Internet Engineering Task Force (IETF)
Request for Comments: 6212
Category: Informational
ISSN: 2070-1721

R. Koodli
Cisco Systems

July 2011

Mobile Networks Considerations for IPv6 Deployment

Mobile Internet access from smartphones and other mobile devices is accelerating the exhaustion of IPv4 addresses. IPv6 is widely seen as crucial for the continued operation and growth of the Internet, and in particular, it is critical in mobile networks. This document discusses the issues that arise when deploying IPv6 in mobile networks. Hence, this document can be a useful reference for service providers and network designers.

Status of This Memo

This document is not an Internet Standards Track specification; it is published for informational purposes.

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Not all documents approved by the IESG are a candidate for any level of Internet Standard; see Section 2 of RFC 5741.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at http://www.rfc-editor.org/info/rfc6312.

Copyright Notice

Table of Contents

1. Introduction

The dramatic growth of the Mobile Internet is accelerating the exhaustion of the available IPv4 addresses. It is widely accepted that IPv6 is necessary for the continued operation and growth of the Internet in general and of the Mobile Internet in particular. While IPv6 brings many benefits, certain unique challenges arise when deploying it in mobile networks. This document describes such challenges and outlines the applicability of the existing IPv6 deployment solutions. As such, it can be a useful reference document for service providers as well as network designers. This document does not propose any new protocols or suggest new protocol specification work.

The primary considerations that we address in this document on IPv6 deployment in mobile networks are:

- Public and Private IPv4 address exhaustion and implications to mobile network deployment architecture;
- Placement of Network Address Translation (NAT) functionality and its implications;
- IPv6-only deployment considerations and roaming implications; and
- Fixed-Mobile Convergence and implications to overall architecture.

In the following sections, we discuss each of these in detail.

For the most part, we assume the Third Generation Partnership Project (3GPP) 3G and 4G network architectures specified in [3GPP.3G] and [3GPP.4G]. However, the considerations are general enough for other mobile network architectures as well [3GPP2.EHRPD].

2. Reference Architecture and Terminology

The following is a reference architecture of a mobile network.

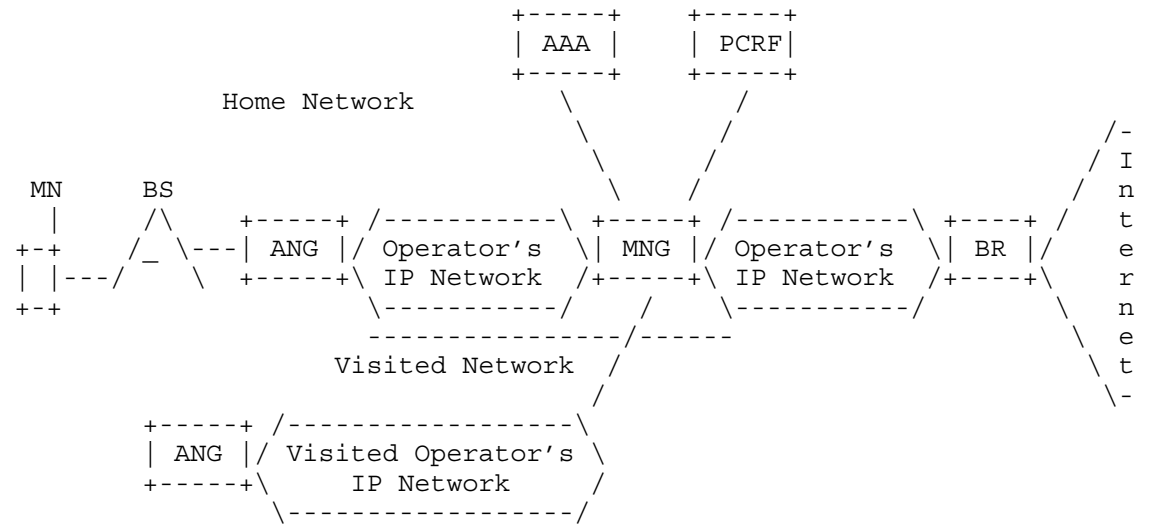

Figure 1: Mobile Network Architecture

A Mobile Node (MN) connects to the mobile network either via its Home Network or via a Visited Network when the user is roaming outside of the Home Network. In the 3GPP network architecture, an MN accesses the network by connecting to an Access Point Name (APN), which maps to a mobile gateway. Roughly speaking, an APN is similar to a Service Set Identifier (SSID) in wireless LAN. An APN is a logical concept that can be used to specify what kinds of services, such as Internet access, high-definition video streaming, content-rich gaming, and so on, that an MN is entitled to. Each APN can specify what type of IP connectivity (i.e., IPv4, IPv6, IPv4v6) is enabled on that particular APN.

