Internet Area
Internet-Draft
Intended status: Informational
Expires: July 15, 2016

E. Baccelli
INRIA
C. Perkins
Futurewei

January 12, 2016

Multi-hop Ad Hoc Wireless Communication
draft-ietf-intarea-adhoc-wireless-com-01

Abstract

This document describes characteristics of communication between interfaces in a multi-hop ad hoc wireless network, that protocol engineers and system analysts should be aware of when designing solutions for ad hoc networks at the IP layer.

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 http://datatracker.ietf.org/drafts/current/.

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

This Internet-Draft will expire on July 15, 2016.

Copyright Notice

Table of Contents

1. Introduction

Experience gathered with ad hoc routing protocol development, deployment and operation, shows that wireless communication presents specific challenges [RFC2501] [DoD01], which Internet protocol designers should be aware of, when designing solutions for ad hoc networks at the IP layer. This document does not prescribe solutions, but instead briefly describes these challenges in hopes of increasing that awareness.

As background, RFC 3819 [RFC3819] provides an excellent reference for higher-level considerations when designing protocols for shared media. From MTU to subnet design, from security to considerations about retransmissions, RFC 3819 provides guidance and design rationale to help with many aspects of higher-level protocol design.

The present document focuses more specifically on challenges in multi-hop ad hoc wireless networking. For example, in that context, even though a wireless link may experience high variability as a communications channel, such variation does not mean that the link is "broken"; indeed many layer-2 technologies serve to reduce error rates by various means. Nevertheless, such errors as noted in this document may still become visible above layer-2 and so become relevant to the operation of higher layer protocols.

2. Multi-hop Ad Hoc Wireless Networks

For the purposes of this document, a multi-hop ad hoc wireless network will be considered to be a collection of devices that each have a radio transceiver (i.e., wireless network interface), and that are moreover configured to self-organize and provide store-and-forward functionality as needed to enable communications. This

document focuses on the characteristics of communications through such a network interface.

Although the characteristics of packet transmission over multi-hop ad hoc wireless networks, described below, are not the typical characteristics expected by IP [RFC6250], it is desirable and possible to run IP over such networks, as demonstrated in certain deployments currently in operation, such as Freifunk [FREIFUNK], and Funkfeuer [FUNKFEUER]. These deployments use routers running IP protocols e.g., OLSR (Optimized Link State Routing [RFC3626]) on top of IEEE 802.11 in ad hoc mode with the same ESSID (Extended Service Set Identification) at the link layer. Multi-hop ad hoc wireless networks may also run on link layers other than IEEE 802.11, and may use routing protocols other than OLSR (for instance, AODV [RFC3561], TBRPF [RFC3684], DSR [RFC4728], or OSPF-MPR [RFC5449]).

Note that in contrast, devices communicating via an IEEE 802.11 access point in infrastructure mode do not form a multi-hop ad hoc wireless network, since the central role of the access point is predetermined, and devices other than the access point do not generally provide store-and-forward functionality.

3. Common Packet Transmission Characteristics in Multi-hop Ad Hoc Wireless Networks

In the following, we will consider several devices in a multi-hop ad hoc wireless network N. Each device will be considered only through its own wireless interface to network N. For conciseness and readability, this document uses the expressions "device A" (or simply "A") as a synonym for "the wireless interface of device A to network N".

Let A and B be two devices in network N. Suppose that, when device A transmits an IP packet through its interface on network N, that packet is correctly and directly received by device B without requiring storage and/or forwarding by any other device. We will then say that B can "detect" A. Note that therefore, when B detects A, an IP packet transmitted by A will be rigorously identical to the corresponding IP packet received by B.

Let S be the set of devices that detect device A through its wireless interface on network N. The following section gathers common characteristics concerning packet transmission over such networks, which were observed through experience with MANET routing protocol development (for instance, OLSR[RFC3626], AODV[RFC3561], TBRPF[RFC3684], DSR[RFC4728], and OSPF-MPR[RFC5449]), as well as deployment and operation (Freifunk[FREIFUNK], Funkfeuer[FUNKFEUER]).

3.1. Asymmetry, Time-Variation, and Non-Transitivity

First, even though a device C in set S can (by definition) detect device A, there is no guarantee that C can, conversely, send IP packets directly to A. In other words, even though C can detect A (since it is a member of set S), there is no guarantee that A can detect C. Thus, multi-hop ad hoc wireless communications may be "asymmetric". Such cases are common.

Second, there is no guarantee that, as a set, S is at all stable, i.e. the membership of set S may in fact change at any rate, at any time. Thus, multi-hop ad hoc wireless communications may be "time-variant". Time variation is often observed in multi-hop ad hoc wireless networks due to variability of the wireless medium, and to device mobility.

Now, conversely, let V be the set of devices which A detects. Suppose that A is communicating at time t0 through its interface on network N. As a consequence of time variation and asymmetry, we observe that A:

1. cannot assume that S = V,

2. cannot assume that S and/or V are unchanged at time t1 later than t0.

Furthermore, transitivity is not guaranteed over multi-hop ad hoc wireless networks. Indeed, let’s assume that, through their respective interfaces within network N:

1. device B and device A can detect one another (i.e. B is a member of sets S and V), and,

2. device A and device C can also detect one another (i.e. C is a also a member of sets S and V).

These assumptions do not imply that B can detect C, nor that C can detect B (through their interface on network N). Such "non-transitivity" is common on multi-hop ad hoc wireless networks.

In a nutshell: multi-hop ad hoc wireless communications can be asymmetric, non-transitive, and time-varying.

