6MAN
Internet-Draft
Intended status: Informational
Expires: 20 November 2025

P. Thubert, Ed.

M. Richardson
Sandelman

19 May 2025

Architecture and Framework for IPv6 over Non-Broadcast Access
draft-ietf-6man-ipv6-over-wireless-08

Abstract

This document presents an architecture and framework for IPv6 access networks that decouples the network-layer concepts of Links, Interface, and Subnets from the link-layer concepts of links, ports, and broadcast domains, and limits the reliance on link-layer broadcasts. This architecture is suitable for IPv6 over any network, including non-broadcast networks, which is typically the case for intangible media such as wireless and virtual networks such as overlays. A study of the issues with IPv6 ND over intangible media is presented, and a framework to solve those issues within the new architecture is proposed.

Status of This Memo

This Internet-Draft is submitted 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 https://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 20 November 2025.

Table of Contents

1. Introduction

IEEE Std. 802.1 [IEEE Std. 802.1] Ethernet Bridging provides an efficient and reliable broadcast service for wired networks; applications and protocols have been built that heavily depend on that feature for their core operation. Unfortunately, Low-Power Lossy Networks (LLNs) and Wireless Local Area Networks (WLANs) generally do not benefit from the same reliable and cheap broadcast capabilities as legacy Ethernet yellow wires, and protocols that rely on broadcasts are less suited in those environments.

Similarly, the use of broadcast is discouraged in large Data Center (DC) fabrics and DC Interconnect (DCI) that extend the lower-layer links in large and physically distributed topologies, e.g., as meshes of point-to-point (P2P) tunnels. In such case, an overlay broadcast service is typically emulated as ingress (or reflector) replication and generates massive amounts of underlay unicast messages, possibly over expensive Wide Area Network (WAN) links.

All in all, as IPv6 [RFC8200] networks migrate from a physical wires to virtual or intangible media, the common requirement shows to decouple the abstractions that are manipulated at the network layer from physical properties such as broadcast capabilities and transitivity that are handled at the lower layers.

The original IPv6 Neighbor Discovery Protocol [RFC4861], [RFC4862] (IPv6 ND) relies heavily on broadcast operation for Router Advertisement (RA), Address Resolution (AR), and Duplicate address Detection (DAD). In modern networks, this may be inefficient (due to many replications over constrained links), unreliable (broadcast may be lost in transmission), counterproductive to network operations (broadcast storms), prone to impersonation and multiplication attacks (a unicast from the outside may cause a broadcast inside), and may be detrimental to privacy (an observer inside the network can discover other onlink addresses).

This document presents an architecture for IPv6 access networks that 1) decouples the network-layer concepts of Links, Interface, and Subnets from the link-layer concepts of links, ports, and broadcast domains, and thereby 2) limits the reliance on (and impact thereof) link-layer broadcasts inside the Subnet. This architecture is suitable for IPv6 over any network, including modern Non-Broadcast MultiAccess (NBMA) and non-transitive Point-to-Multipoint (P2MP) networks. A study of the issues with IPv6 ND over wireless media is presented, and a framework to solve those issues within the new architecture is proposed.

2. Acronyms

This document uses the following abbreviations:

6BBR: Backbone Router
6LN: (6LoWPAN) Node
6LR: (6LoWPAN) Router
ARO: address Registration Option
BGP: Border Gateway Protocol
BLE: Bluetooth(R) Low Energy
DAC: Duplicate address Confirmation (message)
DC: Data Center
DAD: Duplicate address Detection
DAR: Duplicate address Request (message)
DOS: Denial of Service (attack)
EDAC: Extended Duplicate address Confirmation
EDAR: Extended Duplicate address Request
EVPN: Ethernet VPN
FDM: Frequency-Division Multiplexing
IGP: Interior Gateway Protocol
IPSP: Internet Protocol Support Profile
LAN: Local Area Network
LISP: Locator/ID Separation Protocol
LLN: Low-Power and Lossy Network
LLA: link-local address
LoWPAN: Low-Power WPAN
MAC: Medium Access Control
MLSN: Multi-link Subnet
MLD: multicast Listener Discovery
NA: Neighbor Advertisement (message)
NBMA: Non-Broadcast Multi-Access (full mesh)
NCE: Neighbor Cache Entry
ND: Neighbor Discovery (protocol)
NDP: Neighbor Discovery Protocol
NS: Neighbor Solicitation (message)
P2P: Point-to-Point
P2MP: Point-to-Multipoint (partial mesh)

RPL: IPv6 Routing Protocol for LLNs
RA: Router Advertisement (message)
RS: Router Solicitation (message)
SGP: Subnet Gateway Protocol
SPL: Subnet Prefix Length
TDM: Time-Division Multiplexing
TSCH: Time-Slotted Channel Hopping
ULP: Upper-Layer Protocol
VLAN: Virtual LAN
VxLAN: Virtual Extensible LAN
VPN: Virtual Private Network
WAN: Wide Area Network
SND: Subnet Neighbor Discovery (protocol)
WLAN: Wireless Local Area Network
WPAN: Wireless Personal Area Network

3. Issues with IPv6 ND-Based Access

3.1. IPv6 ND and ND-Proxies

Though IPv6 ND was the state of the art when designed for early Ethernet links at the end of the twentieth century, it is less appropriate for modern networks such as wireless and overlays that cannot provide the same cheap and reliable broadcast as a shared yellow wire. The reactive AR operation was designed to limit the amount of memory that is needed for the ND cache, at times where memory was scarce in the adapters. This trade-off, broadcast bandwidth vs. memory in the adapters, should be reevaluated for networks where router memory is aplenty but broadcast has become expensive.

The IPv6 ND Neighbor Solicitation (NS) [RFC4861] message is used as a multicast IP packet for AR and DAD [RFC4862]. While the AR message is intended for one node that owns the Target address, the expectation for DAD is that there’s no node at all with that address. A message that is intended for at most one node is certainly a poor match for a broadcast operation.

The NS messages used in AR and DAD exchanges are sent at the network layer to a Solicited-Node multicast address (SNMA) [RFC4291] and should in theory only reach a very small group of nodes. But to support SNMA, the host must also support multicast Listener Discovery (MLD) [RFC3810], which may be an additional burden to a constrained stack, and the switches should support their own multicast routing and state, which is pushing the real problems to the lower layers.

Also, if implemented, the SNMA procedure would entail close to one state per address is every switch since there is often only one address with the same SNMA in the network - though the birthday paradox applies. This amount of memory may not have been available in early switches, and still makes little economical sense today when a complete ND cache requires one state per address in every router only, as opposed to one in every switch, when the Subnet prefix is advertised as not-onlink.

This is why, in practice, IPv6 ND messages to a SNMA are mapped to link-layer broadcast on Ethernet and Wireless networks, and to full ingress replication in overlays. multicast NS transmissions may occur when a node joins the network, moves, or wakes up and reconnects to the network. Over a very large fabric, this can generate hundreds of broadcasts per second.

If the broadcasts were blindly copied over Wi-Fi links, the link-layer broadcast traffic associated to ND network-layer multicast could consume enough bandwidth to cause a substantial degradation to the unicast service [MCAST EFFICIENCY]. This is why ND Proxies are deployed and charged to reduce the resulting flood; sadly, ND-Proxies are not fully reliable for the lack of a deterministic state on all existing addresses, which leads to unpredictable failures in IPv6 ND operations.

The IPv6 ND Neighbor Advertisement (NA) [RFC4861] message can also be sent as a multicast to all nodes, as a gratuitous information that can be used to override the address mapping in nodes with an existing Neighbor Cache Entry (NCE) for the Target address. If this is done, all nodes in the broadcast domain are impacted though there’s probably none or very few with an NCE. When it is not done, nodes with an NCE will be unable to reach the Target IP address until Neighbor Unreachability Detection (NUD) discovers the issue. Both alternatives are unsatisfactory, meaning that the whole approach should be revisited.

This problem can be alleviated by reducing the size of the broadcast domain that encompasses wireless access links. This has been done in the art of IP subnetting by partitioning the subnets and by routing between them, at the extreme by assigning a prefix, say a /64, to each wireless node (see [RFC8273]).

Another way to split the broadcast domain within a Subnet is to proxy the network-layer protocols that rely on link-layer broadcast operations at the boundary of the split broadcast domains, e.g., using the [IEEE Std. 802.11] "ARP proxy" at the Access Point. The correct operation of the proxy requires the exhaustive list of the IP addresses for which proxying is provided. Forming and maintaining

that knowledge is a hard problem in the general case of radio connectivity, which keeps changing with movements and variations in the environment that alter the range of transmissions. It is achieved in Wi-Fi networks through the proactive method of the wireless association, which is akin to the registration procedure in this architecture.

[SAVI] suggests discovering the addresses by snooping the IPv6 ND protocol, but that can also be unreliable. An IPv6 address may not be discovered immediately due to a packet loss. It may never be discovered in the case of a "silent" node that is not currently using one of its addresses, e.g., a printer that waits in wake-on-lan state. A change of anchor, e.g. due to a movement, may be missed or received out of order, leading to unreliable connectivity and an incomplete list of addresses. Bottom line: snooping IPv6 ND is not appropriate to form and maintain a deterministic knowledge of the IPv6 addresses of all the neighbors that are reachable over a network port.

3.2. The case of Wireless

Like Transparent Bridging, the IPv6 ND operation is reactive, and relies on IP multicast for the AR and DAD procedures. As discussed in Section 3, the network-layer multicast operation is typically implemented as a link-layer broadcast for the lack of an adapted and scalable link-layer multicast operation on most WLANs and Low-Power Personal Area Networks (LoWPANs). It results that on wireless, IPv6 ND multicast messages are typically broadcasted.

As opposed to unicast transmissions, the broadcast transmissions over wireless links are not subject to automatic retries (ARQ) and therefore are not reliable. Reducing the speed at the physical (PHY) layer for broadcast transmissions can increase the reliability, at the expense of a higher relative cost of broadcast on the overall available bandwidth.

Excessive use of broadcast by protocols such as IPv6 ND and Bonjour/ mDNS (see [RFC6762]) led network administrators to install multicast rate limiting to protect the network. Experimentally, this proved to have a dramatic effect on ND performance in large wireless networks. From some testing done at an IETF meeting a few years ago, it seemed that up to 90% of IPv6 link-local multicasts were dropped. The impact on user experience is usually limited since for the most part, the users will connect to addresses outside the Subnet and will not attempt to locate one another. The impact on the NOC might still be significant, since a failure related to ND operations might be transient and difficult to debunk after the fact.

Another experiment conducted during an IETF meeting consisted in manually forcing address duplicates to observe the DAD behavior. This experiment showed that DAD often (up to 80% of times was observed) fails to discover the duplication of IPv6 addresses, at least in a large wireless access networks, see [DAD ISSUES] for more. In practice, IPv6 addresses very rarely conflict, not because the address duplications are detected and resolved by the DAD operation, but thanks to the entropy of the typically 64-bit Interface IDs (IIDs) that makes a collision quasi-impossible for randomized IIDs. This is why, even when DAD fails, the user experience is rarely affected.

Excessive use of broadcast also places a toll on the battery of wireless devices such as IoT sensors and smartphones. On paper, a Wi-Fi station must keep its radio turned on to listen to the periodic series of broadcast frames. Most of those broadcasts are dropped at the network layer when the receiving node finds it is not interested, as is the case of NS messages when the node is not the Target. In order to protect the battery lifetime, a typical smartphone will listen at a multiple of the broadcast period, blindly ignoring a large ratio of the broadcasts, and making IPv6 ND operations even less reliable.

Net-net: broadcast transmissions are not reliable on wireless. Protocols designed for bridged networks that rely on broadcast transmissions often exhibit disappointing behaviors when employed unmodified on a local wireless medium (more in [RFC9119]). Even though there is at most one intended Target for a broadcast AR or DAD message, the broadcast impacts many wireless nodes over the whole Subnet (e.g., the ESS fabric), and yet the chances that intended Target receives the packet are limited. The fact that the user experience for IPv6 ND is not so dramatically affected only shows that those broadcasts are, for a large part, a useless waste of expensive resources.

Wi-Fi Access Points [IEEE Std. 802.11] (APs) deployed in an Extended Service Set (ESS) act as [IEEE Std. 802.1] bridges between the wireless stations (STA) and the wired backbone. As opposed to the classical Transparent (aka Learning) Bridge operation that installs the forwarding state reactively to traffic, the bridging state in the AP is established proactively, at the time of association. The association process registers the link-layer (MAC) address of the STA to the AP proactively, i.e., before it is needed. Based on that information, the AP maintains the exhaustive list of the associated MAC addresses and blocks the link-layer lookups for destination MAC addresses that are not associated to this AP. While the association procedure protects the wireless medium against broadcast-intensive Transparent Bridging lookups, the same problem remains at the network layer for the lack of a similar procedure in IPv6 ND.

3.3. The case of Overlays

Link-layer Overlays (e.g., VLANs) reduce the broadcast domain from the physical one, whereas network-layer overlays (e.g., VxLAN) can extend the Subnet beyond the limits of the physical network, enabling to deploy a Subnet over a large physical domain. The success of network-layer overlays in Cloud DC deployments illustrate the need to decouple the Subnet from the limits of the physical network.

A network-layer overlay is typically a partial or full mesh of point to point tunnels between routers. In case of a full mesh, BUM (Broadcast, Unknown, and multicast) forwarding across the overlay can be implemented as a replication at the ingress router or at a dedicated reflector, but either way entails a growing congestion and latency as the overlay grows in size. BUM forwarding has become so detrimental to the network operations that some operators decide to turn it off. This means that a silent node, whose location is unknown to the fabric control plane and possibly forgotten, cannot be rediscovered by AR procedures, and will not be reachable again until it volunteers its own sign of life. To avoid this, modern protocols should be designed such that the use of overlay-wide broadcasts are limited to operations where a real distribution operation is desired, e.g., every node is interested in receiving the packet.

While a multicast IPv6 ND RA message may be of interest within a site or a subsite to all local nodes, it is probably of little interest to other sites that are served by other routers, even when the overlay spans across both sites. The same goes for a local printer, the broadcast mDNS [RFC6762] lookup should reach local printers but not necessarily faraway ones. This is another indication of the need to decouple the span of the Subnet from the lower layer broadcast domain, and dedicate the broadcast service to local operations such as discovery of really local resources, regardless of the actual span of the Subnet.

As discussed in Section 3.2, multicast IPv6 ND messages are in fact broadcasted across the overlay, meaning that they contribute to the BUM traffic and are treated as full broadcast. Yet, those messages are used for AR and DAD and intend to reach at most one node, the owner of the Target address, if any. Using a BUM transmission that reaches every nodes in the overlay to communicate with at most one is a misuse of the overlay resources and should be replaced by unicast-based methods.