While an APN directs an MN to an appropriate gateway, the MN needs an end-to-end "link" to that gateway. In the Long-Term Evolution (LTE) networks, this link is realized through an Evolved Packet System (EPS) bearer. In the 3G Universal Mobile Telecommunications System (UMTS) networks, such a link is realized through a Packet Data Protocol (PDP) context. The end-to-end link traverses multiple nodes, which are defined below:

- Base Station (BS): The radio Base Station provides wireless connectivity to the MN.

- Access Network Gateway (ANG): The ANG forwards IP packets to and from the MN. Typically, this is not the MN's default router, and the ANG does not perform IP address allocation and management for the mobile nodes. The ANG is located either in the Home Network or in the Visited Network.

- The Mobile Network Gateway (MNG): The MNG is the MN's default router, which provides IP address management. The MNG performs functions such as offering Quality of Service (QoS), applying subscriber-specific policy, and enabling billing and accounting; these functions are sometimes collectively referred to as "subscriber-management" operations. The mobile network architecture, as shown in Figure 1, defines the necessary protocol interfaces to enable subscriber-management operations. The MNG is typically located in the Home Network.

- Border Router (BR): As the name implies, a BR borders the Internet for the mobile network. The BR does not perform subscriber management for the mobile network.

- Authentication, Authorization, and Accounting (AAA): The general functionality of AAA is used for subscriber authentication and authorization for services as well as for generating billing and accounting information.

  In 3GPP network environments, the subscriber authentication and the subsequent authorization for connectivity and services is provided using the "Home Location Register" (HLR) / "Home Subscriber Server" (HSS) functionality.

- Policy and Charging Rule Function (PCRF): The PCRF enables applying policy and charging rules at the MNG.

In the rest of this document, we use the terms "operator", "service provider", and "provider" interchangeably.

3. IPv6 Considerations

3.1. IPv4 Address Exhaustion

It is generally agreed that the pool of public IPv4 addresses is nearing its exhaustion. The IANA has exhausted the available '/8' blocks for allocation to the Regional Internet Registries (RIRs). The RIRs themselves have either "run out" of their blocks or are projected to exhaust them in the near future. This has led to a heightened awareness among service providers to consider introducing technologies to keep the Internet operational. For providers, there are two simultaneous approaches to addressing the run-out problem:

delaying the IPv4 address pool exhaustion (i.e., conserving their existing pool) and introducing IPv6 in operational networks. We consider both in the following.

Delaying public IPv4 address exhaustion for providers involves assigning private IPv4 addressing for end-users or extending an IPv4 address with the use of port ranges, which requires tunneling and additional signaling. A mechanism such as the Network Address Translator (NAT) is used at the provider premises (as opposed to customer premises) to manage the private IP address assignment and access to the Internet. In the following, we primarily focus on translation-based mechanisms such as NAT44 (i.e., translation from public IPv4 to private IPv4 and vice versa) and NAT64 (i.e., translation from public IPv6 to public IPv4 and vice versa). We do this because the 3GPP architecture already defines a tunneling infrastructure with the General Packet Radio Service (GPRS) Tunneling Protocol (GTP), and the architecture allows for dual-stack and IPv6-only deployments.

In a mobile network, the IPv4 address assignment for an MN is performed by the MNG. In the 3GPP network architecture, this assignment is performed in conjunction with the Packet Data Network (PDN) connectivity establishment. A PDN connection implies an end-end link (i.e., an EPS bearer in 4G LTE or a PDP context in 3G UMTS) from the MN to the MNG. There can be one or more PDN connections active at any given time for each MN. A PDN connection may support both IPv4 and IPv6 traffic (as in a dual-stack PDN in 4G LTE networks), or it may support only one of the two traffic types (as in the existing 3G UMTS networks). The IPv4 address is assigned at the time of PDN connectivity establishment or is assigned using DHCP after the PDN connectivity is established. In order to delay the exhaustion of public IPv4 addresses, this IP address needs to be a private IPv4 address that is translated into a shared public IPv4 address. Hence, there is a need for a private-public IPv4 translation mechanism in the mobile network.

