arbitrary location, mixing of the Layer 2 and the Layer 3 solutions may cause the VM traffic traverse the Data Center core multiple times before reaching the destination host.

It must be also possible to send the traffic directly from one VM to another VM (within or between subnets) without traversing through a midpoint router. This is important given that most of the traffic in a Data Center is within the VPNs.

9. Virtual Network Provisioning Requirements

IP-VPN DC has to provide fast and secure provisioning (with low operational complexity) of VPN connectivity for a VM within a Data Center and across Data Centers. This includes interconnecting VMs within and across physical Data Centers in the context of a virtual networking. It also includes the ability to connect a VM to a customer VPN outside the Data Center, thus requiring the ability to provision the communication path within the Data Center to the customer VPN.

The VM provisioning should be performed by an orchestration system. The orchestration system should have a notion of a closer user-group/tenant and the information about the services the tenant is allowed to access. The orchestration system should allocate an IP address to a VM. When the VM is provisioned, its IP address and the closed user-group/VPN identifier (VPN-ID) should be communicated to the host OS on the end-system. There should a centralized database system (possibly with a distributed implementation) that will contain the provisioning information regarding VPN-IDs and the services the corresponding VPNs could

access. This information should be accessible to the virtual network control plane.

The orchestration system should be able to support the specification of fine grain forwarding policies (such as filtering, redirection, rate limiting) to be injected as the traffic flow rules into the virtual network.

Common APIs can be a simple and a useful step to facilitate the provisioning processes. Authentication is required when a VM is being provisioned to join an IP VPN.

An IP-VPN Data Center networking solution should seamlessly support VM connectivity to other network devices (such as service appliances or routers) that use the traditional BGP/MPLS VPN technology.

10. Application of BGP/MPLS VPN Technology to Data Center Network

BGP IP VPN technologies (based on [RFC 4364]) have proven to be able to scale to a large number of VPNs (tens of thousands) and customer routes (millions) while providing for aggregated management capability. Data Center networks could use the same transport mechanisms as used today in many Service Provider networks, specifically the MPLS/BGP VPNs that often overlay huge transport areas.

MPLS/BGP VPNs use BGP as a signaling protocol to exchange VPN routes. IP-VPN DC solution should consider that it might not be feasible to run BGP protocol on a hypervisor or external switch such as top-of-rack. This includes functions like BGP route selection and processing of routing policies, as well as handling MP-BGP structures like Route Distinguishers and Route Targets. Rather, it might be preferable to use a signaling mechanism that is more familiar and compatible with the methods used in the application software development. While network devices (such as routers and appliances) may choose to receive VPN signaling information directly via BGP, the end-systems/switches may choose other type of interface or protocol to exchange virtual end-point information. The IP VPN solution for Data Center should specify the mapping between the signaling messages used by the hypervisors/switches and the MP-BGP routes used by MP-BGP speakers participating in the virtual network.

In traditional WAN deployments of BGP IP VPNs [RFC 4364], the forwarding function and control function of a Provider Edge (PE) device have co-existed within a single physical router. In a Data

Center network, the PE plays a role of the first-hop router, in a

virtual domain. The signaling exchanged between forwarding and control planes in a PE has been proprietary to a specific PE router/vendor. When BGP IP VPNs are applied to a Data Center network, the signaling used between the control plane and forwarding should be open to provisioning and standardization. We explore this requirement in more detail below.

When MPLS/BGP VPNs [RFC 4364] are used to connect VMs or application endpoints, it might be desirable for a hypervisor's host or an external switch (such as TOR) to support only the forwarding aspect of a Provider Edge (PE) function. The VMs or applications would act as Customer Edges (CEs) and the virtual networks interfaces associated with the VMs/applications as CE interfaces. More specifically, a hypervisor/first-hop switch would support only the creation and population of VRF tables that store the forwarding information to the VMs and applications. The forwarding information should include 20-bit label associated with a virtual interface (i.e., a specific VM/application endpoint) and assigned by the destination PE. This label has only a local significance within a destination PE. A hypervisor/first-hop switch would not need to support BGP, a protocol familiar to network devices.