In data centers, overlays are typically combined with server multihoming at the edge. An advanced Network Interface Card (NIC) is equipped with more than one Ethernet ports for redundancy, which connect to different leaf switches (aka ToR) to the same cloud network. The server (say, using Kubernetes) needs a single address and would rather use that address on both ports to the same Subnet, and use either network port for its own reasons, e.g., load balancing, independently of IP addressing considerations. This means that the IP Interface abstraction that the server needs is a logical construct, decoupled from the network ports, and capable to encompass more than one.

3.4. Power and Sustainability

In the wireless case, broadcasts are sent at the slowest speed available, which can be a hundred time slower than unicast transmissions, in order to maximize the chances that all nodes in the BSS will receive the frame. For that relatively long duration, the broadcast transmission holds the spectrum locally, which adds to the latency of pending unicast frames and generates interferences remotely, which may cascade in losses, exponential backoff delays, and retries in adjacent networks. In the process, the broadcast transmission consumes up to a hundred time more power than unicast transmission of equivalent payload size. Power is also wasted when replicating a multicast packet to span an overlay, each time the packet is ultimately dropped by the recipient.

Constrained IoT devices conserve power by placing themselves in deep sleep for most of their time. While a device is sleeping, it cannot answer IPv6 ND messages for AR and DAD. This makes IPv6 ND unsuitable to IoT devices. The 6LoWPAN WG has determined that the most appropriate model for a constrained network is a pull model where the device wakes up, negotiates with its router(s) for addresses and connectivity for some amount of time in the future, and goes back to sleep. The router can then perform a role of sleep proxy that defends the address(es) and holds traffic for the device till the device wakes up and pulls it from the router. Additionally, broadcast operations become rapidly inacceptable in terms of power and bandwidth when the constrained network grows into a mesh. In an LLN, hosts generally do not perform AR for one another, and the not-onlink prefix model is mandatory.

In all cases, to make the Internet greener, we must reconsider the use of broadcast over large access networks. To maintain the capability to build the Subnets we want, the IPv6 architecture must enable to decouple the Subnet from the broadcast domain, make the broadcast domain small and local, and refocus the use of broadcast to the cases where all nodes are interested. Additionally, the IPv6 architecture must enable a model where the network serves and protects low-power devices that sleep most of the time and wake up on their own schedule.

3.5. Security and Privacy

Broadcasting IPv6 ND messages exposes the source and the Target addresses to the whole network, including nodes that do not need to know that those addresses are present in the network. A passive listener may discover the addresses without any observable action or possibility of control by the source. Once it has discovered the presence of neighbor addresses, the onlink attacker can impersonate any host in the network, either by sourcing packets with a stolen address, or by overriding the neighbor caches with a NA that indicates the attacker’s link-layer address.

It is thus desirable to avoid exposing the host IPv6 address in broadcast ND messages. A more private approach has each node see and connect to only a subset of the routers using the not-onlink model. The source may still shortcut to the destination when the destination is effectively on link, based on redirect messages from the router, when desirable. Additionally, it is desirable that only the address owner can source packets with that address, and that another party may neither be able to claim traffic for nor source packets from that address.

The reactive NS lookup method can also be abused from the outside of the network to perform DOS attacks on constrained resources in the network. The attacker just needs to forge random addresses from the Subnet prefix and send one packet to each of these random address. The Subnet ingress router will have to store each of those packets for the duration of the lookup time, and broadcast an NS to lookup each of the forged address. This both locks memory in the router and consumes bandwidth and energy inside the Subnet. All nodes in the Subnet are also impacted, at the minimum to check that they do not own the Target address.

To avoid the lookup delays and associated attacks, it makes sense to use a proactive method whereby the router knows all the addresses present in the network and their link-layer address mapping in advance. When this is achieved, if the destination address of an incoming packet is not present in the router tables, then it can be safely assumed that the destination does not exist in the network and the packet may be dropped by the forwarding engine, e.g., in hardware.

3.6. More Middleboxes

The above problems have been observed at least since the early 2000s, and a number of remediation actions were attempted. The IEEE std 802.11 [IEEE Std. 802.11] mandates the support of a middlebox operation called "ARP proxy" for IPv4 and IPv6 in the Access Point (AP). The "ARP Proxy" cancels broadcasts over a BSS when the IP Target of the ARP/ND message is not owned by a STA associated to the AP. With IPv4, the expectation is that the non-AP station (STA) owns exactly one IP address, and that the address is obtained via DHCP right after the association, so it is simple and deterministic to snoop the address in the DHCP exchange (as long as it remains in the clear), and with that deterministic knowledge of all IPv4 addresses in the BSS, cancel ARP lookups that do not match.

In contrast to IPv4, IPv6 enables a node to form multiple addresses per Link, some of them temporary and with a particular attention paid to privacy. Addresses may be formed and deprecated asynchronously to the association. Even if the knowledge of IPv6 addresses used by a STA can be obtained by snooping protocols such as IPv6 ND and DHCPv6, or by observing data traffic sourced at the STA, such methods provide only an imperfect knowledge of the state of the STA at the AP. This may result in a loss of connectivity for some IPv6 addresses, in particular for addresses rarely used and in a situation of mobility. This may also result in undesirable state persistence in the AP when a STA ceases to use an IPv6 address. It follows that snooping protocols is not a recommended technique and that it should only be used as last resort.

Because IPv6 ND is so easy to attack, some vendors have deployed undocumented proprietary counter measures as middlebox operation in the switches and routers. Those middleboxes snoop the IPv6 ND messages, filter them or modify them, for instance to change their link-layer scope from broadcast to unicast. Based on the snooped information, the middlebox may for instance drop an RA message that appears to be coming from a host (e.g., as inferred because it is received on a wireless adapter), or an NA message coming from a device that does not appear to be the owner. But, for the lack of an explicit contract between the host and the middlebox, the middlebox cannot determine who the real owner is, and it may deny a rightful user.

In a managed network, IP addresses are an expensive and thus a limited resource. To ensure a fair use and protect against DOS attacks, the middlebox may also block Stateless address Autoconfiguration (SLAAC) from a host above a fixed number of addresses. When that happens, the host believes that it can use the address but fails to connect with it. This might happen even if the host has ceased to use other addresses and is now within the allowed quota. IPv6 ND lacks both a method for the host to know how many addresses it can own, and a method for the router to know which addresses the host uses at any point of time. The infrastructure needs a deterministic knowledge of the addresses in use, and for that a contract must be passed between the host and the network to ensure that all the addresses are known and usable.

Maybe the most insidious side effect of those middleboxes is that as opposed to NAT, their operation is obfuscated and proprietary. From one vendor to the next, and from one product generation to the next, their behavior may evolve and affect IPv6 ND in different fashions. In the short term, this may only impact some specific deployments, which may have to work around the issues. In the longer term, this may affect the capability we have to evolve the protocol, like firewalls impact our capability to develop new transports in parallel to TCP and UDP. We must either standardize (e.g., for ND proxy) or eliminate those middlebox activities, and for that, the IPv6 ND protocol must evolve to a model where proprietary middleboxes are not needed anymore. This demands a model that minimizes the use of broadcasts, and where a contract provides mutual guarantees for the hosts that need IPv6 addresses and the routers that provide reachability and protection for these addresses.

3.7. More Operational Issues

"Selectively Isolating Hosts to Prevent Potential Neighbor Discovery Protocol Issues" [ND CONSIDERATIONS] reviews a number of IPv6 ND issues and their possible mitigation in various scenarios. The considered issues include:

1. Impact of multicast on performance and reliability

2. Security exposures from the implicit trust of all on-link nodes

3. On-Demand Neighbor Cache management as a source of performance issues, including exhaustion problems

4. Consequences of the lack of subscriber management capabilities

Table 1 "Which solution solves which issue(s)" of [ND CONSIDERATIONS] indicates that the evaluated issues can be alleviated with the framework detailed in Section 5.

3.8. Summary of Issues

IPv6 ND inherited 2 majors design points from IPv4, a strong coupling of logical and physical concepts, which creates unacceptable constraints on modern deployments with virtual and intangible links, and a reactive operation for AR and DAD that requires an extensive use of broadcast spanning the Subnet. While IPv4 and IPv6 behaviors are similar for addresses obtained via DHCP, the cost of AR and DAD makes IPv6 significantly more expensive than IPv4 when SLAAC is enabled.

And while IPv4 supported NBMA and P2MP models (e.g., on Frame Relay leveraging OSPFv2), the IPv6 promise to support NBMA (for ATM) remains unmet to this day, as only P2P and Transit links are properly supported by IPv6 ND. For those reasons, as well as inherent complexity and unpredictability, IPv6 with SLAAC can be significantly less attractive than IPv4 to some network administrators.

IPv6 ND exposes all addresses to all nodes in the network, which is unfit for privacy. It is prone to DOS attacks from outside the network and impersonation attacks from the inside, with no method to prove the address ownership and perform Source address Validation (SAVI) later on the data traffic. To protect against such threats, the vendors had to introduce middleboxes that interfere with the protocol operation and affect the capability to evolve the protocol in the future.

IPv6 ND lacks a support for mobility (which typically entails a sequence counter maintained by the host and the deprecation of state that is based on older sequences) and for anycast. This makes it very hard for the network to defend the addresses on behalf of the owner, e.g., when the owner is temporarily disconnected. It results that the operation of the middleboxes is unsatisfactory and may cause discontinuities in connectivity.

The extensive use of broadcast operations in IPv6 ND is not only detrimental to bandwidth, it is also an issue for energy conservation and sustainability. Devices must be always attached and always powered on to answer NS messages, which makes IPv6 ND inapplicable to power-conserving devices such as IoT sensors that sleep for the vast majority of their time.

4. IPv6 over Non-Broadcast Networks Architecture

4.1. Basic Concepts

This document introduces an alternate architecture for IPv6 access networks that is designed to apply to the WLANs and LoWPANs types of networks as well as other NBMA networks such as Data-Center overlays and P2MP networks such as IoT radio meshes. It may be used as a replacement to the IPv6 ND reactive model in any network where the issues discussed in Section 3 are detrimental to the network operation.

The key design points in this architecture derive from the original observations made at the 6LoWPAN WG for constrained devices and networks, and focus on avoiding waste of limited resources such as spectrum and energy, by using broadcasts only when broadcast is really needed, and decoupling the IP abstraction of a Subnet from the broadcast domains to avoid Subnet-wide broadcast storms. To that effect, this architecture leverages the not-onlink model and routing inside the Subnet, which enables to form potentially large MLSNs without creating a large broadcast domain at the link layer.

To support the deployment agility that virtual (e.g., VxLAN overlays and pseudowires) and intangible (e.g., wireless, laser, and quantum) links enable, the IP abstractions of Interface, Link, and Subnet are decoupled from their classical physical counterparts of port, link, and broadcast domains. The Subnet is defined by a prefix length called Subnet Prefix Length (SPL), as the longest aggregation that can be advertised in the IGP. An SPL of 64 is typical though the architecture does not mandate it. Host routes and prefixes longer than SPL are advertised inside the Subnet only, using a separate Subnet Gateway protocol (SGP). The SGP is only required in the case of P2MP networks where routers need to relay packets inside the subnet.

Any device that owns an address within the Subnet prefix belongs to the Subnet, this is now decoupled from the physical connectivity and broadcast domain. Instead, the IPv6 routers that serve a Subnet must form a connected dominating set such that every host in the Subnet is connected to at least one router and the routers are connected to one another directly (classical NBMA, aka full mesh) or indirectly via other routers (Point to MultiPoint, P2MP, aka partial mesh). The not-onlink model is used throughout, so hosts do not look each other up, saving all the associated broadcast. Instead, they rely on the routers to forward the packets inside and outside the Subnet. This way, the Subnet can have any structure needed for the deployment, where hosts can move from router to router in the Subnet, or anywhere in the Internet provided they can lay a mobility tunnel to one of the routers for use as IP Link to the Subnet.

All IP Links are abstracted as Point-to-Point, though a lower-layer broadcast service may be used by the router to send RAs to a subset of local hosts in the Subnet, or by the host to send an RS message to a subset of the routers. An IP Interface bundles one or more subInterfaces, one per Subnet that can be reached through that Interface. A Global IPv6 address is installed on the subInterface that connects to the Subnet from which the address derives. The IP Interface connects to one or more IP Links (to different neighbors) over the same or over different physical ports (they are decoupled). A link-local address is associated to the IP Link directly. Each SubInterface connects to the subset of those IP Links that reach other nodes in the Subnet.

In a fashion similar to a IEEE std 802.11 [IEEE Std. 802.11] Association, IPv6 nodes register their addresses to one or more neighbor router(s), which may reject the registration, e.g., in case of a duplication. With the registration, the routers collectively build a complete knowledge of the hosts they serve and in return, hosts obtain guaranteed routing services for their registered addresses for a contractual lifetime.

To support distributed routers in the Subnet, an abstract registrar service maintains the state of all active registrations in the Subnet and answers queries to lookup mappings, validate ownership, and avoid duplications. The registration is abstract to the routing service and the registrar service, and it can be protected to prevent impersonation attacks. The registration enables the network to know deterministically all the IPv6 addresses and link-layer address mapping currently in use, and eliminates the need for lookups and DAD, and for the associated broadcasts.

The abstract routing service allows an ingress router to find a path to the destination address within the Subnet. It can be a simple reflection in a Hub-and-Spoke Subnet that emulates an IEEE Std. 802.11 Infrastructure BSS at the network layer. It can also be a full-fledge routing protocol, e.g., RPL (see [RFC9010]), which is designed to adapt to various LLNs such as WLAN and WPAN radio meshes, or RIFT (see [RFC9692]) or BGP/EVPN (see [RFC8365]), for application in data centers. It can be based on overlay tunnels between ingress router and egress router leveraging a resolver service such as LISP, see [RFC7834] for more. Finally, the routing service can also be an ND proxy that emulates an IEEE Std. 802.11 Infrastructure ESS at the network layer, as specified in the IPv6 Backbone Router [RFC8929].

The abstract registrar service maintains the mapping between the registered node link-layer address and the registered IPv6 address. It contains meta data that enables to ascertain that the second registration for the same address is performed for the same registered node, so it also binds the registered node with the registered IPv6 address. The registrar service provides APIs to look up a link-layer address for an IPv6 address as well as validate IPv6 address ownership. The registrar can be implemented as a mapping server ala LISP [RFC6830], a distributed state ala ND proxy [RFC8929], or a synchronized state ala EVPN [RFC7432]. In the former case, this enables the reactive lookups to be performed as unicast requests to the map resolver. In the latter, the address mapping is synchronized by the routing protocol and known to all the routers for all nodes in the IP Subnet, so there is never a need for a reactive lookup.