3.2. Radio Range and Wireless Irregularities

Section 3.1 presents an abstract description of some common characteristics concerning packet transmission over multi-hop ad hoc wireless networks. This section describes practical examples, which

illustrate the characteristics listed in Section 3.1 as well as other common effects.

Wireless communications are subject to limitations to the distance across which they may be established. The range-limitation factor creates specific problems on multi-hop ad hoc wireless networks. In this context, the radio ranges of several devices often partially overlap. Such partial overlap causes communication to be non-transitive and/or asymmetric, as described in Section 3.1. Moreover, the range may vary from one device to another, depending on location and environmental factors. This is in addition to the time variation of range and signal strength caused by variability in the local environment.

For example, as depicted in Figure 1, it may happen that a device B detects a device A which transmits at high power, whereas B transmits at lower power. In such cases, B detects A, but A cannot detect B. This examplifies the asymmetry in multi-hop ad hoc wireless communications as defined in Section 3.1.

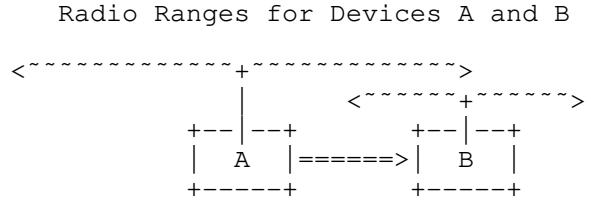


Figure 1: Asymmetric wireless communication: Device A can communicate with device B, but B cannot communicate with A.

Another example, depicted in Figure 2, is known as the "Hidden Terminal" problem. Even though the devices all have equal power for their radio transmissions, they cannot all detect one another. In the figure, devices A and B can detect one another, and devices A and C can also detect one another. On the other hand, B and C cannot detect one another. When B and C simultaneously try to communicate with A, their radio signals may collide. Device A may receive incoherent noise, and may even be unable to determine the source of the noise. The hidden terminal problem illustrates the property of non-transitivity in multi-hop ad hoc wireless communications as described in Section 3.1.

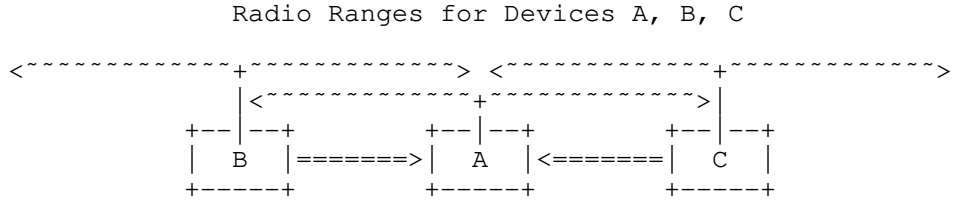


Figure 2: The hidden terminal problem. Devices C and B try to communicate with device A at the same time, and their radio signals collide.

Another situation, shown in Figure 3, is known as the "Exposed Terminal" problem. In the figure, device A and device B can detect each other, and A is transmitting packets to B, thus A cannot detect device C -- but C can detect A. As shown in Figure 3, during the on-going transmission of A, device C cannot reliably communicate with device D because of interference within C's radio range due to A's transmissions. Device C is then said to be "exposed", because it is exposed to co-channel interference from A and is thereby prevented from reliably exchanging protocol messages with D -- even though these transmissions would not interfere with the reception of data sent from A destined to B.

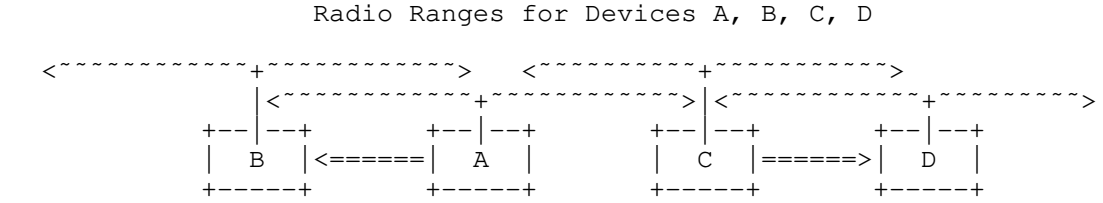


Figure 3: The exposed terminal problem: when device A communicates with device B, device C is "exposed".

Hidden and exposed terminal situations are often observed in multi-hop ad hoc wireless networks. Asymmetry issues with wireless communication may also arise for reasons other than power inequality (e.g., multipath interference). Such problems are often resolved by specific mechanisms below the IP layer, for example, CSMA/CA, which

ensures transmission in periods perceived to be unoccupied by other transmissions. However, depending on the link layer technology in use and the position of the devices, such problems may affect the IP layer due to range-limitation and partial overlap .

Besides radio range limitations, wireless communications are affected by irregularities in the shape of the geographical area over which devices may effectively communicate (see for instance [MC03], [MI03]). For example, even omnidirectional wireless transmission is typically non-isotropic (i.e. non-circular). Signal strength often suffers frequent and significant variations, which are not a simple function of distance. Instead, it is a complex function of the environment including obstacles, weather conditions, interference, and other factors that change over time. Because wireless communications have to encounter different terrain, path, obstructions, atmospheric conditions and other phenomena, analytical formulation of signal strength is considered intractable [VTC99], and the radio engineering community has thus developed numerous radio propagation models, relying on median values observed in specific environments [SAR03].