When a PE forwarding function is implemented on an external switch, such as aggregation or top-of-rack switch, the end-system must be able to communicate the endpoint and its VPN membership information to the external switch. It should be able to convey the endpoint's instantiation as well as removal events.

An IP-VPN Data Center networking solution should be able to support a mixture of internal PEs (implemented in hypervisors/Host OS) and external PEs (implemented on external to the end-system devices).

The IP-VPN DC solution should allow BGP/MPLS VPN-capable network devices, such as routers or appliances, to participate directly in a virtual network with the Virtual Machines and applications. Those network devices can participate in isolated collections of VMs, i.e., in isolated VPNs, as well as in overlapping VPNs (called "extranets" in BGP/MPLS VPN terminology).

The device performing PE forwarding function should be capable of supporting multiple Virtual Routing and Forwarding (VRF) tables representing distinct "close user groups". It should also be able to associate a virtual interface (corresponding to a VM or application endpoint) with a specific VRF.

The first-hop router has to be capable of encapsulating outgoing traffic (end-system towards Data Center network) in IP/GRE or MPLS envelopes, including the per-prefix 20-bit VPN label. The first-hop

router has to be also capable of associating incoming packets from

a Data Center network with a virtual interface, based on the 20-bit VPN label contained in the packets.
The protocol used by the VPN first-hop routers to signal VPNs should be independent of the transport network protocol as long as the transport encapsulation has the ability to carry a 20-bit VPN label.

10.1. Data Center Transport Network

MPLS/VPN technology based on [RFC 4364] specifies several different encapsulation methods for connecting PE routers, namely Label Switched Paths (LSPs), IP tunneling, and GRE tunneling. If LSPs are used in the transport network they could be signaled with LDP, in which case host (/32) routes to all PE routers must be propagated throughout the network, or with RSVP-TE, in which case a full mesh of RSVP-TE tunnels is required, generating a lot of state in the network core. If the number of LSPs is expected to be high, due to a large size of Data Center network, then IP or GRE encapsulation can be used, where the above mentioned scalability is not a concern due to route aggregation property of IP protocols.

10.2. BGP Requirements in a Data Center Environment

10.2.1. BGP Convergence and Routing Consistency

BGP was designed to carry very large amount of routing information but it is not a very fast converging protocol. In addition, the routing protocols, including BGP, have traditionally favored convergence (i.e., responsiveness to route change due to failure or policy change) over routing consistency. Routing consistency means that a router forwards a packet strictly along the path adopted by the upstream routers. When responsiveness is favored, a router applies a received update immediately to its forwarding table before propagating the update to other routers, including those that potentially depend upon the outcome of the update. The route change responsiveness comes at the cost of routing blackholes and loops.

Routing consistency across Data Center is important because in large Data Centers thousands of Virtual Machines can be simultaneously moved between server racks due to maintenance, for example. If packets sent by the Virtual Machines that are being moved are dropped (because they do not follow a live path), the active network connections on those VMs will be dropped. To minimize the disruption to the established communications during VM migration, the live path continuity is required.

10.2.2. VM Mobility Support

To overcome BGP convergence and route consistency limitations, the forwarding plane techniques that support fast convergence should be used. In fact, there exist forwarding plane techniques that support fast convergence by removing from the forwarding table a locally learn route and instantaneously using already installed new routing information to a given destination. This technique is often referred to as "local repair". It allows to forward traffic (almost) continuously to a VM that has migrated to a new physical location using an indirect forwarding path or tunnel via VM's old location (i.e., old VM forwarder). The traffic path is restored locally at the VM's old location while the network converges to the new location of the migrated VM. Eventually, the network converges to optimal path and bypasses the local repair.
BGP should assist in the local repair techniques by advertizing multiple and not only the best path to a given destination.

10.2.3. Optimizing Route Distribution

When virtual networks are triggered based on the IP communication (as proposed in this document), the Route Target Constraint extension [RFC 4684] of BGP should be used to optimize the route distribution for sparse virtual network events. This technique ensures that only those VPN forwarders that have local participants in a particular data plane event receive its routing information. This also decreases the total load on the upstream BGP speakers.