On the one hand, the Architecture proposed in this document avoids the use of broadcast operation for DAD and AR, and on the other hand, it supports use cases where Subnet and link-layer domains are not congruent, which is common in wireless networks unless a specific link-layer emulation is provided.

The address registration establishes a contract between the nodes and the routers where nodes can ask for addresses which will be guaranteed to be operational for a contractual lifetime, and the

network may accept or refuse granting additional addresses based on state (e.g., duplicate address) as well as policy (e.g., quota). This ways hosts and routers agree deterministically on which addresses will be served to which nodes in the Subnet. The registration is agnostic to the router to router and router to registrar interfaces. The latter interface can be implemented in various fashions that can blend in existing technologies such as legacy IPv6 ND network through ND proxy, as well as EVPN-based and LISP-based overlays.

4.2. Terminology

4.2.1. IP Links

The term "link" refers to layer 2 (comprising MAC and link layers) communication medium that can be leveraged at layer 3 (aka IP layer, aka network layer) to instantiate one IP hop (see section 2 of [RFC8200]. In this document we conserve that term (lowercase) but differentiate it from an IP Link, which is a network-layer abstraction that somehow represents the link but is not the link, like the map is not the country.

With IPv6, IP has moved to network-layer abstractions for its operations, e.g., with the use of a link-local address (LLA), and that of IP multicast for link-scoped operations. At the same time, the concept of an IP Link emerged as an abstraction that represents how the network layer considers the link:

* An IP Link connects an IP node to one or more other IP nodes using a lower-layer subnetwork. The lower-layer subnetwork may comprise multiple links, e.g., in the case of a switched fabric or a mesh-under LLN.

* an IP Link defines the scope of an LLA, and defines the domain in which the LLA must be unique

* An IP Link is attached to a physical port, and one link-local address is associated to the IP Link.

* an IP Link provides a subset of the connectivity that is offered by the physical link at the lower layer; if the IP Link is narrower than the link-layer reachable domain, then network-layer filters must restrict the link-scoped communication to remain between peers on a same IP Link. More than one IP Link may be installed on the same network port to connect to different peers.

* an IP Link can be Point to Point (P2P), Point to Point (P2MP, forming a partial mesh and non-transitive), NBMA (non-broadcast multi-access, fully meshed), or transit (broadcast-capable and any-to-any).

It is a network design decision to use one IP Link model or another over a given lower-layer subnetwork, e.g., to map a Frame Relay network as a P2MP IP Link, or as a collection of P2P IP Links. As another example, an Ethernet fabric may be bridged, in which case the nodes that interconnect the lower-layer links are L2 switches, and the fabric can be abstracted as a single transit IP Link; or the fabric can be routed, in which case the P2P IP Links are congruent with the link-layer links, and the nodes that interconnect the links are routers.

This architecture only uses P2P Link abstractions as shown in Figure 1, where an IP Link is identified by a pair of local and remote link-layer (MAC) address. A network port may enable to reach to more than one peer at the link layer; in that case, this architecture maps each peer relationship as a different IP Link. A link-local address only needs to be unique within that peer to peer relationship.

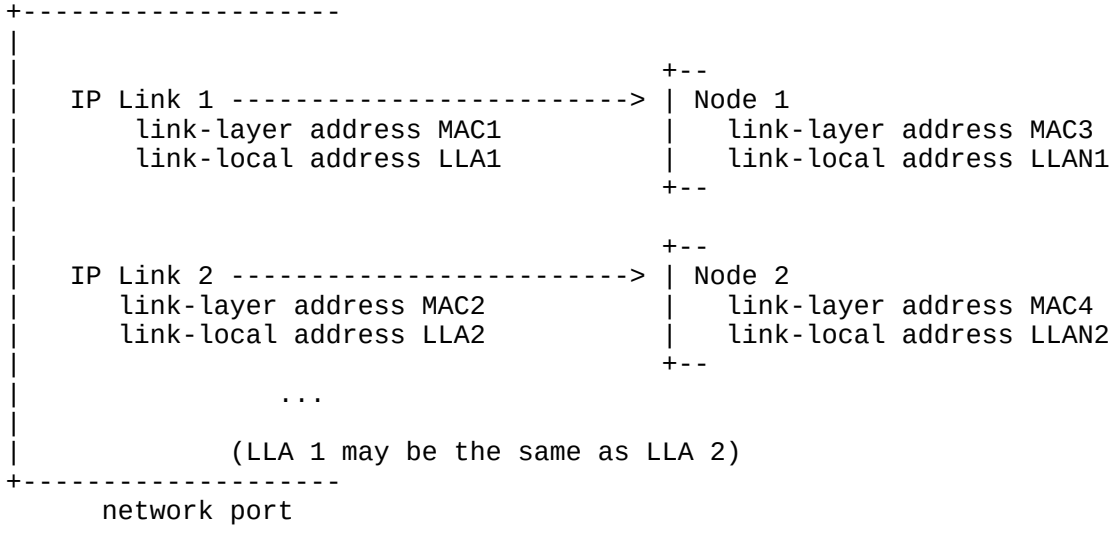

Figure 1: P2P Link Abstraction

If only the network port but not the link-layer address of the peer is visible from the network layer when processing a message, then the network layer cannot discriminate the IP Link of packets arriving on the same network port, and for that reason, it will reject a second

registration for the same link-local address by a second peer, meaning that a link-local address will have to be unique on a network port across IP Links. In that case, the link-local address of the peer is used to identify the IP Link, and all the addresses registered to this node with the same peer link-local address as source will be associated to the same IP Link to that peer.

4.2.2. IP Interfaces

As is the case for links, the term interface has been historically confused between the network port that provides physical connectivity, and the network-layer abstraction that connects the host with the IP Link:

* an IP Interface is an abstraction that connects the host with a collection of IP Links (for the purpose of link-local communication) and bundles the interfaces for each IP Subnet as subInterfaces. The host installs at least one link-local address on an IP Interface for each IP Link that is connected through that Interface, and one subInterface per Subnet. The same link-local address may be reused over different IP Links as long as it is not a collision for the peer on that IP Link. Similarly, the host installs one or more global scope unicast address(es) on an IP subInterface for the associated Subnet, and the address is advertised over each IP Link in the SubInterface.

* an IP Interface can be P2P, in which case it connects to a single IP Link, or P2MP, in which case it aggregates multiple IP Links. In a multihomed host, a single IP Interface can be installed to connect to the IP Links associated to different network ports, in which case the same IPv6 address may be advertised on more than one network port. Conversely, when more than one Subnet is reachable over a network port, more than one IP Interface may leverage that network port for transmission.

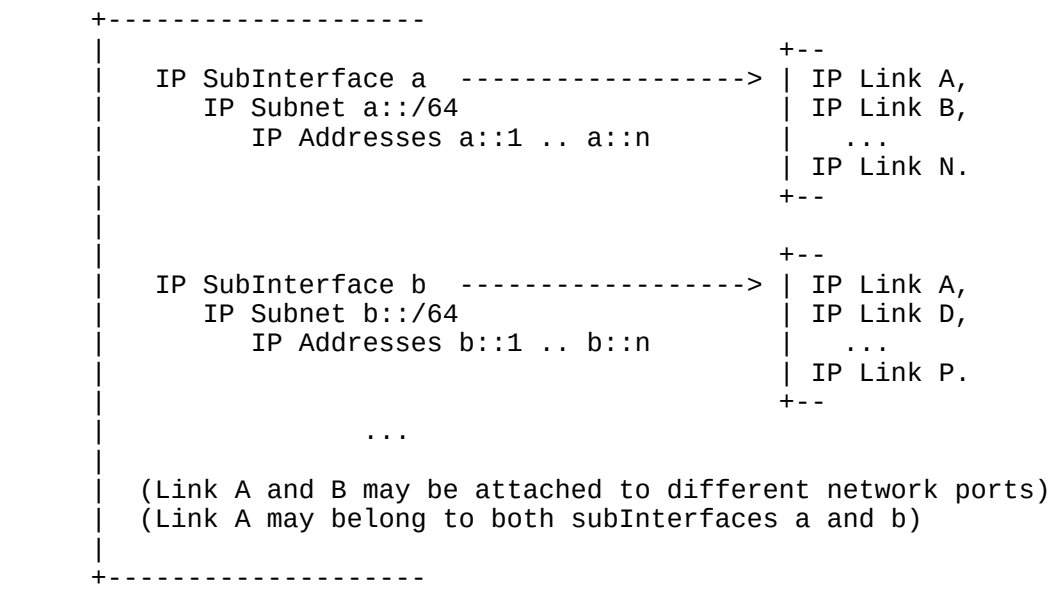

Figure 2: Interface Abstraction

4.2.3. IP Subnets

IPv6 builds another abstraction, the IP Subnet, over one shared IP Link or over a collection IP Links, forming a MLSN in the latter case. An MLSN is formed over IP Links (e.g., P2P or P2MP) that are interconnected by routers that either inject hosts routes in an SGP, in which case the topology can be anything, or perform ND proxy operations, in which case the structure of links must be strictly hierarchical to avoid loops.

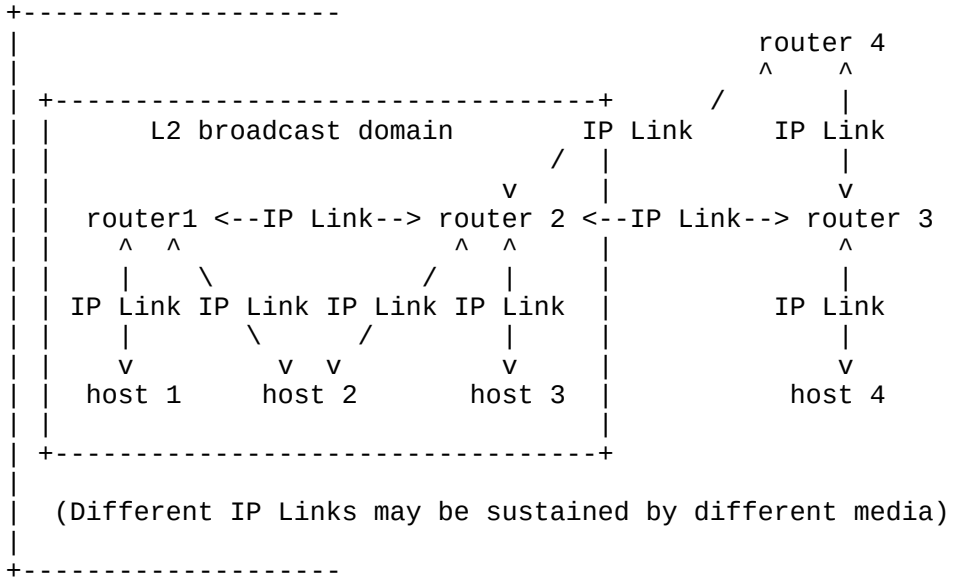

Figure 3: Subnet Abstraction

It is a network design decision to use one IP Subnet model or another over a given lower-layer network. A switched fabric can host one or more IP subnets, in which case the IP Links can reach all and beyond one Subnet. On the other hand, a Subnet can encompass a collection of links; in that case, the scope of the link-local addresses, which is the IP Link, is narrower than the span of the Subnet.

A Subnet prefix is associated with the IP Subnet, and a node is a member of an IP Subnet when it has an IP address that derives from that prefix. The IP address has Global Unicast scope (in the formal sense of [RFC4291]), and, as opposed to link-local Addresses, the scope of the address is not limited to the IP Link.

The switched and routed fabric above could be the exact same network of physical links and boxes, what changes is the way the networking abstractions are mapped onto the system, and the implication of such decision include the capability to reach another node at the link layer, and the size of the broadcast domain and related broadcast storms.

4.2.4. ND Proxies

[RFC8929] defines bridging and routing IPv6 ND proxies for registering nodes / registered addresses. Both forms of ND proxies interconnect IP Links and enable to isolate the link-layer broadcast domains. But in the case of a bridging proxy, the link-layer unicast communication can still exist between the link-layer domains that are covered by network-layer links, whereas in the base of a routing proxy, they are isolated, and packets must be routed back and forth. Bridging proxies are possible between compatible technologies and translational bridges (e.g., Wi-Fi to Ethernet), whereas routing proxies are required between non-bridgeable technologies and desirable to avoid exposing the link-layer addresses across, e.g., for reasons of stability and scalability.

ND proxies can also serve IPv6 nodes that still rely on IPv6 ND in a coexistence scenario. The ND proxy intercepts (snoops) the multicast NS messages from the nodes and, in case or AR or DAD, polls the registrar to lookup whether an active mapping exists for the Target. When that is the case, the ND proxy may forward the NS message as a link-layer unicast to the node that owns the binding, else it may either drop the multicast or broadcast it at the link layer. Once the node formed an address, the ND proxy fills the registrar to associate the IPv6 address with the node. The method is brittle, since there is no contract with the node to guarantee the ownership, no "contract", as discussed in Section 3.6, so for those addresses, the registrar may be inaccurate.

4.2.5. Subnet Gateway Protocols

The SPL boundary creates a wall between the traditional Interior Gateway Protocols (IGP) that operate between Subnets and manipulate shorter than SPL prefixes, and Subnet Gateway Protocols (SGP) that operate inside a Subnet and manipulate longer than SPL prefixes, typically /128 host routes, and possibly more specific data like link-layer address mappings and address Proof of Ownership.

As opposed to classical IGPs, an SGP must support rapid mobility of addresses to cope with wireless devices and virtual machines mobility. In that regard, an SGP operates mores as a MANET protocol than as a classical IGP. Ideally, there should be no stale route, and no microloop. A classical method in MANETs to achieve this is to sequence the movements and advertise the sequence in the routing protocol, so only routes with the most recent sequence can be followed, and once a packet starts following a route with a certain sequence, it must be discarded rather to have to follow a path with an older sequence. To support this approach, the node that registers an address must be the owner of a mobility sequence number and update that sequence when it moves.

Multihoming being a classical requirement in DC environments, the SGP must be able to differentiate not only address duplication from movement, but also from anycast addresses, which can be advertised from multiple places in a coordinated (same mobility sequence) or uncoordinated fashion. For unicast addresses, an token that identifies the address owner can be used for address duplication avoidance, and if that token is cryptographic, it can be used as registration ownership verifier as well.

4.3. IP Models

4.3.1. Physical Broadcast Domain

At the physical (PHY) layer, a node’s broadcast domain is the set of nodes that may receive a transmission that the node sends over a network port, for instance the set of nodes in range of the radio transmission. This set can comprise a single peer on a serial cable used as point-to-point link. It may also comprise multiple peer nodes on a broadcast radio or a shared physical resource such as the Ethernet wires and hubs for which IPv6 ND was initially designed.