The above irregularities also cause communications on multi-hop ad hoc wireless networks to be non-transitive, asymmetric, or time-varying, as described in Section 3.1, and may impact protocols at the IP layer and above. There may be no indication to the IP layer when a previously established communication channel becomes unusable; "link down" triggers are generally absent in multi-hop ad hoc wireless networks, since the absence of detectable radio energy (e.g., in carrier waves) may simply indicate that neighboring devices are not currently transmitting. Such an absence of detectable radio energy does not therefore indicate whether or not transmissions have failed to reach the intended destination.

4. Alternative Terminology

Many terms have been used in the past to describe the relationship of devices in a multi-hop ad hoc wireless network based on their ability to send or receive packets to/from each other. The terms used in previous sections of this document have been selected because the authors believe they are unambiguous, with respect to the goal of this document (see Section 1).

In this section, we exhibit some other terms that describe the same relationship between devices in multi-hop ad hoc wireless networks. In the following, let network N be, again, a multi-hop ad hoc wireless network. Let the set S be, as before, the set of devices that can directly receive packets transmitted by device A through its interface on network N. In other words, any device B belonging to S

can detect packets transmitted by A. Then, due to the asymmetric nature of wireless communications:

- We may say that device A "reaches" device B. In this terminology, there is no guarantee that B reaches A, even if A reaches B.

- We may say that device B "hears" device A. In this terminology, there is no guarantee that A hears B, even if B hears A.

- We may say that device A "has a link" to device B. In this terminology, there is no guarantee that B has a link to A, even if A has a link to B.

- We may say that device B "is adjacent to" device A. In this terminology, there is no guarantee that A is adjacent to B, even if B is adjacent to A.

- We may say that device B "is downstream from" device A. In this terminology, there is no guarantee that A is downstream from B, even if B is downstream from A.

- We may say that device B "is a neighbor of" device A. In this terminology, there is no guarantee that A is a neighbor of B, even if B a neighbor of A. As it happens, terminology based on "neighborhood" is quite confusing for multi-hop wireless communications. For example, when B can detect A, but A cannot detect B, it is not clear whether B should be considered a neighbor of A at all, since A would not necessarily be aware that B was a neighbor, as it cannot detect B. It is thus best to avoid the "neighbor" terminology, except for when some level of symmetry has been verified.

This list of alternative terminologies is given here for illustrative purposes only, and is not suggested to be complete or even representative of the breadth of terminologies that have been used in various ways to explain the properties mentioned in Section 3. We do not discuss bidirectionality, but as a final observation it is worthwhile to note that bidirectionality is not synonymous with symmetry. For example, the error statistics in either direction are often different for a link that is otherwise considered bidirectional.

5. Security Considerations

Section 18 of RFC 3819 [RFC3819] provides an excellent overview of security considerations at the subnetwork layer. Beyond the material there, multi-hop ad hoc wireless networking (i) is not limited to

subnetwork layer operation, and (ii) makes use of wireless communications.

On one hand, a detailed description of security implications of wireless communications in general is outside of the scope of this document. Notably, however, eavesdropping on a wireless link is much easier than for wired media (although significant progress has been made in the field of wireless monitoring of wired transmissions). As a result, traffic analysis attacks can be even more subtle and difficult to defeat in this context. Furthermore, such communications over a shared media are particularly prone to theft of service and denial of service (DoS) attacks.

On the other hand, the potential multi-hop aspect of the networks we consider in this document goes beyond traditional scope of subnetwork design. In practice, unplanned relaying of network traffic (both user traffic and control traffic) happens routinely. Due to the physical nature of wireless media, Man in the Middle (MITM) attacks are facilitated, which may significantly alter network performance. This highlights the need to stick to the "end-to-end principle": L3 security, end-to-end, becomes a primary goal, independently of securing layer-2 and layer-1 protocols (though L2 and L1 security can indeed help to reach this goal).

6. IANA Considerations

This document does not have any IANA actions.

7. Informative References

[RFC2501] Corson, S. and J. Macker, "Mobile Ad hoc Networking (MANET): Routing Protocol Performance Issues and Evaluation Considerations", RFC 2501, DOI 10.17487/RFC2501, January 1999, <http://www.rfc-editor.org/info/rfc2501>.

[RFC3561] Perkins, C., Belding-Royer, E., and S. Das, "Ad hoc On-Demand Distance Vector (AODV) Routing", RFC 3561, DOI 10.17487/RFC3561, July 2003, <http://www.rfc-editor.org/info/rfc3561>.

[RFC3626] Clausen, T., Ed. and P. Jacquet, Ed., "Optimized Link State Routing Protocol (OLSR)", RFC 3626, DOI 10.17487/RFC3626, October 2003, <http://www.rfc-editor.org/info/rfc3626>.

[RFC3684] Ogier, R., Templin, F., and M. Lewis, "Topology Dissemination Based on Reverse-Path Forwarding (TBRPF)", RFC 3684, DOI 10.17487/RFC3684, February 2004, <http://www.rfc-editor.org/info/rfc3684>.

[RFC3819] Karn, P., Ed., Bormann, C., Fairhurst, G., Grossman, D., Ludwig, R., Mahdavi, J., Montenegro, G., Touch, J., and L. Wood, "Advice for Internet Subnetwork Designers", BCP 89, RFC 3819, DOI 10.17487/RFC3819, July 2004, <http://www.rfc-editor.org/info/rfc3819>.

[RFC4728] Johnson, D., Hu, Y., and D. Maltz, "The Dynamic Source Routing Protocol (DSR) for Mobile Ad Hoc Networks for IPv4", RFC 4728, DOI 10.17487/RFC4728, February 2007, <http://www.rfc-editor.org/info/rfc4728>.