10.2.4. Inter-operability with MPLS/BGP VPNs

As was stated in section 10, the IP-VPN DC solution should be fully inter-operable with MPLS/BGP VPNs. MPLS/BGP VPN technology is widely supported on routers and other appliances. When connecting a Data Center virtual network with other services/networks, it is not necessary to advertize the specific VM host routes but rather the aggregated routing information. A router or appliance within a Data Center can be used to aggregate VPN's IP routing information and advertize the aggregated prefixes. The aggregated prefixes would be advertized with the router/appliance IP address as BGP next-hop and with locally assigned aggregate 20-bit label. The aggregate label will trigger a destination IP lookup in its corresponding VRF on all the packets entering the virtual network.

11. Virtual Machine Migration Requirement

The "Virtual Machine live migration" (a.k.a. VM mobility) is highly desirable for many reasons such as efficient and flexible resource sharing, Data Center migration, disaster recovery, server redundancy, or service bursting. VM live migration consists in moving a virtual machine from one physical server to another, while preserving the VM's active network connections (e.g., TCP and higher-level sessions).

VM live mobility primarily happens within the same physical Data Center but VM live mobility between Data Centers might be also required. The IP-VPN Data Center solutions need to address both intra-Data Center and inter-Data Center VM live mobility.

Traditional Data Center deployments have followed IP subnet boundary, i.e., hosts often stayed in the same IP subnet and a host had to change its IP address when it moved to a different location. Such architecture have worked well when hosts were dedicated to an application and resided in physical proximity to each other. These assumptions are not true in the IaaS environment where compute resources associated with a given application can be spread and dynamically move across a large Data Center.

Many DC design proposals are trying to address the VM mobility with data-center wide VLANs using Data Center-wide Layer 2 broadcast domains. With data-center wide VLANs, a VM move is handled by generating gratuitous ARP reply to update all ARP caches and switch learning tables. Since a virtual subnet locality cannot be preserved in a large Data Center, a virtual subnet (VLAN) must be present on every Data Center switch, limiting the number of virtual networks to 4094. Even if a Layer 2 Data Center solution is able to minimize or eliminate the ARP flooding across Data Center core, all edge switches still have to perform dynamic VM MAC learning and maintain VM's MAC-to-IP mappings.

Since in large Data Centers physical proximity of computing resources cannot be assumed, grouping of hosts into subnets does not provide any VM mobility benefits. Rather, VM mobility in a large Data Center should be based on a collection of host routes spread randomly across a large physical area.

When dealing with IP-only applications it is not only sufficient but optimal to forward the traffic based on Layer 3 rather than on Layer 2 information. The MAC addresses of Virtual Machines are irrelevant to IP services and introduce unnecessary overhead (i.e., maintaining ARP caches of VM MACs) and complications when VMs move (e.g., when VM's MAC address is changed in its new location). IP-based VPN

connectivity solution is a cost effective and scalable approach to

solve VM mobility problem. In IP-VPN DC a VM move is handled by a route advertisement.

To accommodate live migration of Virtual Machines, it is desirable to assign a permanent IP address to a VM that remains with the VM after it moves. Typically, a VM/application reaches the off-subnet destinations via a default gateway, which should be the first-hop router (in the virtual topology). A VM/application should reach the on-subnet destinations via an ARP proxy which again should be the VPN first-hop router. A VM/application cannot change the default gateway's IP and MAC addresses during live migration, as it would require changes to TCP/IP stack in the guest OS. Hence, the first-hop VPN router should use a common, locally significant IP address and a common virtual MAC address to support VM live mobility. More specifically, this IP address and the MAC address should be the same on all first-hop VPN routers in order to support the VM moves between different physical machines. Moreover, in order to preserve virtual network and infrastructure separation, the IP and MAC addresses of the first-hop routers should be shared among all virtual IP-subnets/VPNs. Since the first-hop router always performs an IP lookup on every packet destination IP address, the VM traffic is forwarded on the optimal path and traverses the Data Center network only once.

The VM live migration has to be transparent to applications and any external entity interacting with the applications. This implies that the VM's network connectivity restoration time is critical. The transport sessions can typically survive over several seconds of disruption, however, applications may have sub-second latency requirement for their correct operation.