On WLAN and LoWPAN radios, the physical broadcast domain is defined relative to a particular transmitter, as the set of nodes that can receive what this transmitter is sending. Literally every frame defines its own broadcast domain since the chances of reception of a given frame are statistical. In average and in stable conditions, the broadcast domain of a particular node can be still be seen as mostly constant and can be used to define a closure of nodes on which an upper-layer abstraction can be built.

A physical-layer communication can be established between two nodes if the physical broadcast domains of their unicast transmissions include one another. On WLAN and LoWPAN radios, that relation is usually not reflexive, since nodes disable the reception when they transmit; still, they may retain a copy of the transmitted frame, so

it can be seen as reflexive at the MAC layer. It is often symmetric, meaning that if B can receive a frame from A, then A can receive a frame from B. But there can be asymmetries due to power levels, interferers near one of the receivers, or differences in the quality of the hardware (e.g., crystals, PAs and antennas) that may affect the balance to the point that the connectivity becomes mostly uni-directional, e.g., A to B but practically not B to A.

It takes a particular effort to place a set of devices in a fashion that all their physical broadcast domains fully overlap, and that specific situation cannot be assumed in the general case. In other words, the relation of radio connectivity is generally not transitive, meaning that A in range of B and B in range of C does not necessarily imply that A is in range of C.

4.3.2. Link-layer Broadcast Emulations

We call Direct MAC Broadcast (DMB) the transmission mode where the broadcast domain that is usable at the MAC layer is directly the physical broadcast domain. IEEE Std. 802.15.4 [IEEE Std. 802.15.4] and IEEE Std. 802.11 [IEEE Std. 802.11] OCB (for Out of the Context of a BSS) are examples of DMB radios. DMB networks provide mostly symmetric and non-transitive transmission. This contrasts with a number of link-layer Broadcast Emulation (LLBE) schemes that are described in this section.

In the case of Ethernet, while a physical broadcast domain is constrained to a single shared wire, the IEEE Std. 802.1 [IEEE Std. 802.1] bridging function emulates the broadcast properties of that wire over a whole physical mesh of Ethernet links. For the upper layer, the qualities of the shared wire are essentially conserved, with a reliable and cheap broadcast operation over a transitive closure of nodes defined by their connectivity to the emulated wire.

In large switched fabrics, overlay techniques enable a limited connectivity between nodes that are known to a Map Resolver. The emulated broadcast domain is configured to the system, e.g., with a VXLAN network identifier (VNID). Broadcast operations on the overlay can be emulated but can become very expensive, and it makes sense to proactively install the relevant state in the mapping server as opposed to rely on reactive broadcast lookups to do so.

An IEEE Std. 802.11 [IEEE Std. 802.11] Infrastructure Basic Service Set (BSS) also provides a transitive closure of nodes as defined by the broadcast domain of a central AP. The AP relays both unicast and broadcast packets and provides the symmetric and transitive emulation of a shared wire between the associated nodes, with the capability to signal link-up/link-down to the upper layer. Within a BSS, the

physical broadcast domain of the AP serves as emulated broadcast domain for all the nodes that are associated to the AP. Broadcast packets are relayed by the AP and are not acknowledged. To increase the chances that all nodes in the BSS receive the broadcast transmission, AP transmits at the slowest PHY speed. This translates into maximum co-channel interferences for others and the longest occupancy of the medium, for a duration that can be a hundred times that of the unicast transmission of a frame of the same size.

For that reason, upper-layer protocols (ULPs) should tend to avoid the use of broadcast when operating over IEEE std 802.11 [IEEE Std. 802.11] as they already typically do over IEEE std 802.15.4 [IEEE Std. 802.15.4]. To cope with these problems, APs may implement strategies such as turn a broadcast into a series of unicast transmissions, or drop the message altogether, which may impact the upper-layer protocols. For instance, some APs may not copy Router Solicitation (RS) messages under the assumption that there is no router across the wireless network. This assumption may be correct at some point of time and may become incorrect in the future. Another strategy used in Wi-Fi APS is to proxy protocols that heavily rely on broadcast, such as the address Resolution in ARP and IPv6 ND, and either respond on behalf or preferably forward the broadcast frame as a unicast to the intended Target.

In an IEEE Std. 802.11 [IEEE Std. 802.11] Infrastructure Extended Service Set (ESS), infrastructure BSSes are interconnected by a bridged network, typically running Transparent Bridging and the Spanning tree Protocol or a more advanced link-layer Routing (L2R) scheme. In the original model of learning bridges, the forwarding state is set by observing the source MAC address of the frames. When a state is missing for a destination MAC address, the frame is broadcasted with the expectation that the response will populate the state on the reverse path. This is a reactive operation, meaning that the state is populated reactively to the need to reach a destination. It is also possible in the original model to broadcast a gratuitous frame to advertise self throughout the bridged network, and that is also a broadcast.

The process of the Wi-Fi association prepares a bridging state proactively at the AP, which avoids the need for a reactive broadcast lookup over the wireless access. In an ESS, the AP may also generate a gratuitous broadcast sourced at the MAC address of the STA to prepare or update the state in the learning bridges so they point towards the AP for the MAC address of the STA. This framework emulates that proactive method at the network layer for the operations of AR, DAD and ND proxy.

In some instances of WLANs and LoWPANs, a Mesh-Under technology (e.g., a IEEE Std. 802.11s or IEEE Std. 802.15.10) provides meshing services that are similar to bridging, and the broadcast domain is well-defined by the membership of the mesh. Mesh-Under emulates a broadcast domain by flooding the broadcast packets at the link layer. When operating on a single frequency, this operation is known to interfere with itself, and requires inter-frame gaps to dampen the collisions, which reduces further the amount of available bandwidth.

As the cost of broadcast transmissions becomes increasingly expensive, there is a push to rethink the upper-layer protocols to reduce the dependency on broadcast operations.

4.3.3. Mapping the IP Link Abstraction

As introduced in Section 4.2.1, IPv6 defines a concept of IP Link, link scope and link-local Addresses (LLA), an LLA being unique and usable only within the scope of an IP Link. The IPv6 ND [RFC4861] DAD [RFC4862] process uses a multicast transmission to detect a duplicate address, which requires that the owner of the address is connected to the link-layer broadcast domain of the sender.

On a wired medium, the IP Link is often confused with the physical broadcast domain because both are determined by the serial cable or the Ethernet shared wire. Ethernet Bridging reinforces that illusion with a link-layer broadcast domain that emulates a physical broadcast domain over the mesh of wires. But the difference shows on legacy P2MP and NBMA networks such as ATM and Frame-Relay, on shared links, and on newer types of NBMA networks such as radio and composite radio-wires networks. It also shows when private VLANs or link-layer cryptography restrict the capability to read a frame to a subset of the connected nodes.

In Mesh-Under and Infrastructure BSS, the IP Link extends beyond the physical broadcast domain to the emulated link-layer broadcast domain. Relying on multicast for the ND operation remains feasible but becomes highly detrimental to the unicast traffic, and becomes less and less energy-efficient and reliable as the network grows.

On DMB radios, IP Links between peers come and go as the individual physical broadcast domains of the transmitters meet and overlap. The DAD operation cannot provide once and for all guarantees over the broadcast domain defined by one radio transmitter if that transmitter keeps meeting new peers on the go.

The scope on which the uniqueness of an LLA must be checked is each new pair of nodes for the duration of their conversation. As long as there’s no conflict, a node may use the same LLA with multiple peers

but it has to perform DAD again with each new peer. A node may need to form a new LLA to talk to a new peer, and multiple LLAs may be present in the same radio network to talk to different peers. In this framework, each pair of nodes defines a P2P IP Link, and define the domain where an LLA must be unique.

The DAD and AR procedures in IPv6 ND expect that a node in a Subnet is reachable within the broadcast domain of any other node in the Subnet when that other node attempts to form an address that would be a duplicate or attempts to resolve the MAC address of this node. This is why ND is applicable for P2P and transit links, but requires extensions for more complex topologies.

4.3.4. Mapping the IPv6 Subnet Abstraction

As introduced in Section 4.2.3, IPv6 also defines the concept of a IP Subnet for IPv6 unicast addresses with a global scope, Global and Unique Local Addresses (GUA and ULA). All the addresses in the same Subnet share the same prefix, and by extension, a node belongs to an IP Subnet if it has an address that derives from the prefix of the Subnet. That address must be topologically correct, meaning that it must be installed on a sub-Interface that connects to the Subnet, for use with routers that expose the Subnet in their RA messages (see [RFC5942]).

Unless intently replicated in different locations for very specific purposes, a Subnet prefix is unique within a routing system; for ULAs, the routing system is typically a limited domain, whereas for GUAs, it is the whole Internet.

For that reason, it is sufficient to validate that an address that is formed from a Subnet prefix is unique within the scope of that Subnet to guarantee that it is globally unique within the whole routing system. Note that a Subnet may become partitioned due to the loss of a wired or wireless link, so even that operation is not necessarily obvious, more in [DAD APPROACHES].

The IPv6 aggregation model relies on the property that a packet from the outside of a Subnet can be routed to any router that belongs to the Subnet, and that this router will be able to either resolve the destination link-layer address and deliver the packet, or, in the case of an MLSN, route the packet to the destination within the Subnet.

If the Subnet is known as on-link, then any node may also resolve the destination link-layer address and deliver the packet, but if the Subnet is not on-link, then a host in the Subnet that does not have a Neighbor Cache Entry (NCE) for the destination will also need to pass the packet to a router, more in [RFC5942].

On Ethernet, an IP Subnet is often congruent with an IP Link because both are determined by the physical attachment to a shared wire or an IEEE Std. 802.1 bridged domain. In that case, the connectivity over the IP Link is both symmetric and transitive, the Subnet can appear as on-link, and any node can resolve a destination MAC address of any other node directly using IPv6 ND.

But an IP Link and an IP Subnet are not always congruent. In the case of a Shared Link, individual subnets may each encompass only a subset of the nodes connected to the link. Conversely, in Route-Over Multi-link subnets (MLSN) [RFC4903], routers federate the links between nodes that belong to the Subnet, the Subnet is not on-link and it extends beyond any of the federated links.

4.4. Subnet Neighbor Discovery and Routing

This Architecture defines a new operation for ND that is based on 2 major paradigm changes, a proactive address registration by hosts to their attachment routers and routing to host routes (/128) within the Subnet. This allows ND to avoid the expectations of transit links and Subnet-wide broadcast domains.

The proactive address registration, called Stateful address Autoconfiguration (SFAAC) by opposition to SLAAC, is agnostic to the method used for address Assignment, e.g., Manual, Semantically Opaque Autoconfiguration [RFC7217], randomized [RFC8981], or DHCPv6 [RFC8415]. It does not change the IPv6 addressing [RFC4291] or the current practices of assigning prefixes, with typically a SPL of 64, to a Subnet. But the DAD operation is performed as a unicast exchange with the abstract registrar service.

This Architecture combines SFAAC with the not-onlink model on the IP Interfaces. Hosts do not expect the IP Subnet to be reachable over the L2 broadcast domain and rely on their routers to forward the packets inside and outside the Subnet. In turn, the routers expose to each other all the IPv6 addresses that are either owned or registered to it as host routes over a Subnet Gateway Protocol, a routing protocol that is specialized in routing inside the Subnet and can be decoupled with the IGP, that is the routing protocol used between subnets.

5. A Framework for Address and Prefix Registration with Subnet Routing

5.1. Stateful address Autoconfiguration

Stateful address Autoconfiguration (SFAAC) was initially standardized for IoT and wireless links as [RFC6775], [RFC8505], [RFC8928], [I-D.ietf-6lo-updating-rfc-8928], and [RFC9685]. The central operation in SFAAC is a new Address Registration mechanism that allows for SFAAC but is not limited to autoconfigured addresses. Indeed, all addresses that the node has should be registered to its router(s). When the node operates as a router, e.g., for a stub or a virtual network, it can even register a full prefix leveraging with the extension specified in [PREFIX REGISTRATION].

To enable Address Registration, a new option in NS/NA messages, the Extended address Registration Option (EARO) signals that the Target address is being registered and provides the registration parameters [RFC8505]. This method allows to prepare and maintain the host routes in the routers and avoids the reactive address Resolution in IPv6 ND and the associated link-layer broadcast transmissions. [RFC8928] adds the capability to verify in future Registrations the ownership of the address or prefix being registered for the first time. This verification relies on a cryptographic token called a Registration Ownership Verifier (ROVR) that is associated with the address at the first registration and persisted in the network. The ROVR derives from a keypair that can also be autoconfigured by the host, and does not require a Public Key Infrastructure (PKI).

The EARO provides information to the router that is agnostic to both the way the address (or prefix) is obtained (e.g., SFAAC or DHCPv6) and the routing operation in the Subnet. This allows to effectively decouple the host operation from the router operation. Routing can take multiple forms, from an SGP to a collapsed Hub-and-Spoke model where only one router owns and advertises the prefix. [RFC8505] is already referenced as the registration interface to "RIFT: Routing in Fat Trees" [RFC9692] and "RPL: IPv6 Routing Protocol for Low-Power and Lossy Networks" [RFC6550] with [RFC9010].

Subnet ND (SND) instantiates the Architecture presented in Section 4; it combines SFAAC with a Backbone Router (6BBR) ND proxy function (more in [RFC8929]) operating as a network-layer Access Point. Multiple 6BBRs placed along the wireless edge of a Backbone link handle IPv6 Neighbor Discovery and forward packets over the backbone on behalf of the registered nodes on the wireless edge. This enables to span a Subnet over an MLSN that federates edge wireless links with a high-speed, typically Ethernet, backbone (as a network-layer ESS). The ND proxy maintains the reachability for Global Unicast and link-local Addresses within the federated MLSN, either as a routing proxy

where it replies with its own MAC address or as a bridging proxy that typically forwards the multicast ND messages as unicast link-layer frames to their target. The wireless nodes can form any address they want and move freely from a wireless edge link to another, without renumbering. In that case, the registrar is distributed between the 6BBR, each 6BBR maintaining only a state for the subset of the addresses that were registered to it and for which it is authoritative. When the 6BBR is not currently authoritative for a new address being registered to it, it relies on IPv6 ND that is used reactively over the backbone to obtain an existing registration state in the disaggregated registrar that the 6BBRs form collectively.

This framework allows other implementations of the abstract concept of the registrar. For instance, [EVPN-SFAAC] allows to distribute the registrar in every router, and leverages EVPN as the method to synchronize the registrar state between routers. In that case, BGP acts both as the SGP to announce the reachability of the addresses and as the synchronization protocol between the distributed registrar. All the routers know proactively the mapping for all the addresses, and there is no need for a reactive lookup as is the case for SND. As another example, a Locator/ID Separation Protocol (LISP) Map-Resolver [RFC6830] could support the EDAR/EDAC exchange either directly or via a proxy, and serve as registrar.