[RFC5449] Baccelli, E., Jacquet, P., Nguyen, D., and T. Clausen, "OSPF Multipoint Relay (MPR) Extension for Ad Hoc Networks", RFC 5449, DOI 10.17487/RFC5449, February 2009, <http://www.rfc-editor.org/info/rfc5449>.

[RFC6250] Thaler, D., "Evolution of the IP Model", RFC 6250, DOI 10.17487/RFC6250, May 2011, <http://www.rfc-editor.org/info/rfc6250>.

[DoD01] Freebersyser, J. and B. Leiner, "A DoD perspective on mobile ad hoc networks", Addison Wesley C. E. Perkins, Ed., 2001, pp. 29--51, 2001.

[FUNKFEUER] "Austria Wireless Community Network, http://www.funkfeuer.at", 2013.

[MC03] Corson, S. and J. Macker, "Mobile Ad hoc Networking: Routing Technology for Dynamic, Wireless Networks", IEEE Press Mobile Ad hoc Networking, Chapter 9, 2003.

[SAR03] Sarkar, T., Ji, Z., Kim, K., Medour, A., and M. Salazar-Palma, "A Survey of Various Propagation Models for Mobile Communication", IEEE Press Antennas and Propagation Magazine, Vol. 45, No. 3, 2003.

[VTC99] Kim, D., Chang, Y., and J. Lee, "Pilot power control and service coverage support in CDMA mobile systems", IEEE Press Proceedings of the IEEE Vehicular Technology Conference (VTC), pp.1464-1468, 1999.

[MI03] Kotz, D., Newport, C., and C. Elliott, "The Mistaken Axioms of Wireless-Network Research", Dartmouth College Computer Science Technical Report TR2003-467, 2003.

[FREIFUNK]
"Freifunk Wireless Community Networks, http://www.freifunk.net", 2013.

Appendix A. Acknowledgements

This document stems from discussions with the following people, in alphabetical order: Jari Arkko, Teco Boot, Carlos Jesus Bernardos Cano, Ian Chakeres, Thomas Clausen, Robert Cragie, Christopher Dearlove, Ralph Droms, Brian Haberman, Ulrich Herberg, Paul Lambert, Kenichi Mase, Thomas Narten, Erik Nordmark, Alexandru Petrescu, Stan Ratliff, Zach Shelby, Shubhranshu Singh, Fred Templin, Dave Thaler, Mark Townsley, Ronald Velt in't, and Seung Yi.

Authors' Addresses

Emmanuel Baccelli
INRIA

EMail: Emmanuel.Baccelli@inria.fr
URI: http://www.emmanuelbaccelli.org/

Charles E. Perkins
Futurewei

Phone: +1-408-330-4586
EMail: charlie.perkins@huawei.com

INTAREA
Internet-Draft
Intended status: Standards Track
Expires: September 2, 2016

E. Nordmark
Arista Networks

Mar 2016

IP over Intentionally Partially Partitioned Links
draft-nordmark-intarea-ippl-03

Abstract

IP makes certain assumptions about the L2 forwarding behavior of a multi-access IP link. However, there are several forms of intentional partitioning of links ranging from split-horizon to Private VLANs that violate some of those assumptions. This document specifies that link behavior and how IP handles links with those properties.

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 http://datatracker.ietf.org/drafts/current/.

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

This Internet-Draft will expire on September 2, 2016.

described in the Simplified BSD License.

Table of Contents

1. Introduction

IPv4 and IPv6 can in general handle two forms of links; point-to-point links when only have two IP nodes (self and remote), and multi-access links with one or more nodes attached to the link. For the multi-access links IP in general, and particular protocols like ARP and IPv6 Neighbor Discovery, makes a few assumptions about transitive and reflexive connectivity i.e., that all nodes attached to the link can send packets to all other nodes.

There are cases where for various reasons and deployments one wants what looks like one link from the perspective of IP and routing, yet the L2 connectivity is restrictive. A key property is that an IP subnet prefix is assigned to the link, and IP routing sees it as a regular multi-access link. But a host attached to the link might not be able to send packets to all other hosts attached to the link. The motivation for this is outside the scope of this document, but in summary the motivation to preserve the subnet view as seen by IP routing is to conserve IP(v4) address space, and the motivation to restrict communication on the link could be due to (security) policy or potentially wireless connectivity approaches.

This intentional and partial partition appears in a few different forms. For DSL [TR-101] and Cable [DOCSIS-MULPI] the pattern is to have a single access router on the link, and all the hosts can send and receive from the access router, but host-to-host communication is blocked. A richer set of restrictions are possible for Private VLANs (PVLAN) [RFC5517], which has a notion of three different ports i.e. attachment points: isolated, community, and promiscuous. Note that other techniques operate at L2/L3 boundary like [RFC4562] but those are out of scope for this document.

The possible connectivity patterns for PVLAN appears to be a superset of the DSL and Cable use of split horizon, thus this document specifies the PVLAN behavior, shows the impact on IP/ARP/ND, and specifies how IP/ARP/ND must operate to work with PVLAN.

If private VLANs, or the split horizon subset, has been configured at layer 2 for the purposes of IPv4 address conservation, then that layer 2 configuration will affect IPv6 even though IPv6 might not have the same need for address conservation.

2. Keywords and Terminology

The keywords MUST, MUST NOT, REQUIRED, SHALL, SHALL NOT, SHOULD, SHOULD NOT, RECOMMENDED, MAY, and OPTIONAL, when they appear in this document, are to be interpreted as described in [RFC2119].