In the Long-Term Evolution (LTE) 4G network, there is a requirement for an always-on PDN connection in order to reliably reach a mobile user in the All-IP network. This requirement is due to the need for supporting Voice over IP service in LTE, which does not have circuit-based infrastructure. If this PDN connection were to use IPv4 addressing, a private IPv4 address is needed for every MN that attaches to the network. This could significantly affect the availability and usage of private IPv4 addresses. One way to address this is by making the always-on PDN (that requires voice service) to be IPv6. The IPv4 PDN is only established when the user needs it.

The 3GPP standards also specify a deferred IPv4 address allocation on a dual-stack IPv4v6 PDN at the time of connection establishment. This has the advantage of a single PDN for IPv6 and IPv4 along with deferring IPv4 address allocation until an application needs it. The deferred address allocation requires support for a dynamic configuration protocol such as DHCP as well as appropriate triggers to invoke the protocol. Such a support does not exist today in mobile phones. The newer iterations of smartphones could provide such support. Also, the tethering of smartphones to laptops (which typically support DHCP) could use deferred allocation depending on when a laptop attaches to the smartphone. Until appropriate triggers and host stack support is available, the applicability of the address-deferring option may be limited.

On the other hand, in the existing 3G UMTS networks, there is no requirement for an always-on connection even though many smartphones seldom relinquish an established PDP context. The existing so-called pre-Release-8 deployments do not support the dual-stack PDP connection. Hence, two separate PDP connections are necessary to support IPv4 and IPv6 traffic. Even though some MNs, especially the smartphones, in use today may have IPv6 stack, there are two remaining considerations. First, there is little operational experience and compliance testing with these existing stacks. Hence, it is expected that their use in large deployments may uncover software errors and interoperability problems that inhibit providing services based on IPv6 for such hosts. Second, only a fraction of current phones in use have such a stack. As a result, providers need to test, deploy, and operationalize IPv6 as they introduce new handsets, which also continue to need access to the predominantly IPv4 Internet.

The considerations from the preceeding paragraphs lead to the following observations. First, there is an increasing need to support private IPv4 addressing in mobile networks because of the public IPv4 address run-out problem. Correspondingly, there is a greater need for private-public IPv4 translation in mobile networks. Second, there is support for IPv6 in both 3G and 4G LTE networks already in the form of PDP context and PDN connections. To begin with, operators can introduce IPv6 for their own applications and services. In other words, the IETF's recommended model of dual-stack IPv6 and IPv4 networks is readily applicable to mobile networks with the support for distinct APNs and the ability to carry IPv6 traffic on PDP/PDN connections. The IETF dual-stack model can be applied using a single IPv4v6 PDN connection in Release-8 and onwards but requires separate PDP contexts in the earlier releases. Finally, operators can make IPv6 the default for always-on mobile connections using either the IPv4v6 PDN or the IPv6 PDN and use IPv4 PDNs as necessary.

3.2. NAT Placement in Mobile Networks

In the previous section, we observed that NAT44 functionality is needed in order to conserve the available pool and delay public IPv4 address exhaustion. However, the available private IPv4 pool itself is not abundant for large networks such as mobile networks. For instance, the so-called NET10 block [RFC1918] has approximately 16.7 million private IPv4 addresses starting with 10.0.0.0. A large mobile service provider network can easily have more than 16.7 million subscribers attached to the network at a given time. Hence, the private IPv4 address pool management and the placement of NAT44 functionality becomes important.