To minimize the disruption to the established communications during VM migration, the control plane of a DC solution should be able to differentiate between VM activation in a new location from advertising its host route to the network. This will enable the VPN first-hop routers forwarders to install a route to VM's new location prior to its migration, allowing the traffic to be tunneled via the first-hop router at the VM's old location. There are techniques available in BGP as well as in forwarding plane that support fast convergence due to withdrawal or replacement of current or less preferred forwarding information (see section 10.2 for more detailed description of such technique).

12. IP-VPN Data Center Use Case: Virtualization of Mobile Network

Application access is being done increasingly from clients such as

cell phones or tablets connecting via private or public WiFi access

points, or 3G/LTE wireless access. Enterprises with a mobile workforce need to access resources in the enterprise VPN while they are traveling, e.g., sales data from a corporate database. The mobile workforce might also, for security reasons, be equipped with disk-less notebooks which rely on the enterprise VPN for all file accesses. The mobile workforce applications may occasionally need to utilize the compute resources and other functions (e.g., storage) that the enterprise hosts on the infrastructure of a cloud computing provider. The mobile devices might require simultaneous access to resources in both, the cloud infrastructure as well as the enterprise VPN.
The enterprise wide area network may use a provider-based MPLS/BGP VPN service. The wireless service providers already use MPLS/BGP VPNs for enterprise customer isolation in the mobile packet core elements. Using the same VPN technology in the service provider Data Center network (or in a public Data Center network) is a natural extension.

Furthermore, there is a need to instantiate mobile applications themselves as virtual networks in order to improve application performance (e.g., latency, Quality-of-Service) or to enable new applications with specialized requirements. In addition it might be required that the application's computing resource is made to be part of the mobility network itself and placed as close as possible to a mobile user. Since LTE data and voice applications use IP protocols only, the IP-VPN solution to virtualization of compute resources in mobile networks would be the optimal approach.

The infrastructure of a large scale mobility network could itself be virtualized and made available in the form of virtual private networks to organizations that do not want to spend the required capital. The Mobile Core functions can be realized via software running on virtual machines in a service-provider-class compute environment. The functional entities such as Service-Gateways (S-GW), Packet-Gateways (P-GW), or Policy Charging and Enforcement Function (PCEF) of the LTE system can be run as applications on virtual machines, coordinated by an orchestrator and managed by a hypervisor. Virtualized packet core network elements (PCEF, S-GW, P-GW) could be placed anywhere in the mobile network infrastructure, as long as the IP connectivity is provided. The virtualization of the Mobile Core functions running on a private computing environment has many benefits, including faster service delivery, better economies of scale, simpler operations.
Since the LTE (Long Term Evolution) and Evolved Packet Core (EPC) system are all-IP networks, the IP-VPN solution to mobile network virtualization is the best fit.

13. Security Considerations

The document presents the problems need to be addressed in the L3VPN for Data Center space. The requirements and solutions will be documented separately.

The security considerations for general requirements or individual solutions will be documented in the relevant documents.

14. IANA Considerations

This document contains no new IANA considerations.

15. Normative References

[RFC 4364] Rosen, E. and Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPNs)", RFC 4364, February 2006.

[RFC 4023] Worster, T., Rekhter, Y. and E. Rosen, "Encapsulating in IP or Generic Routing Encapsulation (GRE)", RFC 4023, March 2005.

[RFC 4684] Marques, P., Bonica, R., Fang, L., Martini, L., Raszuk, R., Patel, K. and J. Guichard, "Constrained Route Distribution for Border Gateway Protocol/Multiprotocol Label Switching (BGP/MPLS) Internet Protocol (IP) Virtual Private Networks (VPNs)", RFC 4684, November 2006.

16. Informative References

[RFC 2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

17. Authors' Addresses

Maria Napierala
AT&T
200 Laurel Avenue
Middletown, NJ 07748
Email: mnapierala@att.com

Luyuan Fang
Cisco Systems
111 Wood Avenue South

Iselin, NJ 08830, USA
Email: lufang@cisco.com

Dennis Cai
Cisco Systems
725 Alder Drive
Milpitas, CA 95035, USA
Email: dcai@cisco.com

18. Acknowledgements

The authors would like to thank Pedro Marques for his helpful comments and input.