The following terms from [RFC4861] are used without modifications:

node	a device that implements IP.
router	a node that forwards IP packets not explicitly addressed to itself.
host	any node that is not a router.
link	a communication facility or medium over which nodes can communicate at the link layer, i.e., the layer immediately below IP. Examples are Ethernets (simple or bridged), PPP links, X.25, Frame Relay, or ATM networks as well as Internet-layer (or higher-layer) "tunnels", such as tunnels over IPv4 or IPv6 itself.
interface	a node's attachment to a link.
neighbors	nodes attached to the same link.

This document defines the following set of terms:

bridge	a layer-2 device which implements 802.1Q
port	a bridge's attachment to another bridge or to a node.

3. Private VLAN

A private VLAN is a structure which uses two or more 802.1Q (VLAN) values to separate what would otherwise be a single VLAN, viewed by IP as a single broadcast domain, into different types of ports with different L2 forwarding behavior between the different ports. A private VLAN consists of a single primary VLAN and multiple secondary VLANs.

From the perspective of both a single bridge and a collection of interconnected bridges there are three different types of ports use to attach nodes plus an inter-bridge port:

- Promiscuous: A promiscuous port can send packets to all ports that are part of the private VLAN. Such packets are sent using the primary VLAN ID.
- Isolated: Isolated VLAN ports can only send packets to promiscuous ports. Such packets are sent using an isolated VLAN ID.
- Community: A community port is associated with a per-community VLAN ID, and can send packets to both ports in the same community VLAN and promiscuous ports.
- Inter-bridge: A port used to connect a bridge to another bridge.

3.1. Bridge Behavior

Once a bridge or a set of interconnected bridges have been configured with both the primary and isolated VLAN ID, and zero or more community VLAN IDs associated with the private VLAN, the following forward behaviors apply to the bridge:

- o A packet received on an isolated port MUST NOT be forwarded out an isolated or community port; it SHOULD (subject to bandwidth/ resource issues) be forwarded out promiscuous and inter-bridge ports.
- o A packet received on a community port MUST NOT be forwarded out an isolated port or a community port with a different VLAN ID; it SHOULD be forwarded out promiscuous and inter-bridge ports as well as community ports that have the same community VLAN ID.
- o A packet received on a promiscuous port SHOULD be forwarded out all types of ports in the private VLAN.
- o A packet received on an inter-bridge port with an isolated VLAN ID should be forwarded as a packet received on an isolated port.
- o A packet received on an inter-bridge port with a community VLAN ID should be forwarded as a packet received on a community port associated with that VLAN ID.
- o A packet received on an inter-bridge port with a promiscuous VLAN ID should be forwarded as a packet received on a promiscuous port.

In addition to the above VLAN filtering and implied MAC address learning rules, the packet forwarding is also subject to the normal 802.1Q rules with blocking ports due to spanning-tree protocol etc.

4. IP over IPPL

When IP is used over Intentionally Partially Partitioned links like private VLANs the normal usage is to attached routers (and potentially other shared resources like servers) to promiscuous ports, while attaching other hosts to either community or isolated ports. If there is a single host for a given tenant or other domain of separation, then it is most efficient to attach that host to an isolated port. If there are multiple hosts in the private VLAN that should be able to communicate at layer 2, then they should be assigned a common community VLAN ID and attached to ports with that VLAN ID.

The above configuration means that hosts will not be able to communicate with each other unless they are in the same community. However, mechanisms outside of the scope of this document can be used to allow IP communication between such hosts e.g., by having firewall or gateway in or beyond the routers connected to the promiscuous ports. When such a policy is in place it is important that all packets which cross communities are sent to a router, which can have access-control lists or deeper firewall rules to decide which packets to forward.

5. IPv6 over IPPL

IPv6 Neighbor Discovery [RFC4861] can be used to get all the hosts on the link to send all unicast packets except those send to link-local destination addresses to the routers. That is done by setting the L-flag (on-link) to zero for all of the Prefix Information options. Note that this is orthogonal to whether SLAAC (Stateless Address Auto-Configuration) [RFC4862] or DHCPv6 [RFC3315] is used for address autoconfiguration. Setting the L-flag to zero is RECOMMENDED configuration for private VLANs.

If the policy includes allowing some packets that are sent to link-local destinations to cross between different tenants, then some for of NS/NA proxy is needed in the routers, and the routers need to forward packets addressed to link-local destinations out the same interface as REQUIRED in [RFC2460]. If the policy allows for some packets sent to global IPv6 address to cross between tenants then the routers would forward such packets out the same interface. However, with the L=0 setting those global packets will be sent to the default router, while the link-local destinations would result in a Neighbor Solicitation to resolve the IPv6 to link-layer address binding. Handling such a NS when there are multiple promiscuous ports hence multiple routers risks creating loops. If the router already has a neighbor cache entry for the destination it can respond with an NA on behalf of the destination. However, if it does not it MUST NOT send a NS on the link, since the NA will be received by the other router(s) on the link which can cause an unbounded flood of multicast NS packets (all with hoplimit 255), in particular of the host IPv6 address does not respond. Note that such an NS/NA proxy is defined in [RFC4389] under some topological assumptions such as there being a distinct upstream and downstream direction, which is not the case of two or more peer routers on the same IPPL. For that reason NS/NA packet proxies as in [RFC4389] MUST NOT be used with IPPL.

IPv6 includes Duplicate Address Detection [RFC4862], which assumes that a link-local IPv6 multicast can be received by all hosts which share the same subnet prefix. That is not the case in a private VLAN, hence there could potentially be undetected duplicate IPv6 addresses. However, the DAD proxy approach [RFC6957] defined for split-horizon behavior can safely be used even when there are multiple promiscuous ports hence multiple routers attached to the link, since it does not rely on sending Neighbor Solicitations instead merely gathers state from received packets. The use of [RFC6957] with private VLAN is RECOMMENDED.