In addition to the developments cited above, NAT placement is important for other reasons as well. Access networks generally need to produce network and service usage records for billing and accounting. This is true also for mobile networks where "subscriber management" features (i.e., QoS, Policy, and Billing and Accounting) can be fairly detailed. Since a NAT introduces a binding between two addresses, the bindings themselves become necessary information for subscriber management. For instance, the offered QoS on private IPv4 address and the (shared) public IPv4 address may need to be correlated for accounting purposes. As another example, the Application Servers within the provider network may need to treat traffic based on policy provided by the PCRF. If the IP address seen by these Application Servers is not unique, the PCRF needs to be able to inspect the NAT binding to disambiguate among the individual MNs. The subscriber session management information and the service usage information also need to be correlated in order to produce harmonized records. Furthermore, there may be legal requirements for storing the NAT binding records. Indeed, these problems disappear with the transition to IPv6. For now, it suffices to assert that NAT introduces state, which needs to be correlated and possibly stored with other routine subscriber information.

Mobile network deployments vary in their allocation of IP address pools. Some network deployments use the "centralized model" where the pool is managed by a common node, such as the PDN's BR, and the pool shared by multiple MNGs all attached to the same BR. This model has served well in the pre-3G deployments where the number of subscribers accessing the Mobile Internet at any given time has not exceeded the available address pool. However, with the advent of 3G networks and the subsequent dramatic growth in the number of users on the Mobile Internet, service providers are increasingly forced to consider their existing network design and choices. Specifically, providers are forced to address private IPv4 pool exhaustion as well as scalable NAT solutions.

In order to tackle the private IPv4 exhaustion in the centralized model, there would be a need to support overlapped private IPv4 addresses in the common NAT functionality as well as in each of the gateways. In other words, the IP addresses used by two or more MNs (which may be attached to the same MNG) are very likely to overlap at the centralized NAT, which needs to be able to differentiate traffic. Tunneling mechanisms such as Generic Routing Encapsulation (GRE) [RFC2784] [RFC2890], MPLS [RFC3031] VPN tunnels, or even IP-in-IP encapsulation [RFC2003] that can provide a unique identifier for a NAT session can be used to separate overlapping private IPv4 traffic as described in [GI-DS-LITE]. An advantage of centralizing the NAT and using the overlapped private IPv4 addressing is conserving the limited private IPv4 pool. It also enables the operator's enterprise network to use IPv6 from the MNG to the BR; this (i.e., the need for an IPv6-routed enterprise network) may be viewed as an additional requirement by some providers. The disadvantages include the need for additional protocols to correlate the NAT state (at the common node) with subscriber session information (at each of the gateways), suboptimal MN-MN communication, absence of subscriber-aware NAT (and policy) function, and, of course, the need for a protocol from the MNG to BR itself. Also, if the NAT function were to experience failure, all the connected gateway service will be affected. These drawbacks are not present in the "distributed" model, which we discuss in the following.

In a distributed model, the private IPv4 address management is performed by the MNG, which also performs the NAT functionality. In this model, each MNG has a block of 16.7 million unique addresses, which is sufficient compared to the number of mobile subscribers active on each MNG. By distributing the NAT functionality to the edge of the network, each MNG is allowed to reuse the available NET10 block, which avoids the problem of overlapped private IPv4 addressing at the network core. In addition, since the MNG is where subscriber management functions are located, the NAT state correlation is readily enabled. Furthermore, an MNG already has existing interfaces to functions such as AAA and PCRF, which allows it to perform subscriber management functions with the unique private IPv4 addresses. Finally, the MNG can also pass-through certain traffic types without performing NAT to the Application Servers located within the service provider's domain, which allows the servers to also identify subscriber sessions with unique private IPv4 addresses. The disadvantages of the "distributed model" include the absence of centralized addressing and centralized management of NAT.

In addition to the two models described above, a hybrid model is to locate NAT in a dedicated device other than the MNG or the BR. Such a model would be similar to the distributed model if the IP pool supports unique private addressing for the mobile nodes, or it would

be similar to the centralized model if it supports overlapped private IP addresses. In any case, the NAT device has to be able to provide the necessary NAT session binding information to an external entity (such as AAA or PCRF), which then needs to be able to correlate those records with the user's session state present at the MNG.