The framework allows for mixed environments with registrations and IPv6 ND, using [RFC8929] to perform ND proxy operations on behalf of registered address and respond to DAD and lookups from legacy nodes, and prevent registering nodes from autoconfiguring addresses that exist in legacy nodes by performing DAD on behalf of the registering nodes, more in Section 7.

5.2. Links and Link-Local Addresses

Link-local Addresses are typically autoconfigured, though it it is possible to set them some other ways such as configuration. With this Architecture, SFAAC is the recommended way to form a link-local address using a P2P Link abstraction.in that case, DAD is performed between communicating pairs of nodes and an NCE can be populated on both sides with a single unicast exchange, as shown in Figure 4. In the case of a bridging proxies, though, the link-local traffic is bridged over the backbone and the DAD must proxied there as well.

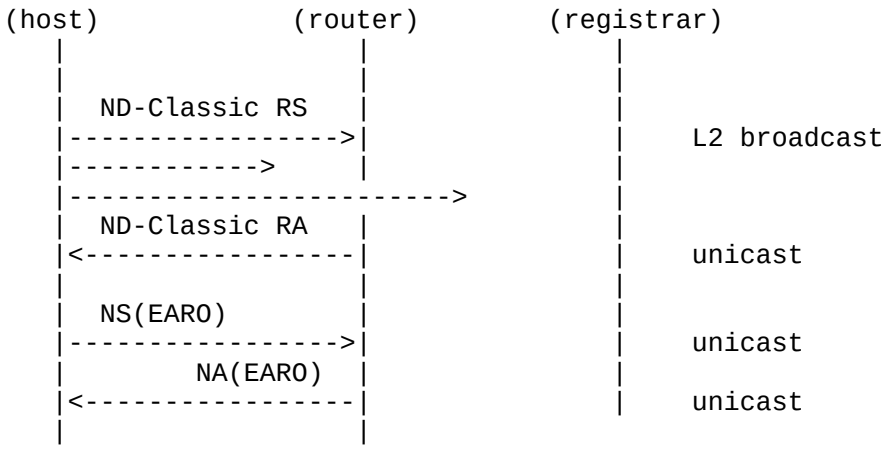

Figure 4: Registration for a Link-Local Address

For instance, in the case of Bluetooth(R) Low Energy (BLE) [RFC7668][IEEEstd802151], the uniqueness of link-local Addresses needs only to be verified between the pair of communicating nodes, the central router and the peripheral host. In that example, 2 peripheral hosts connected to the same central router cannot have the same link-local address because the addresses would collision at the central router which could not talk to both over the same network port, unless it can separate the IP Links, e.g., based on the remote MAC address. The DAD operation from SFAAC is appropriate for that use case, but the one from ND is not, because the peripheral hosts are not on the same broadcast domain.

On the other hand, the uniqueness of GUAs and ULAs is validated at the Subnet Level, using a logical registrar that is global to the Subnet.

5.3. Subnets and Global Addresses

As opposed to Link-Local Addresses that are typically autoconfigured, Subnets and Global Addresses may be obtained via DHCPv6 [RFC8415]. In that case, it makes sense to use a ROVR [RFC8928] as device ID (DUID) in the DHCP exchange as illustrated in Figure 5.

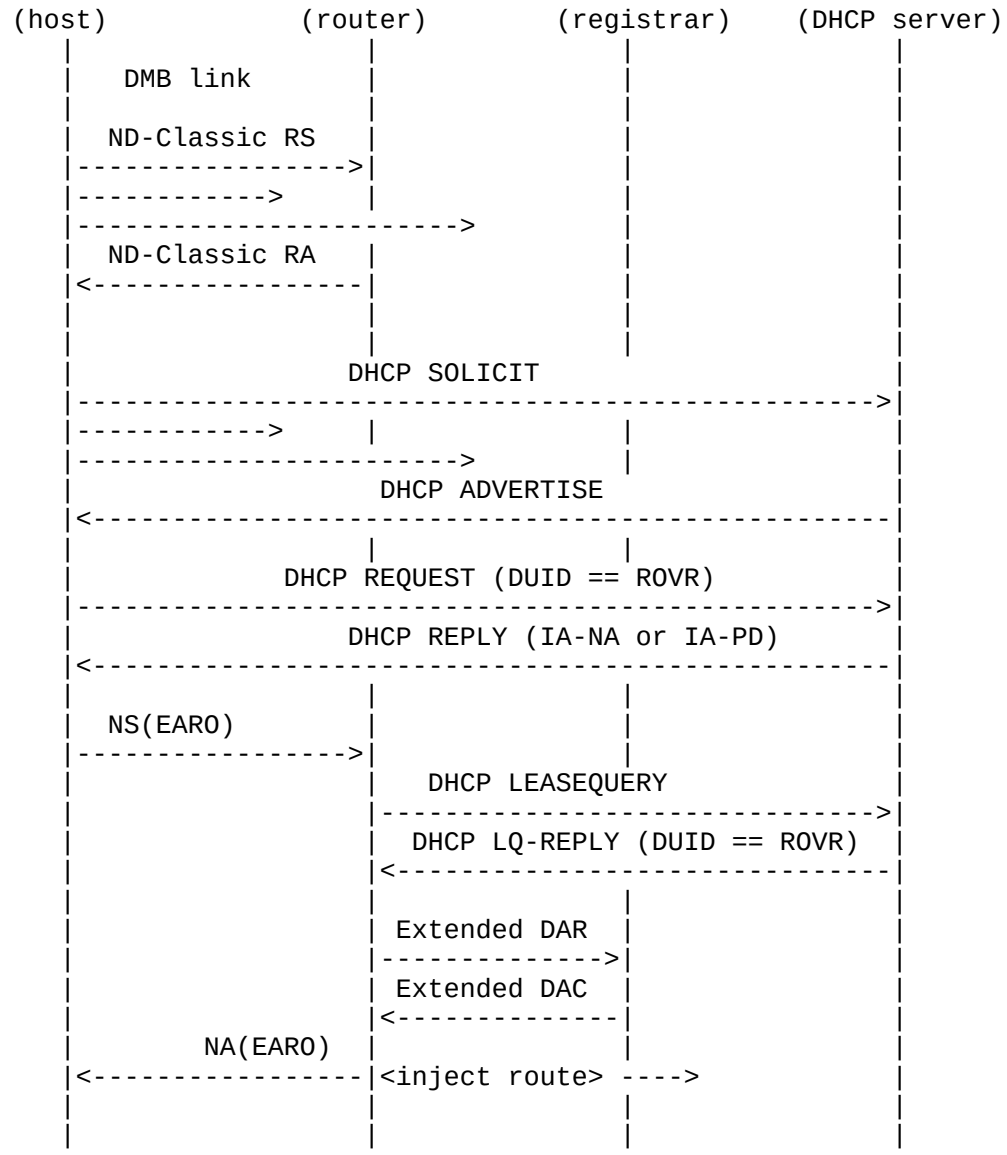

Figure 5: Registration for an Address or Prefix obtained via DHCPv6

5.4. Anycast and Multicast Addresses

While IPv6 ND was initially defined for unicast addresses only, [RFC9685] extends [RFC8505] for anycast and multicast IPv6 addresses as well. Though RPL [RFC6550], which is extended in that document, is the SGP of choice in a Low-power Lossy Network (LLN), the registration is agnostic to the SGP and the same model applies to any SGP that is capable of advertising multicast and/or anycast addresses as well as unicast.

[RFC9685] can be used as a replacement for MLD [RFC3810] for use cases where broadcast are not desirable, and when a device push model such as SFAAC is preferred over a network pull such as MLD and IPv6 ND. With [RFC8505], the host does not need to define SNMAs for its unicast addresses and does not perform the associated MLDv2 operation. With [RFC9685], MLDv2 and its extensive use of broadcast can be totally eliminated.

In the case of anycast, the signal enables the 6BBRs to accept more than one registration for the same address, and collectively elect the registering host receives a packet for a given anycast address.

5.5. Hub and Spoke Networks

Address and Prefix Registration extends IPv6 ND for Hub-and-Spoke (e.g., BLE) and Route-Over (e.g., RPL) Multi-link subnets (MLSNs).

In the Hub-and-Spoke case, each Hub-Spoke pair is a distinct IP Link, and a Subnet can be mapped on a collection of links that are connected to the Hub. The Subnet prefix is associated to the Hub.

Acting as a router, the Hub advertises the prefix as not-on-link to the spokes in RA messages Prefix Information Options (PIO). Acting as hosts, the Spokes autoconfigure addresses from that prefix and register them to the Hub with a corresponding lifetime.

Acting as a registrar, the Hub maintains a binding table of all the registered IP addresses and rejects duplicate registrations, thus ensuring a DAD protection for a registered address even if the registering node is sleeping.

The Hub also maintains an NCE for the registered addresses and can deliver a packet to any of them during their respective lifetimes. It can be observed that this design builds a form of network-layer Infrastructure BSS.

A Route-Over MLSN is considered as a collection of Hub-and-Spoke where the Hubs form a connected dominating set of the member nodes of the Subnet, and IPv6 routing takes place between the Hubs within the Subnet. A single logical registrar is deployed to serve the whole mesh.

The registration in [RFC8505] is abstract to the routing protocol and provides enough information to feed a routing protocol such as RPL as specified in [RFC9010]. In a degraded mode, all the Hubs are connected to a same high speed backbone such as an Ethernet bridging domain where IPv6 ND is operated. In that case, it is possible to federate the Hub, Spoke and Backbone nodes as a single Subnet, operating ND proxy operations [RFC8929] at the Hubs, acting as 6BBRs. It can be observed that this latter design builds a form of network-layer Infrastructure ESS.

5.6. P2MP Networks

In the case of P2MP networks, there is not necessarily a direct IP Link between a source and a destination that are in the same subnet. The expectation is that routers running a common SGP form a connected dominating set for the IP Subnet, meaning that every router can reach every other router directly or via some other routers, and that every host that has addresses in the IP Subnet is connected to at least one router that participates to the SGP.

The connected assumption allows any host to reach any other node in the subnet via one or more SGP routers, relying on the SGP alone for the forwarding decision inside the MLSN. Outreaching gateways for the subnet also inject a default route and more specific routes to external destinations.

In the resulting topology, IPv6 forms P2P IP Links between peer routers and hosts using [RFC8505]. A P2P IP Link is established between 2 nodes when one registers its link local address to the other. The procedure for registering a link local address only requires that the link local address of each peer is unique from the perspective of the other peer, so it does not require any broadcast or third party validation. The registration is typically triggered when receiving an RA messages that indicates a link local address for the router, and advertises that the router supports address registrations by [RFC8505].

Global unicast, anycast and multicast addresses are advertised by hosts to routers using [RFC8505], and propagated router to router using the SGP under the control of the host. A special "R" flag in the EARO indicates that the host expects the router to provide return reachability. Upon that flag, the router advertises the registered address in the SGP, with a lifetime that is in the same order as that the router has accepted for the registration.

Depending on its scope [RFC7346], a multicast message may reach up to different boundaries in the Subnet. Admin-Local scope reaches all nodes in the Subnet and leverages the SGP. Realm-Local scope reaches all the peers in a collection of IP Links, in which case ingress replication and MAC unicast transmission are used. Link-Local scope reaches all nodes in the broadcast domain of an IP interface, in which case MAC-level broadcast may be used. For instance, an RA message that is sent over the broadcast domain of an IP interface with a Link-Local scope and it will reach beyond the known IP Links over that interface. This is required because the RA message can be the trigger for the creation of the IP Link.

When a node roams outside the subnet, meaning that it is physically disconnected from all the routers is the IP Subnet, it may open a tunnel to one anchor router in the IP Subnet

5.7. Advertising Prefixes

By definition, prefixes longer than SPL are inside a Subnet and do not leak outside the SGP. Still, it is valid for a node to register a prefix of any size longer than SPL, and for the router to advertise the registered prefix in the SGP. This can be useful for instance to expose a /96 Prefix that is used to transport IPv4 mapped traffic [RFC6052].

In a number of situation (e.g., [I-D.ietf-v6ops-dhcp-pd-per-device] and [I-D.ietf-snac-simple]) it becomes desirable for a node to advertise that it owns a prefix to neighbor routers, and that packets sourced at (in case of multihoming) or destinated to (in case of a stub) that prefix should be passed to this node. At the time of this writing, the prefix is typically delegated by DHCP, and proprietary code ties the DHCP server that delegates a prefix, or a DHCP relay, and the routers that install a route for the delegated prefix via the delegated node. it is desirable to provide a standard way for the delegated node to advertise the prefix to the neighbor routers and avoid proprietary extensions.

[PREFIX REGISTRATION] extends [RFC8505] to enable a node that owns or is directly connected to a Prefix to register that Prefix to neighbor routers. The registration indicates that the registered Prefix can

be reached via the advertising node without a loop. The prefix registration also provides a protocol-independent interface for the node to request the router to redistribute the prefix in the SGP.

6. SND Applicability

SND applies equally to physical links that are P2P, transit, P2MP Hub-and-Spoke, to links that provide link-layer Broadcast Domain Emulation such as Mesh-Under and Wi-Fi BSS, and to Route-Over meshes. In either cases, the IP Link abstraction in SND is always P2P.

There is an intersection where The IP Link and the IP Subnet are congruent and where both ND and SND could apply. These includes P2P, the MAC emulation of a PHY broadcast domain, and the particular case of always on, fully overlapping physical radio broadcast domain. But even in those cases where both are possible, SND is preferable vs. ND because it reduces the need of broadcast; for more details, see the introduction of [RFC8929].

There are also a number of practical use cases in the wireless world where links and subnets are not congruent:

* The IEEE Std. 802.11 infrastructure BSS enables one Subnet per AP, and emulates a broadcast domain at the link layer. The Infrastructure ESS extends that model over a backbone and recommends the use of an ND proxy [IEEE Std. 802.11] to interoperate with Ethernet-connected nodes. SND incorporates an ND proxy to serve that need, which was missing so far, , more in Section 6.2

* Bluetooth is Hub-and-Spoke at the link layer. It would make little sense to configure a different Subnet between the central and each individual peripheral node (e.g., sensor). Rather, [RFC7668] allocates a prefix to the central node acting as router, and each peripheral host (acting as a host) forms one or more address(es) from that same prefix and registers it, more in Section 6.4.3.

* A typical SmartGrid networks puts together Route-Over MLSNs that comprise thousands of IPv6 nodes. The 6TiSCH architecture [RFC9030] presents the Route-Over model over an IEEE Std. 802.15.4 Time-Slotted Channel-Hopping (TSCH) [IEEE Std. 802.15.4] mesh, and generalizes it for multiple other applications, more in Section 6.4.4.