The Router Advertisements in a private VLAN MUST be sent out on a promiscuous VLAN ID so that all nodes on the link receive them.

6. IPv4 over IPPL

IPv4 [RFC0791] and ARP [RFC0826] do not have a counterpart to the Neighbor Discovery On-link flag. Hence nodes attached to isolated or community ports will always ARP for any destination which is part of its configured subnet prefix, and those ARP request packets will not be forwarded by the bridges to the target nodes. Thus the routers attached to the promiscuous ports MUST provide a robust proxy ARP mechanism if they are to allow any (firewalled) communication between nodes from different tenants or separation domains.

For the ARP proxy to be robust it MUST avoid loops where router1 attached to the link sends an ARP request which is received by router2 (also attached to the link), resulting in an ARP request from router2 to be received by router1. Likewise, it MUST avoids a similar loop involving IP packets, where the reception of an IP packet results in sending a ARP request from router1 which is proxied by router2. At a minimum, the reception of an ARP request MUST NOT result in sending an ARP request, and the routers MUST either be configured to know each others MAC addresses, or receive the VLAN tagged packets so they can avoid proxying when the packet is received on with the promiscuous VLAN ID. Note that should there be an IP forwarding loop due to proxying back and forth, the IP TTL will expire avoiding unlimited loops.

Any proxy ARP approach MUST work correctly with Address Conflict Detection [RFC5227]. ACD depends on ARP probes only receiving responses if there is a duplicate IP address, thus the ARP probes MUST NOT be proxied. These ARP probes have a Sender Protocol Address of zero, hence they are easy to identify.

When proxying an ARP request (with a non-zero Sender Protocol Address) the router needs to respond by placing its own MAC address in the Sender Hardware Address field. When there are multiple routers attached to the private VLAN this will not only result in multiple ARP replies for each ARP request, those replies would have a different Sender Hardware Address. That might seem surprising to the requesting node, but does not cause an issue with ARP implementations that follow the pseudo-code in [RFC0826].

If the two or more routers attached to the private VLAN implement VRRP [RFC5798] the routers MAY use their VRRP MAC address as the Sender Hardware Address in the proxied ARP replies, since this reduces the risk nodes that do not follow the pseudo-code in [RFC0826]. However, if they do so it can cause flapping of the MAC tables in the bridges between the routers and the ARPing node. Thus such use is NOT RECOMMENDED in general topologies of bridges but can be used when there are no intervening bridges.

7. Multiple routers

In addition to the above issues when multiple routers are attached to the same PVLAN, the routers need to avoid potential routing loops for packets entering the subnet. When such a packet arrives the router might need to send a ARP request (or Neighbor Solicitation) for the host, which can trigger the other router to send a proxy ARP (or Neighbor Advertisement). The host, if present, will also respond to the ARP/NS. This issue is described in [PVLAN-HOSTING] in the particular case of HSRP.

When multiple routers are attached to the same PVLAN, wheter they are using VRRP, HSRP, or neither, they SHOULD NOT proxy ARP/ND respond to a request from another router. At a minimum a router MUST be configurable with a list of IP addresses to which it should not proxy respond. Thus the user can configure that list with the IP address(es) of the other router(s) attached to the PVLAN.

8. Multicast over IPPL

Layer 2 multicast or broadcast is used by protocols like ARP [RFC0826], IPv6 Neighbor Discovery [RFC4861] and Multicast DNS [RFC6762] with link-local scope. The first two have been discussed above.

Multicast DNS can be handled by implementing using some proxy such as [I-D.ietf-dnssd-hybrid] but that is outside of the scope of this document.

IP Multicast which spans across multiple IP links and that have senders that are on community or isolated ports require additional forwarding mechanisms in the routers that are attached to the promiscuous ports, since the routers need to forward such packets out to any allowed receivers in the private VLAN without resulting in packet duplication. For multicast senders on isolated ports such forwarding would result in the sender potentially receiving the packet it transmitted. For multicast senders on community ports, any receivers in the same community VLAN are subject to receiving duplicate packets; one copy directly from layer 2 from the sender and a second copy forwarded by the multicast router.

For that reason it is NOT RECOMMENDED to configure outbound multicast forwarding from private VLANs.

9. Security Considerations

In general DAD is subject to a Denial of Service attack since a malicious host can claim all the IPv6 addresses [RFC3756]. Same issue applies to IPv4/ARP when Address Conflict Detection [RFC5227] is implemented.

10. IANA Considerations

There are no IANA actions needed for this document.

11. Acknowledgements

The author is grateful for the comments from Mikael Abrahamsson, Fred Baker, Wes Beebee, Hemant Singh, Dave Thaler, and Sowmini Varadhan.

12. References

12.1. Normative References

[RFC0791] Postel, J., "Internet Protocol", STD 5, RFC 791, DOI 10.17487/RFC0791, September 1981, <http://www.rfc-editor.org/info/rfc791>.

[RFC0826] Plummer, D., "Ethernet Address Resolution Protocol: Or Converting Network Protocol Addresses to 48.bit Ethernet Address for Transmission on Ethernet Hardware", STD 37, RFC 826, DOI 10.17487/RFC0826, November 1982, <http://www.rfc-editor.org/info/rfc826>.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <http://www.rfc-editor.org/info/rfc2119>.