The foregoing discussion can be summarized as follows. First, the management of the available private IPv4 pool has become important given the increase in Mobile Internet users. Mechanisms that enable reuse of the available pool are required. Second, in the context of private IPv4 pool management, the placement of NAT functionality has implications to the network deployment and operations. The centralized models with a common NAT have the advantages of continuing their legacy deployments and the reuse of private IPv4 addressing. However, they need additional functions to enable traffic differentiation and NAT state correlation with subscriber state management at the MNG. The distributed models also achieve private IPv4 address reuse and avoid overlapping private IPv4 traffic in the operator's core, but without the need for additional mechanisms. Since the MNG performs (unique) IPv4 address assignment and has standard interfaces to AAA and PCRF, the distributed model also enables a single point for subscriber and NAT state reporting as well as policy application. In summary, providers interested in readily integrating NAT with other subscriber management functions, as well as conserving and reusing their private IPv4 pool, may find the distributed model compelling. On the other hand, those providers interested in common management of NAT may find the centralized model more compelling.

3.3. IPv6-Only Deployment Considerations

As we observed in the previous section, the presence of NAT functionality in the network brings up multiple issues that would otherwise be absent. NAT should be viewed as an interim solution until IPv6 is widely available, i.e., IPv6 is available for mobile users for all (or most) practical purposes. Whereas NATs at provider premises may slow down the exhaustion of public IPv4 addresses, expeditious and simultaneous introduction of IPv6 in the operational networks is necessary to keep the Internet "going and growing". Towards this goal, it is important to understand the considerations in deploying IPv6-only networks.

There are three dimensions to IPv6-only deployments: the network itself, the mobile nodes, and the applications, represented by the 3-tuple {nw, mn, ap}. The goal is to reach the coordinate {IPv6, IPv6, IPv6} from {IPv4, IPv4, IPv4}. However, there are multiple paths to arrive at this goal. The classic dual-stack model would traverse the coordinate {IPv4v6, IPv4v6, IPv4v6}, where each

dimension supports co-existence of IPv4 and IPv6. This appears to be the path of least disruption, although we are faced with the implications of supporting large-scale NAT in the network. There is also the cost of supporting separate PDP contexts in the existing 3G UMTS networks. The other intermediate coordinate of interest is {IPv6, IPv6, IPv4}, where the network and the MN are IPv6-only, and the Internet applications are recognized to be predominantly IPv4. This transition path would, ironically, require interworking between IPv6 and IPv4 in order for the IPv6-only MNs to be able to access IPv4 services and applications on the Internet. In other words, in order to disengage NAT (for IPv4-IPv4), we need to introduce another form of NAT (i.e., IPv6-IPv4) to expedite the adoption of IPv6.

It is interesting to consider the preceeding discussion surrounding the placement of NAT for IPv6-IPv4 interworking. There is no overlapping private IPv4 address problem because each IPv6 address is unique and there are plenty of them available. Hence, there is also no requirement for (IPv6) address reuse, which means no protocol is necessary in the centralized model to disambiguate NAT sessions. However, there is an additional requirement of DNS64 [RFC6147] functionality for IPv6-IPv4 translation. This DNS64 functionality must ensure that the synthesized AAAA record correctly maps to the IPv6-IPv4 translator.

IPv6-only deployments in mobile networks need to reckon with the following considerations. First, both the network and the MNs need to be IPv6 capable. Expedited network upgrades as well as rollout of MNs with IPv6 would greatly facilitate this. Fortunately, the 3GPP network design for LTE already requires the network nodes and the mobile nodes to support IPv6. Even though there are no requirements for the transport network to be IPv6, an operational IPv6 connectivity service can be deployed with appropriate existing tunneling mechanisms in the IPv4-only transport network. Hence, a service provider may choose to enforce IPv6-only PDN and address assignment for their own subscribers in their Home Networks (see Figure 1). This is feasible for the newer MNs when the mobile network is able to provide IPv6-only PDN support and IPv6-IPv4 interworking for Internet access. For the existing MNs, however, the provider still needs to be able to support IPv4-only PDP/PDN connectivity.

Migration of applications to IPv6 in MNs with IPv6-only PDN connectivity brings challenges. The applications and services offered by the provider obviously need to be IPv6-capable. However, an MN may host other applications, which also need to be IPv6-capable in IPv6-only deployments. This can be a "long-tail" phenomenon; however, when a few prominent applications start offering IPv6, there can be a strong incentive to provide application-layer (e.g., socket

interface) upgrades to IPv6. Also, some IPv4-only applications may be able to make use of alternative access such as WiFi when available. A related challenge in the migration of applications is the use of IPv4 literals in application layer protocols (such as XMPP) or content (as in HTML or XML). Some Internet applications expect their clients to supply IPv4 addresses as literals, and this will not be possible with IPv6-only deployments. Some of these experiences and the related considerations in deploying an IPv6-only network are documented in [ARKKO-V6]. In summary, migration of applications to IPv6 needs to be done, and such a migration is not expected to be uniform across all subsets of existing applications.