  Each node in a SmartGrid network may have tens to a hundred others nodes in range. A key problem for the routing protocol is which other node(s) should this node peer with, because most of the

possible peers do not provide added routing value. When both energy and bandwidth are constrained, talking to them is a waste of resources and most of the possible P2P links are not even used. Peerings that are actually used come and go with the dynamics of radio signal propagation. It results that allocating prefixes to all the possible P2P links and maintain as many addresses in all nodes is not even considered.

6.1. Case of LPWANs

LPWANs are by nature so constrained that the addresses and subnets are fully pre-configured and operate as P2P or Hub-and-Spoke. This saves the steps of neighbor Discovery and enables a very efficient stateful compression of the IPv6 header. So neither IPv6 ND nor SND is really used in that space.

6.2. Case of Infrastructure IEEE std 802.11 BSS and ESS

In contrast to IPv4, IPv6 enables a node to form multiple addresses, some of them temporary to elusive, and with a particular attention paid to privacy. Addresses may be formed and deprecated asynchronously to the association.

Snooping protocols such as IPv6 ND and DHCPv6 and observing data traffic sourced at the STA provides an imperfect knowledge of the state of the STA at the AP. Missing a state or a transition may result in the loss of connectivity for some of the addresses, in particular for an address that is rarely used, belongs to a sleeping node, or one in a situation of mobility. This may also result in undesirable remanent state in the AP when the STA ceases to use an IPv6 address while remaining associated. It results that snooping protocols is not a recommended technique and that it should only be used as last resort, when the SND registration is not available to populate the state.

The recommended alternative method is to use the SND Registration for IPv6 Addresses. This way, the AP exposes its capability to proxy ND to the STA in Router Advertisement messages. In turn, the STA may request proxy ND services from the AP for all of its IPv6 addresses, using the Extended address Registration Option, which provides the following elements:

* The registration state has a lifetime that limits unwanted state remanence in the network.

* The registration is optionally secured using [RFC8928] to prevent address theft and impersonation.

* The registration carries a sequence number, which enables to figure the order of events in a fast mobility scenario without loss of connectivity.

The ESS mode requires a "ARP-Proxy" operation at the AP. This includes a proxy ND operation that must cover Duplicate Address Detection, Neighbor Unreachability Detection, Address Resolution and Address Mobility to transfer a role of ND proxy to the AP where a STA is associated following the mobility of the STA. New text in the 802.11me revision (WIP) encourages to turn link-layer broadcast into unicast and let the STA respond for itself, as opposed to responding on behalf, unless operating as a sleep proxy.

The SND proxy ND specification that associated to the address Registration is [RFC8929]. With that specification, the AP participates to the protocol as a Backbone Router, typically operating as a bridging proxy though the routing proxy operation is also possible. As a bridging proxy, the backbone router either replies to NS lookups with the MAC address of the STA, or preferably forwards the lookups to the STA as link-layer unicast frames to let the STA answer. For the data plane, the backbone router acts as a normal AP and bridges the packets to the STA as usual. As a routing proxy, the backbone router replies with its own MAC address and then routes to the STA at the network layer. The routing proxy reduces the need to expose the MAC address of the STA on the wired side, for a better stability and scalability of the bridged fabric.

6.3. Case of Mesh Under Technologies

The Mesh-Under provides a broadcast domain emulation with symmetric and Transitive properties and defines a transit link for IPv6 operations. It results that the model for IPv6 operation is similar to that of a BSS, with the root of the mesh operating as an Access Point does in a BSS/ESS.

While it is still possible to operate IPv6 ND, the inefficiencies of the flooding operation make the associated operations even less desirable than in a BSS, and the use of SND is highly recommended.

6.4. Case of DMB radios

IPv6 over DMB radios uses P2P links that can be formed and maintained when a pair of DMB radios transmitters are in range from one another.

6.4.1. Using IPv6 ND only

DMB radios do not provide MAC level broadcast emulation. An example of that is IEEE Std. 802.11 OCB which uses IEEE Std. 802.11 MAC/PHYs but does not provide the BSS functions.

It is possible to form P2P IP Links between each individual pairs of nodes and operate IPv6 ND over those links with link-local addresses. DAD must be performed for all addresses on all P2P IP Links.

If special deployment care is taken so that the physical broadcast domains of a collection of the nodes fully overlap, then it is also possible to build an IP Subnet within that collection of nodes and operate IPv6 ND.

If an external mechanism avoids duplicate addresses and if the deployment ensures the connectivity between peers, a non-transit Hub-and-Spoke deployment is also possible where the Hub is the only router in the Subnet and the Prefix is advertised as not on-link.

6.4.2. Using Subnet ND

Though this can be achieved with IPv6 ND, SND is the recommended approach since it uses unicast communications which are more reliable and less impacting for other users of the medium.

The routers send RAs with a SLLAO at a regular period. The period can be indicated in the RA-Interval Option [RFC6275]. If available, the message can be transported in a compressed form in a beacon, e.g., in OCB Basic Safety Messages (BSM) that are nominally sent every 100ms.

An active beaconing mode is possible whereby the Host sends broadcast RS messages to which a router can answer with a unicast RA.

A router that has Internet connectivity and is willing to serve as an Internet Access may advertise itself as a default router [RFC4191] in its RA messages. The RA is sent over an unspecified IP Link where it does not conflict to anyone, so DAD is not necessary at that stage.

The host instantiates an IP Link where the router’s address is not a duplicate. To achieve this, it forms a link-local address that does not conflict with that of the router and registers to the router using [RFC8505]. If the router sent an RA(PIO), the host can also autoconfigure an address from the advertised prefix and register it.

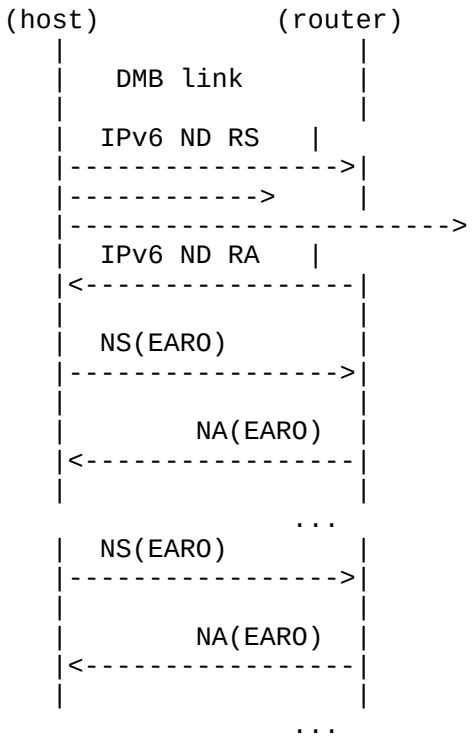

Figure 6: RFC 8505 Registration Flow for a link-local address

The lifetime in the registration should start with a small value (X=RMin, TBD), and exponentially grow with each re-registration to a larger value (X=Rmax, TBD). The IP Link is considered down when (X=NbBeacons, TDB) expected messages are not received in a row. It must be noted that the physical link flapping does not affect the state of the registration and when a physical link comes back up, the active registrations (i.e., registrations for which lifetime is not elapsed) are still usable. Packets should be held or destroyed when the IP Link is down.

P2P links may be federated in Hub-and-Spoke by edge routers, and the Subnet may comprise multiple edge routers, in which case each advertises its registered addresses over the SGP as illustrated in Figure 7. Note that the Extended DAR/DAC exchange can be omitted if it can be replaced with the information that is distributed in the SGP, see for instance [RFC9010] which applies to IoT environments, which needs only the first EDAR/EDAC exchange, and [EVPN-SFAAC], for EVPN-based wireless deployments in enterprise and campus, which does not use EDAR/EDAC at all.

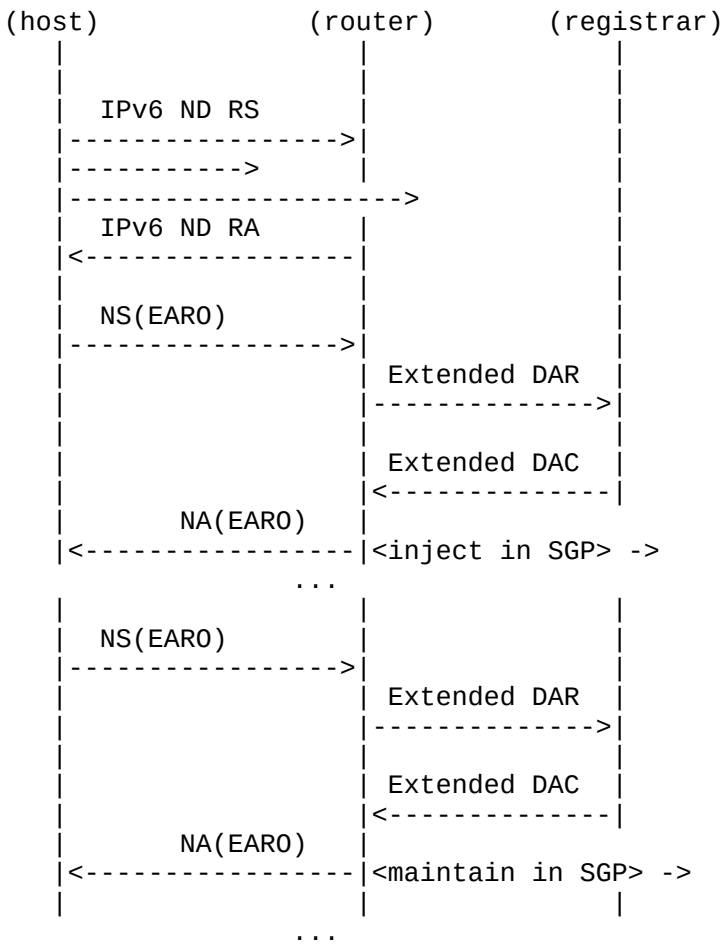

Figure 7: RFC 8505 Registration Flow for a Global address

An example Hub-and-Spoke is an OCB Road-Side Unit (RSU) that owns a prefix, provides Internet connectivity using that prefix to On-Board Units (OBUs) within its physical broadcast domain. An example of Route-Over MLSN is a collection of cars in a parking lot operating RPL to extend the connectivity provided by the RSU beyond its physical broadcast domain. Cars may then operate NEMO [RFC3963] for their own prefix using their address derived from the prefix of the RSU as CareOf address.

As opposed to unicast addresses, there might be multiple registrations from multiple parties for the same address. The router conserves one registration per party per multicast or anycast address, but injects the route into the SGP only once for each

address, asynchronously to the registration, as shown in Figure 8. On the other hand, the validation exchange with the registrar is still needed if the router checks the right for the host to listen to the anycast or multicast address.

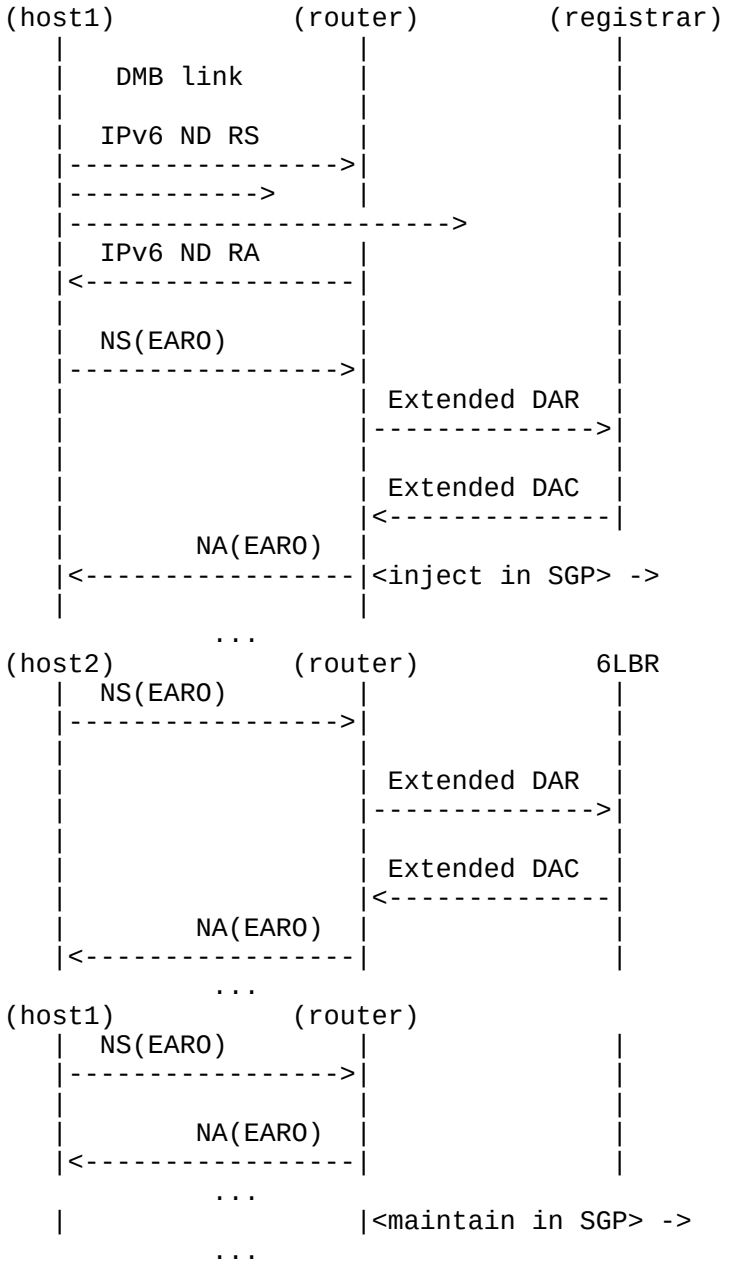

Figure 8: Registration Flow for an anycast or multicast address

6.4.3. Example: BLE and BLE Mesh

"IPv6 over BLUETOOTH(R) Low Energy" [RFC7668] was defined after [RFC6775] but before [RFC8505]. In other words, IPv6 over BLE leverages SFAAC, but the specification needed to be upgraded to add the reference to [RFC8505] so it uses the upgraded method for the registration.

This is updated by "IPv6 Mesh over BLUETOOTH(R) Low Energy Using the Internet Protocol Support Profile (IPSP)" [RFC9159]. [RFC9159] fully leverages SFAAC, referencing [RFC6775], [RFC8505], and [RFC8928]. The IP Subnet and IP Link models defined in section 3.2 and 3.3 of [RFC9159], respectively, are in conformance with this Architecture. As for the SGP, [RFC9159] relies on the router role defined in the IPSP.