[RFC2460] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", RFC 2460, DOI 10.17487/RFC2460, December 1998, <http://www.rfc-editor.org/info/rfc2460>.

[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, <http://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, <http://www.rfc-editor.org/info/rfc4862>.

[RFC6957] Costa, F., Combes, J-M., Ed., Pougnard, X., and H. Li, "Duplicate Address Detection Proxy", RFC 6957, DOI 10.17487/RFC6957, June 2013, <http://www.rfc-editor.org/info/rfc6957>.

12.2. Informative References

[DOCSIS-MULPI] "DOCSIS 3.0: MAC and Upper Layer Protocols Interface Specification", August 2015, <http://www.cablelabs.com/wp-content/uploads/specdocs/CM-SP-MULPIv3.0-I28-150827.pdf>.

[I-D.ietf-dnssd-hybrid] Cheshire, S., "Hybrid Unicast/Multicast DNS-Based Service Discovery", draft-ietf-dnssd-hybrid-03 (work in progress), February 2016.

[PVLAN-HOSTING] "PVLANs in a Hosting Environment", March 2010, <https://puck.nether.net/pipermail/cisco-nsp/2010-March/068469.html>.

[RFC3315] Droms, R., Ed., Bound, J., Volz, B., Lemon, T., Perkins, C., and M. Carney, "Dynamic Host Configuration Protocol for IPv6 (DHCPv6)", RFC 3315, DOI 10.17487/RFC3315, July 2003, <http://www.rfc-editor.org/info/rfc3315>.

[RFC3756] Nikander, P., Ed., Kempf, J., and E. Nordmark, "IPv6 Neighbor Discovery (ND) Trust Models and Threats", RFC 3756, DOI 10.17487/RFC3756, May 2004, <http://www.rfc-editor.org/info/rfc3756>.

[RFC4389] Thaler, D., Talwar, M., and C. Patel, "Neighbor Discovery Proxies (ND Proxy)", RFC 4389, DOI 10.17487/RFC4389, April 2006, <http://www.rfc-editor.org/info/rfc4389>.

[RFC4562] Melsen, T. and S. Blake, "MAC-Forced Forwarding: A Method for Subscriber Separation on an Ethernet Access Network", RFC 4562, DOI 10.17487/RFC4562, June 2006, <http://www.rfc-editor.org/info/rfc4562>.

[RFC5227] Cheshire, S., "IPv4 Address Conflict Detection", RFC 5227, DOI 10.17487/RFC5227, July 2008,

<http://www.rfc-editor.org/info/rfc5227>.

[RFC5517] HomChaudhuri, S. and M. Foschiano, "Cisco Systems' Private VLANs: Scalable Security in a Multi-Client Environment", RFC 5517, DOI 10.17487/RFC5517, February 2010, <http://www.rfc-editor.org/info/rfc5517>.

[RFC5798] Nadas, S., Ed., "Virtual Router Redundancy Protocol (VRRP) Version 3 for IPv4 and IPv6", RFC 5798, DOI 10.17487/RFC5798, March 2010, <http://www.rfc-editor.org/info/rfc5798>.

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

[TR-101] "Migration to Ethernet-Based DSL Aggregation", The Broadband Forum Technical Report TR-101, July 2011, <http://www.broadband-forum.org/technical/download/TR-101_Issue-2.pdf>.

Author's Address

Erik Nordmark
Arista Networks
Santa Clara, CA
USA

Email: nordmark@arista.com

Internet Engineering Task Force R. Winter
Internet-Draft M. Faath
Intended status: Informational F. Weisshaar
Expires: April 21, 2016 University of Applied Sciences Augsburg
October 19, 2015

Considerations for IP broadcast and multicast protocol designers draft-winfaa-broadcast-consider-01

Abstract

A number of application-layer protocols make use of IP broadcasts or multicast messages for functions such as local service discovery or name resolution. Some of these functions can only be implemented efficiently using such mechanisms. When using broadcasts or multicast messages, a passive observer in the same broadcast domain can trivially record these messages and analyze their content. Therefore, broadcast/multicast protocol designers need to take special care when designing their protocols.

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 http://datatracker.ietf.org/drafts/current/.

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

This Internet-Draft will expire on April 21, 2016.

Copyright Notice

Table of Contents

1. Introduction

Broadcast and multicast messages have a large receiver group by design. Because of that, these two mechanisms are vital for a number of basic network functions such as auto-configuration. Application developers use broadcast/multicast messages to implement things like local service or peer discovery and it appears that an increasing number of applications make use of it.

Using broadcast/multicast can become problematic if the information that is being distributed can be regarded as sensitive or when the information that is distributed by multiple of these protocols can be correlated in a way that sensitive data can be derived. This is clearly true for any protocol really, but broadcast/multicast is special in two respects: a) the aforementioned large receiver group which makes it trivial for anybody on a LAN to collect the information without special priviledges or a special location in the network and b) encryption is more difficult when broadcasting/ multicasting messages. This draft documents a number of design considerations for broadcast/multicast protocol designers that are intended to reduce the likelyhood that a broadcast protocol can be misused to collect sensitive data about devices, users and groups of users on a LAN.

1.1. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

2. Design considerations

There are a few obvious and a few not necessarily obvious things designers of broadcast/multicast protocols should consider. Most of these items are based on protocol behaviour observed as part of an experiment on an operational network.

2.1. Message frequency

Frequent broadcast/multicast traffic caused by an application can give user behaviour and online times away. This allows a passive observer to potentially decuct a user’s current activity (e.g. a game) and it allows to create an online profile (i.e. times the user is on the network). The higher the frequency of these messages, the more accurate this profile will be. Given that broadcasts are only visible in the same broadcast domain, these messages also give the rough location of the user away (e.g. a campus or building).