Voice over LTE (VoLTE) also brings some unique challenges. The signaling for voice is generally expected to be available for free while the actual voice call itself is typically charged on its duration. Such a separation of signaling and the payload is unique to voice, whereas an Internet connection is accounted without specifically considering application signaling and payload traffic. This model is expected to be supported even during roaming. Furthermore, providers and users generally require voice service regardless of roaming, whereas Internet usage is subject to subscriber preferences and roaming agreements. This requirement to ubiquitously support voice service while providing the flexibility for Internet usage exacerbates the addressing problem and may hasten provisioning of VoLTE using the IPv6-only PDN.

As seen earlier, roaming is unique to mobile networks, and it introduces new challenges. Service providers can control their own network design but not their peers' networks, which they rely on for roaming. Users expect uniformity in experience even when they are roaming. This imposes a constraint on providers interested in IPv6-only deployments to also support IPv4 addressing when their own (outbound) subscribers roam to networks that do not offer IPv6. For instance, when an LTE deployment is IPv6-only, a roamed 3G network may not offer IPv6 PDN connectivity. Since a PDN connection involves the radio base station, the ANG, and the MNG (see Figure 1), it would not be possible to enable IPv6 PDN connectivity without roamed network support. These considerations also apply when the visited network is used for offering services such as VoLTE in the so-called Local Breakout model; the roaming MN's capability as well as the roamed network capability to support VoLTE using IPv6 determine whether fallback to IPv4 would be necessary. Similarly, there are inbound roamers to an IPv6-ready provider network whose MNs are not capable of IPv6. The IPv6-ready provider network has to be able to support IPv4 PDN connectivity for such inbound roamers. There are encouraging signs that the existing deployed network nodes in the 3GPP architecture already provide support for IPv6 PDP context. It

would be necessary to scale this support for a (very) large number of mobile users and offer it as a ubiquitous service that can be accounted for.

In summary, IPv6-only deployments should be encouraged alongside the dual-stack model, which is the recommended IETF approach. This is relatively straightforward for an operator's own services and applications, provisioned through an appropriate APN and the corresponding IPv6-only PDP or EPS bearer. Some providers may consider IPv6-only deployment for Internet access as well, and this would require IPv6-IPv4 interworking. When the IPv6-IPv4 translation mechanisms are used in IPv6-only deployments, the protocols and the associated considerations specified in [RFC6146] and [RFC6145] apply. Finally, such IPv6-only deployments can be phased-in for newer mobile nodes, while the existing ones continue to demand IPv4-only connectivity.

Roaming is important in mobile networks, and roaming introduces diversity in network deployments. Until IPv6 connectivity is available in all mobile networks, IPv6-only mobile network deployments need to be prepared to support IPv4 connectivity (and NAT44) for their own outbound roaming users as well as for inbound roaming users. However, by taking the initiative to introduce IPv6-only for the newer MNs, the mobile networks can significantly reduce the demand for private IPv4 addresses.

3.4. Fixed-Mobile Convergence

Many service providers have both fixed broadband and mobile networks. Access networks are generally disparate, with some common characteristics but with enough differences to make it challenging to achieve "convergence". For instance, roaming is not a consideration in fixed access networks. An All-IP mobile network service provider is required to provide voice service, whereas this is not required for a fixed network provider. A "link" in fixed networks is generally capable of carrying IPv6 and IPv4 traffic, whereas not all mobile networks have "links" (i.e., PDP/PDN connections) capable of supporting IPv6 and IPv4. Indeed, roaming makes this problem worse when a portion of the link (i.e., the Home Network in Figure 1) is capable of supporting IPv6 and the other portion of the link (i.e., the Visited Network in Figure 1) is not. Such architectural differences, as well as policy and business model differences make convergence challenging.

Nevertheless, within the same provider's space, some common considerations may apply. For instance, IPv4 address management is a common concern for both of the access networks. This implies that the same mechanisms discussed earlier, i.e., delaying IPv4 address