6.4.4. Example: 6TiSCH

The Time-Slotted Channel Hopping Mode (TSCH) of IEEE Std 802.15.4 [IEEE Std. 802.15.4] defines one of the time-sensitive modes of IEEE Std 802.15.4. TSCH is both a Time-Division Multiplexing (TDM) and a Frequency-Division Multiplexing (FDM) technique, whereby a different channel can be used for each transmission. TSCH allows the scheduling of transmissions for deterministic operations and applies to the slower and most energy-constrained deterministic wireless use cases. The scheduled operation provides for a more reliable experience, which can be used to monitor and manage resources, e.g., energy and water, in a more efficient fashion. Proven deterministic networking standards for use in process control, including ISA100.11a [ISA100.11a] and WirelessHART [WirelessHART], have demonstrated the capabilities of the IEEE Std 802.15.4 TSCH MAC for high reliability against interference, low-power consumption on well-known flows, and its applicability for Traffic Engineering (TE) from a central controller. As such, TSCH has become the de facto standard for IPv6 in industrial wireless process control applications.

"An Architecture for IPv6 over the Time-Slotted Channel Hopping Mode of IEEE 802.15.4"[RFC9030] defines how IPv6 is operated over TSCH. As is the case for BLE mesh, [RFC9030] conforms with this Architecture in its concepts of IP Link and IP Subnet, and references the SFAAC RFCs [RFC6775], [RFC8505], and [RFC8928] for the ND procedures.

Additionally, the 6TiSCH Architecture leverages RPL [RFC6550] for the SGP, and allows to use ND Proxy operations [RFC8929] to federate multiple Links over a transit link called the backbone. Some nodes

may use SLAAC and see the prefix as on-link, mapping the backbone as a single IP transit Link, whereas other nodes rely on SFAAC and see the prefix as on-link, mapping multiple P2P IP Links over the backbone. In that case, ND proxy operations handles AR from SLAAC nodes to SFAAC nodes, whereas the routers forward the traffic from SFAAC nodes to SLAAC nodes. Note that MAC addresses being incompatible between IEEE Std 802.15.4 and IEEE Std 802.11, [RFC8929] operates in so called "routing proxy" between the 2, meaning that it impersonates the SLLAO of registered IPv6 addresses with its MAC address over the backbone, receives the traffic, and routes it over the edges links, using the appropriate MAC technology there.

7. Coexistence with IPv6 ND

The framework allows for a mixed environment with both models, IPv6 ND and SFAAC, coexist. With [RFC8929], an ethernet backbone link operating IPv6 ND federates a MultiLink Subnet (MLSN) of wireless links and/or meshes, and routers called Backbone Routers (6BBR) operate as ND proxies.

In a wireless deployments, the Backbone Routers are placed along the wireless edge of a backbone (e.g., in Access Points) and federate multiple wireless links to form a single MLSN, echoing the Wi-Fi ESS structure but at the network layer, as shown in Figure 9. In that example, Optimistic Duplicate address Detection (ODAD) [RFC4429] allows the IPv6 address to be used before completion of DAD, so the whole flow below can happen in the milliseconds that follow the Wi-Fi association.

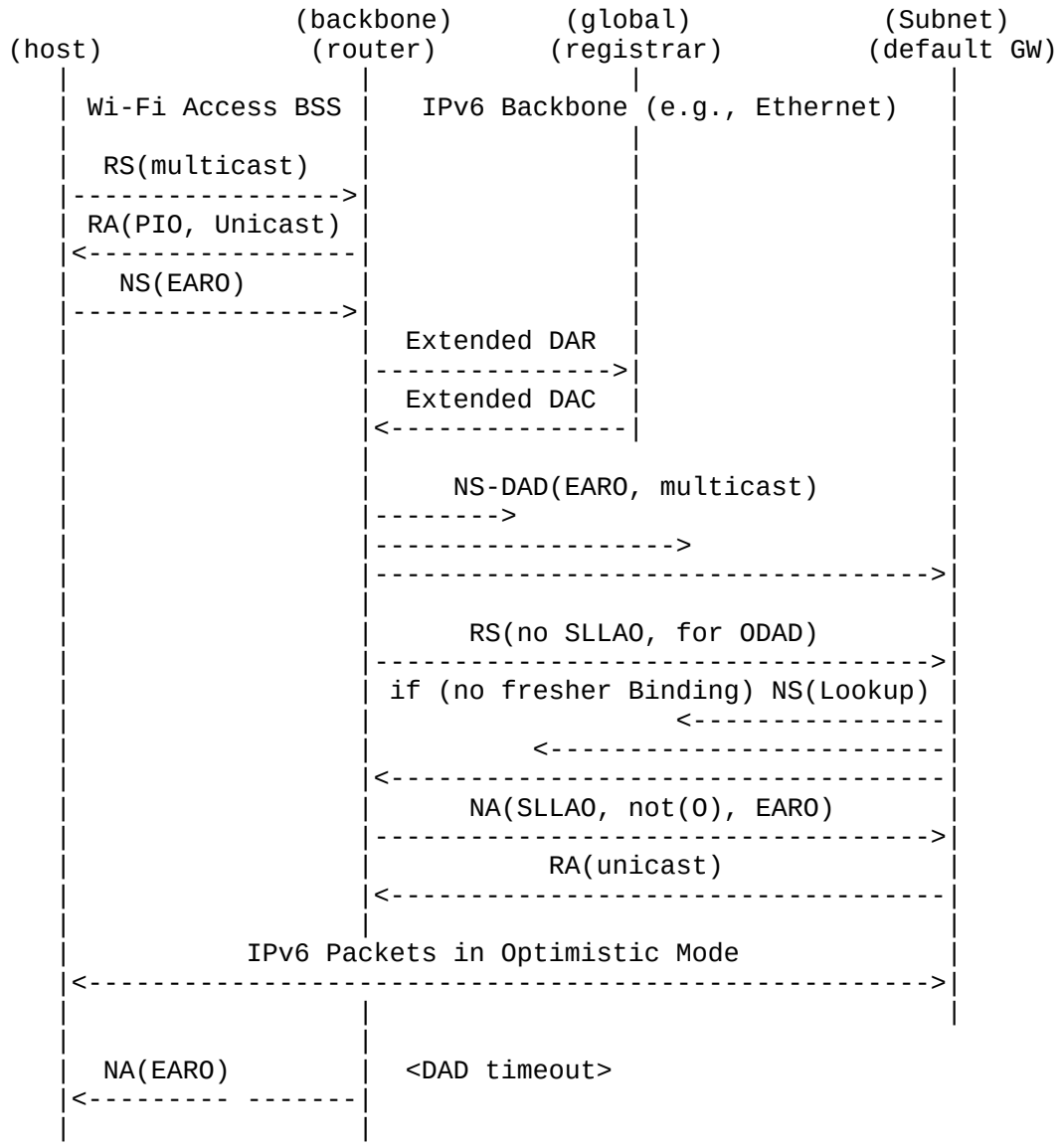

Figure 9: Initial Registration Flow to a 6BBR Acting as a ND Proxy

In use cases such as overlays, a Map Resolver acting as 6LBR may be deployed on the Backbone Link to serve the whole Subnet, and EDAR/ EDAC messages (or equivalent alternates, e.g., using LISP) can be used in combination with DAD to enable coexistence with IPv6 ND over the backbone. The 6LBR proactive operations will then coexist on the Backbone with the reactive IPv6 ND operation. Nodes that support

[UNICAST AR] may query the mappings they look up with the 6LBR before attempting the reactive operation, which may be avoided if the 6LBR is conclusive, either detecting a duplication or returning a mapping. This model also enables a snooping switch acting as ND proxy to intercept Ar and DAD NS messages and perform unicast lookups to the resolver and only broadcast the original NS message when the unicast lookup fails.

Note that the RS sent initially by the 6LN (e.g., a Wi-Fi STA) is transmitted as a multicast, but since it is intercepted by the 6BBR, it is never effectively broadcast at link layer. The multiple arrows in Figure 9 associated to the ND messages on the backbone denote a real link-layer broadcast.

It is not necessary to isolate the registering nodes in separate physical links, but it is preferred with wireless links as it enables to isolate the broadcast domain on the ethernet link from the wireless links at the Access Points. In other words, the 6BBRs collectively form a global registrar for the Subnet that aggregates the information in each local registrar in the 6LBR. The global registrar is distributed between the 6BBRs, which leverage IPv6 ND (AR and DAD) to lookup information that they do not have locally from the other 6BBRs and from nodes that are connected to the backbone.

In the case of wireless meshes, RPL may be used as local SGP in each mesh as shown in Figure 10. More details on the operation of SND and RPL over the MLSN can be found in section 3.1, 3.2, 4.1 and 4.2.2 of [RFC9030].

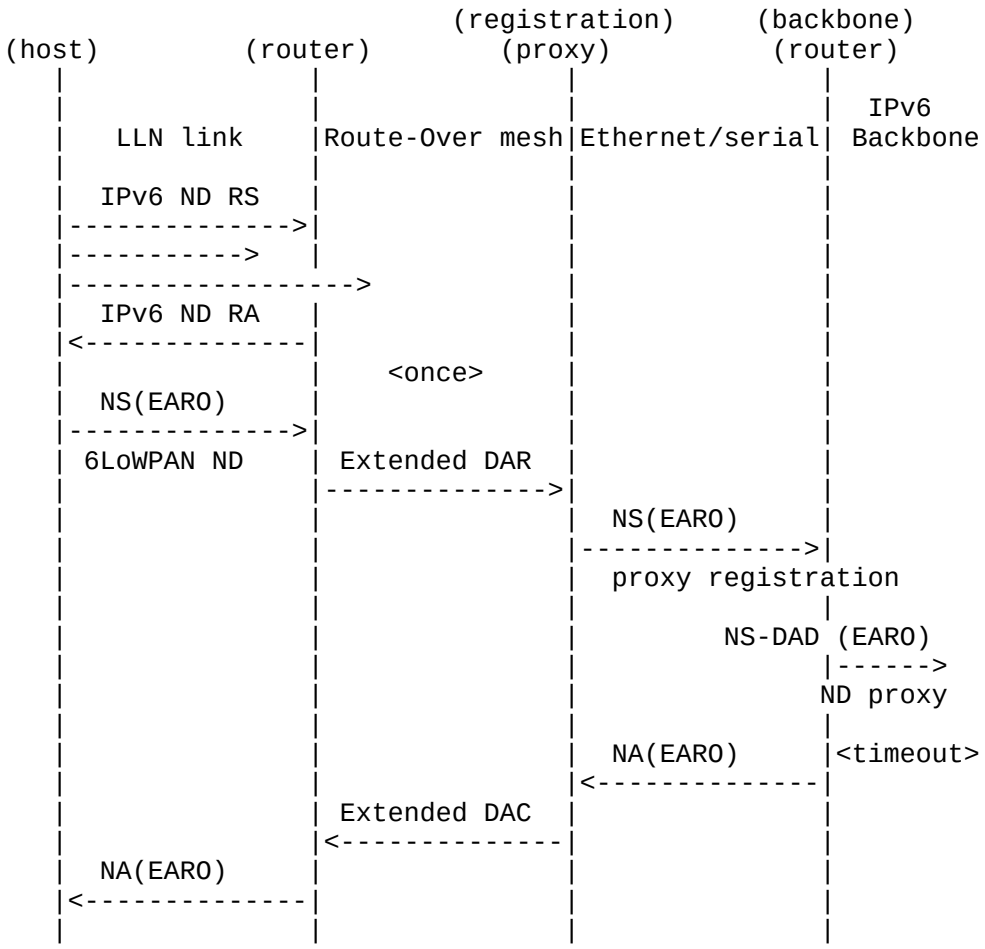

Figure 10: Initial Registration Flow with 6BBR ND-Proxy

8. Privacy Considerations

IPv6 ND exposes all addresses to all nodes in the Subnet, which is a privacy issue and makes impersonation attacks easier. In contrast, in switched and wireless networks, a host is not on-path of the unicast packets for registration and for data for other hosts, so it cannot snoop the other addresses in the network. A rogue host can only discover the existence of an addresses by trying and failing to register that address, but for that it would need to fathom which address to try and that can be very hard in, say, a SPL=64 address space that is used wisely. For that reason, this framework limits that knowledge to on-path snooping switches, to the routers and to the abstract registrar, which are typically more controlled / harder

to hack than the common host. When IPv6 ND and SFAAC coexist within the same Subnet, all addresses in the Subnet, including registered addresses, can be snooped in the broadcast domain where IPv6 ND is operated. It makes sense to reduce that domain to the maximum and control which device connect to it.

The exposure of addresses can be further reduced if the exchanges with the registrar (e.g., EDAR and EDAC) are encrypted, e.g., using a public key associated with the registrar. The registration and routing exchanges could also be encrypted to avoid leaking the addresses to snooping switches, but this is typically not done inside a physical site where the networking gear is tightly controlled. In a DCI environment, the inter-side (SD-WAN) links are typically encrypted, to the exchanges are obfuscated from an on-path listener.

9. Security Considerations

The registration model [RFC8505] implemented by this framework allows for a model where the ingress routers have a full knowledge of all the addresses in the Subnet. The ingress router can thus discard any packet which destination appears to be in the Subnet from its prefix, but is not known, meaning that it does not exist. This mostly defeats the traditional DoS scanning attacks against ND whereby the remote attacker sends volumes of packets to as many non-existent addresses to saturate the Neighbor Cache and clog the Subnet internal bandwidth in broadcasts.

When the ownership verifier is cryptographic, this framework enables a ZeroTrust model whereby only the address owner can advertise an address in ND and as source of data packets, more in [RFC8928]. This defeats the classical impersonation attacks against IPv6 ND and allows to disable the proprietary middlebox software aimed at protecting the address ownership against onlink rogues.

10. IANA Considerations

This specification does not require IANA action.

11. Contributors

Brian Carpenter Provided support, hints, and text snippets throughout the lifetime of the I-Draft

12. Acknowledgments

Many thanks to the participants of the 6lo WG where a lot of the work discussed here happened, following work at ROLL, 6TiSCH, and mainly 6LoWPAN.

Special thanks to Brian Carpenter and Eric Levy-Abegnoli who provided support and useful comments throughout the development of this architecture, and to Erik Nordmark and Zach Shelby with whom this work really started during IETF 72 in Dublin.

Also many thanks to Eduard Vasilenko, XiPeng Xiao, Behcet Sarikaya for their contributions and support to this work at 6MAN, v6Ops, and ETSI IPE.

13. Normative References

[RFC4191] Draves, R. and D. Thaler, "Default Router Preferences and More-Specific Routes", RFC 4191, DOI 10.17487/RFC4191, November 2005, <https://www.rfc-editor.org/info/rfc4191>.

[RFC4861] Narten, T., Nordmark, E., Simpson, W., and H. Soliman, "Neighbor Discovery for IP version 6 (IPv6)", RFC 4861, DOI 10.17487/RFC4861, September 2007, <https://www.rfc-editor.org/info/rfc4861>.

[RFC4862] Thomson, S., Narten, T., and T. Jinmei, "IPv6 Stateless Address Autoconfiguration", RFC 4862, DOI 10.17487/RFC4862, September 2007, <https://www.rfc-editor.org/info/rfc4862>.

[RFC5942] Singh, H., Beebee, W., and E. Nordmark, "IPv6 Subnet Model: The Relationship between Links and Subnet Prefixes", RFC 5942, DOI 10.17487/RFC5942, July 2010, <https://www.rfc-editor.org/info/rfc5942>.

[RFC6052] Bao, C., Huitema, C., Bagnulo, M., Boucadair, M., and X. Li, "IPv6 Addressing of IPv4/IPv6 Translators", RFC 6052, DOI 10.17487/RFC6052, October 2010, <https://www.rfc-editor.org/info/rfc6052>.

[RFC6275] Perkins, C., Ed., Johnson, D., and J. Arkko, "Mobility Support in IPv6", RFC 6275, DOI 10.17487/RFC6275, July 2011, <https://www.rfc-editor.org/info/rfc6275>.

[RFC6830] Farinacci, D., Fuller, V., Meyer, D., and D. Lewis, "The Locator/ID Separation Protocol (LISP)", RFC 6830, DOI 10.17487/RFC6830, January 2013, <https://www.rfc-editor.org/info/rfc6830>.

[RFC7432] Sajassi, A., Ed., Aggarwal, R., Bitar, N., Isaac, A., Uttaro, J., Drake, J., and W. Henderickx, "BGP MPLS-Based Ethernet VPN", RFC 7432, DOI 10.17487/RFC7432, February 2015, <https://www.rfc-editor.org/info/rfc7432>.

[RFC7346] Droms, R., "IPv6 Multicast Address Scopes", RFC 7346, DOI 10.17487/RFC7346, August 2014, <https://www.rfc-editor.org/info/rfc7346>.

[RFC8200] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", STD 86, RFC 8200, DOI 10.17487/RFC8200, July 2017, <https://www.rfc-editor.org/info/rfc8200>.

[RFC8505] Thubert, P., Ed., Nordmark, E., Chakrabarti, S., and C. Perkins, "Registration Extensions for IPv6 over Low-Power Wireless Personal Area Network (6LoWPAN) Neighbor Discovery", RFC 8505, DOI 10.17487/RFC8505, November 2018, <https://www.rfc-editor.org/info/rfc8505>.

[RFC8928] Thubert, P., Ed., Sarikaya, B., Sethi, M., and R. Struik, "Address-Protected Neighbor Discovery for Low-Power and Lossy Networks", RFC 8928, DOI 10.17487/RFC8928, November 2020, <https://www.rfc-editor.org/info/rfc8928>.

[I-D.ietf-6lo-updating-rfc-8928] Thubert, P. and A. Rashid, "Fixing the C-Flag in EARO", Work in Progress, Internet-Draft, draft-ietf-6lo-updating-rfc-8928-03, 17 May 2025, <https://datatracker.ietf.org/doc/html/draft-ietf-6lo-updating-rfc-8928-03>.

[RFC8929] Thubert, P., Ed., Perkins, C.E., and E. Levy-Abegnoli, "IPv6 Backbone Router", RFC 8929, DOI 10.17487/RFC8929, November 2020, <https://www.rfc-editor.org/info/rfc8929>.

[RFC9685] Thubert, P., Ed., "Listener Subscription for IPv6 Neighbor Discovery Multicast and Anycast Addresses", RFC 9685, DOI 10.17487/RFC9685, November 2024, <https://www.rfc-editor.org/info/rfc9685>.

[PREFIX REGISTRATION] Thubert, P., "IPv6 Neighbor Discovery Prefix Registration", Work in Progress, Internet-Draft, draft-ietf-6lo-prefix-registration-10, 17 April 2025, <https://datatracker.ietf.org/doc/html/draft-ietf-6lo-prefix-registration-10>.

14. Informative References

[RFC3963] Devarapalli, V., Wakikawa, R., Petrescu, A., and P. Thubert, "Network Mobility (NEMO) Basic Support Protocol", RFC 3963, DOI 10.17487/RFC3963, January 2005, <https://www.rfc-editor.org/info/rfc3963>.

[RFC3810] Vida, R., Ed. and L. Costa, Ed., "Multicast Listener Discovery Version 2 (MLDv2) for IPv6", RFC 3810, DOI 10.17487/RFC3810, June 2004, <https://www.rfc-editor.org/info/rfc3810>.

[RFC4291] Hinden, R. and S. Deering, "IP Version 6 Addressing Architecture", RFC 4291, DOI 10.17487/RFC4291, February 2006, <https://www.rfc-editor.org/info/rfc4291>.

[RFC4429] Moore, N., "Optimistic Duplicate Address Detection (DAD) for IPv6", RFC 4429, DOI 10.17487/RFC4429, April 2006, <https://www.rfc-editor.org/info/rfc4429>.

[RFC4903] Thaler, D., "Multi-Link Subnet Issues", RFC 4903, DOI 10.17487/RFC4903, June 2007, <https://www.rfc-editor.org/info/rfc4903>.

[RFC6550] Winter, T., Ed., Thubert, P., Ed., Brandt, A., Hui, J., Kelsey, R., Levis, P., Pister, K., Struik, R., Vasseur, JP., and R. Alexander, "RPL: IPv6 Routing Protocol for Low-Power and Lossy Networks", RFC 6550, DOI 10.17487/RFC6550, March 2012, <https://www.rfc-editor.org/info/rfc6550>.

[RFC6762] Cheshire, S. and M. Krochmal, "Multicast DNS", RFC 6762, DOI 10.17487/RFC6762, February 2013, <https://www.rfc-editor.org/info/rfc6762>.

[RFC6775] Shelby, Z., Ed., Chakrabarti, S., Nordmark, E., and C. Bormann, "Neighbor Discovery Optimization for IPv6 over Low-Power Wireless Personal Area Networks (6LoWPANs)", RFC 6775, DOI 10.17487/RFC6775, November 2012, <https://www.rfc-editor.org/info/rfc6775>.

[RFC7668] Nieminen, J., Savolainen, T., Isomaki, M., Patil, B., Shelby, Z., and C. Gomez, "IPv6 over BLUETOOTH(R) Low Energy", RFC 7668, DOI 10.17487/RFC7668, October 2015, <https://www.rfc-editor.org/info/rfc7668>.

[RFC7217] Gont, F., "A Method for Generating Semantically Opaque Interface Identifiers with IPv6 Stateless Address Autoconfiguration (SLAAC)", RFC 7217, DOI 10.17487/RFC7217, April 2014, <https://www.rfc-editor.org/info/rfc7217>.

[RFC7834] Saucez, D., Iannone, L., Cabellos, A., and F. Coras, "Locator/ID Separation Protocol (LISP) Impact", RFC 7834, DOI 10.17487/RFC7834, April 2016, <https://www.rfc-editor.org/info/rfc7834>.

[RFC8273] Brzozowski, J. and G. Van de Velde, "Unique IPv6 Prefix per Host", RFC 8273, DOI 10.17487/RFC8273, December 2017, <https://www.rfc-editor.org/info/rfc8273>.

[RFC8365] Sajassi, A., Ed., Drake, J., Ed., Bitar, N., Shekhar, R., Uttaro, J., and W. Henderickx, "A Network Virtualization Overlay Solution Using Ethernet VPN (EVPN)", RFC 8365, DOI 10.17487/RFC8365, March 2018, <https://www.rfc-editor.org/info/rfc8365>.

[RFC8415] Mrugalski, T., Siodelski, M., Volz, B., Yourtchenko, A., Richardson, M., Jiang, S., Lemon, T., and T. Winters, "Dynamic Host Configuration Protocol for IPv6 (DHCPv6)", RFC 8415, DOI 10.17487/RFC8415, November 2018, <https://www.rfc-editor.org/info/rfc8415>.

[RFC8981] Gont, F., Krishnan, S., Narten, T., and R. Draves, "Temporary Address Extensions for Stateless Address Autoconfiguration in IPv6", RFC 8981, DOI 10.17487/RFC8981, February 2021, <https://www.rfc-editor.org/info/rfc8981>.

[RFC9692] Przygienda, T., Ed., Head, J., Ed., Sharma, A., Thubert, P., Rijsman, B., and D. Afanasiev, "RIFT: Routing in Fat Trees", RFC 9692, DOI 10.17487/RFC9692, April 2025, <https://www.rfc-editor.org/info/rfc9692>.

[RFC9010] Thubert, P., Ed. and M. Richardson, "Routing for RPL (Routing Protocol for Low-Power and Lossy Networks) Leaves", RFC 9010, DOI 10.17487/RFC9010, April 2021, <https://www.rfc-editor.org/info/rfc9010>.

[DAD ISSUES]
Yourtchenko, A. and E. Nordmark, "A survey of issues related to IPv6 Duplicate Address Detection", Work in Progress, Internet-Draft, draft-yourtchenko-6man-dad-issues-01, 3 March 2015, <https://datatracker.ietf.org/doc/html/draft-yourtchenko-6man-dad-issues-01>.

[MCAST EFFICIENCY]
Vyncke, E., Thubert, P., Levy-Abegnoli, E., and A. Yourtchenko, "Why Network-Layer Multicast is Not Always Efficient At Datalink Layer", Work in Progress, Internet-Draft, draft-vyncke-6man-mcast-not-efficient-01, 14 February 2014, <https://datatracker.ietf.org/doc/html/draft-vyncke-6man-mcast-not-efficient-01>.

[RFC9030] Thubert, P., Ed., "An Architecture for IPv6 over the Time-Slotted Channel Hopping Mode of IEEE 802.15.4 (6TiSCH)", RFC 9030, DOI 10.17487/RFC9030, May 2021, <https://www.rfc-editor.org/info/rfc9030>.

[RFC9119] Perkins, C., McBride, M., Stanley, D., Kumari, W., and JC. Z炭単iga, "Multicast Considerations over IEEE 802 Wireless Media", RFC 9119, DOI 10.17487/RFC9119, October 2021, <https://www.rfc-editor.org/info/rfc9119>.

[RFC9159] Gomez, C., Darroudi, S.M., Savolainen, T., and M. Spoerk, "IPv6 Mesh over BLUETOOTH(R) Low Energy Using the Internet Protocol Support Profile (IPSP)", RFC 9159, DOI 10.17487/RFC9159, December 2021, <https://www.rfc-editor.org/info/rfc9159>.

[ND CONSIDERATIONS]
Xiao, X., V, Metz, E., Mishra, G. S., and N. Buraglio, "Neighbor Discovery Considerations in IPv6 Deployments", Work in Progress, Internet-Draft, draft-ietf-v6ops-nd-considerations-12, 28 April 2025, <https://datatracker.ietf.org/doc/html/draft-ietf-v6ops-nd-considerations-12>.

[SAVI] Bi, J., Wu, J., Lin, T., Wang, Y., and L. He, "A SAVI Solution for WLAN", Work in Progress, Internet-Draft, draft-bi-savi-wlan-24, 13 November 2022, <https://datatracker.ietf.org/doc/html/draft-bi-savi-wlan-24>.

[UNICAST AR]
Thubert, P. and E. Levy-Abegnoli, "IPv6 Neighbor Discovery Unicast Lookup", Work in Progress, Internet-Draft, draft-thubert-6lo-unicast-lookup-02, 8 November 2021, <https://datatracker.ietf.org/doc/html/draft-thubert-6lo-unicast-lookup-02>.

[DAD APPROACHES]
Nordmark, E., "Possible approaches to make DAD more robust and/or efficient", Work in Progress, Internet-Draft, draft-nordmark-6man-dad-approaches-02, 19 October 2015, <https://datatracker.ietf.org/doc/html/draft-nordmark-6man-dad-approaches-02>.

[EVPN-SFAAC]
Thubert, P., Przygienda, T., and J. Tantsura, "Secure EVPN MAC Signaling", Work in Progress, Internet-Draft, draft-thubert-bess-secure-evpn-mac-signaling-04, 13 September 2023, <https://datatracker.ietf.org/doc/html/draft-thubert-bess-secure-evpn-mac-signaling-04>.

[I-D.ietf-snac-simple]
Lemon, T. and J. Hui, "Automatically Connecting Stub Networks to Unmanaged Infrastructure", Work in Progress, Internet-Draft, draft-ietf-snac-simple-06, 4 November 2024, <https://datatracker.ietf.org/doc/html/draft-ietf-snac-simple-06>.

[I-D.ietf-v6ops-dhcp-pd-per-device]
Colitti, L., Linkova, J., and X. Ma, "Using DHCPv6-PD to Allocate Unique IPv6 Prefix per Client in Large Broadcast Networks", Work in Progress, Internet-Draft, draft-ietf-v6ops-dhcp-pd-per-device-08, 3 April 2024, <https://datatracker.ietf.org/doc/html/draft-ietf-v6ops-dhcp-pd-per-device-08>.

[WirelessHART]
www.hartcomm.org, "Industrial Communication Networks - Wireless Communication Network and Communication Profiles - WirelessHART - IEC 62591", 2010.

[ISA100.11a]
ISA/ANSI, "Wireless Systems for Industrial Automation: Process Control and Related Applications - ISA100.11a-2011 - IEC 62734", 2011, <http://www.isa.org/Community/SP100WirelessSystemsforAutomation>.

[IEEE Std. 802.15.4]
IEEE standard for Information Technology, "IEEE Std. 802.15.4, Part. 15.4: Wireless Medium Access Control (MAC) and Physical layer (PHY) Specifications for Low-Rate Wireless Personal Area Networks".

[IEEE Std. 802.11]
IEEE standard for Information Technology, "IEEE Standard for Information technology -- Telecommunications and information exchange between systems Local and metropolitan area networks-- Specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical layer (PHY) Specifications".

[IEEEstd802151]
IEEE standard for Information Technology, "IEEE Standard for Information Technology - Telecommunications and Information Exchange Between Systems - Local and Metropolitan Area Networks - Specific Requirements. - Part 15.1: Wireless Medium Access Control (MAC) and Physical layer (PHY) Specifications for Wireless Personal Area Networks (WPANs)".

[IEEE Std. 802.1]
IEEE standard for Information Technology, "IEEE Standard for Information technology -- Telecommunications and information exchange between systems Local and metropolitan area networks Part 1: Bridging and Architecture".

Authors' Addresses

Pascal Thubert (editor)
06330 Roquefort-les-Pins
France
Email: pascal.thubert@gmail.com

Michael C. Richardson
Sandelman Software Works
Email: mcr+ietf@sandelman.ca
URI: http://www.sandelman.ca/