If a protocol relies on frequent or periodic broadcast/multicast messages, the frequency should be chosen conservatively, in particular if the messages contain persisten identifiers.

2.2. Persistent identifiers

A few broadcast/multicast protocols observed in the wild make use of persistent identifiers. This includes the use of hostnames or more abstract persistent identifiers such as a UUID or similar. These IDs e.g. identify the installation of a certain application and might not change across updates of the software. This allows a passive observer to track a user precisely if broadcast/multicast messages are frequent. This is even true, in case the IP and/or MAC address changes. Such identifiers also allow two different interfaces (e.g. Wifi and Ethernet) to be correlated to the same device. If the application makes use of persitent identifiers for multiple installations of the same application for the same user, this even allows to infer that different devices belong to the same user.

If a protocol relies on IDs to be transmitted, it should be considered if frequent ID rotations are possible in order to make user tracking more difficult.

2.3. Anticipate user behaviour

A large number of users name their device after themselves, either using their first name, last name or both. Often a hostname includes the type, model or maker of a device, its function or includes language specific information. Based on gathered data, this appears to currently be prevalent user behaviour. For protocols using the hostname as part of the messages, this clearly will reveal personally identifiable information to everyone on the local network.

Where possible, the use of hostnames in broadcast/multicast protocols should be avoided. If only a persistent ID is needed, this can be generated. An application might want to display the information it will broadcast on the LAN at install/config time, so the user is at least aware of the application's behaviour.

2.4. Rember - You are not alone

A large number of services and applications make use of the broadcast/multicast mechanism. That means there are various sources of information that are easily accessible by a passive observer. In isolation, the information these protocols reveal might seem harmless, but given multiple such protocols, it might be possible to correlate this information. E.g. a protocol that uses frequent messages including a UUID to identify the particular installation does not give the identity of the user away. But a single message including the user's hostname might just do that and it can be correlated using e.g. the MAC address of the device's interface.

A broadcast protocol designer should be aware of the fact that even if the protocol's information seems harmless, there might be ways to correlate that information with other broadcast protocol information to reveal sensitive information about a user.

2.5. Configurability

A lot of applications and services using broadcast protocols do not include the means to declare "safe" environments (e.g. based on the SSID of a WiFi network). E.g. a device connected to a public WiFi will likely broadcast the same information as when connected to the home network. It would be beneficial if certain behaviour could be restricted to "safe" environments.

An application developer making use of broadcasts as part of the application should make the broadcast feature, if possible, configurable, so that potentially sensitive information does not leak on public networks.

3. IANA Considerations

This memo includes no request to IANA.

4. Security Considerations

TBD

5. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

Appendix A. Additional Stuff

This becomes an Appendix.

Authors' Addresses

Rolf Winter
University of Applied Sciences Augsburg
Augsburg
DE

Email: rolf.winter@hs-augsburg.de

Michael Faath
University of Applied Sciences Augsburg
Augsburg
DE

Email: michael.faath@hs-augsburg.de

Fabian Weisshaar
University of Applied Sciences Augsburg
Augsburg
DE

Email: fabian.weisshaar@hs-augsburg.de

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: August 3, 2016

X. Xu
Huawei Technologies
R. Asati
Cisco Systems
T. Herbert
Facebook
L. Yong
Huawei USA
Y. Lee
Comcast
Y. Fan
China Telecom
I. Beijnum
Institute IMDEA Networks
January 31, 2016

Encapsulating IP in UDP
draft-xu-intarea-ip-in-udp-03

Abstract

Existing Softwire encapsulation technologies are not adequate for efficient load balancing of Softwire service traffic across IP networks. This document specifies additional Softwire encapsulation technology, referred to as IP-in-UDP (User Datagram Protocol), which can facilitate the load balancing of Softwire service traffic across IP networks.

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 http://datatracker.ietf.org/drafts/current/.

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

This Internet-Draft will expire on August 3, 2016.

Table of Contents

1. Introduction

To fully utilize the bandwidth available in IP networks and/or facilitate recovery from a link or node failure, load balancing of traffic over Equal Cost Multi-Path (ECMP) and/or Link Aggregation Group (LAG) across IP networks is widely used. [RFC5640] describes a method for improving the load balancing efficiency in a network carrying Softwire Mesh service [RFC5565] over Layer Two Tunneling Protocol - Version 3 (L2TPv3) [RFC3931] and Generic Routing Encapsulation (GRE) [RFC2784] encapsulations. However, this method requires core routers to perform hash calculation on the "load-balancing" field contained in tunnel encapsulation headers (i.e., the Session ID field in L2TPv3 headers or the Key field in GRE headers), which is not widely supported by existing core routers.

Most existing routers in IP networks are already capable of distributing IP traffic "microflows" [RFC2474] over ECMP paths and/or

LAG based on the hash of the five-tuple of User Datagram Protocol (UDP) [RFC0768] and Transmission Control Protocol (TCP) packets (i.e., source IP address, destination IP address, source port, destination port, and protocol). By encapsulating the Softwire service traffic into an UDP tunnel and using the source port of the UDP header as an entropy field, the existing load-balancing capability as mentioned above can be leveraged to provide fine-grained load-balancing of Softwire service traffic traffic over IP networks. This is similar to why LISP [RFC6830] , MPLS-in-UDP [RFC7510] and VXLAN [RFC7348] use UDP encapsulation. Therefore, this specification defines an IP-in-UDP encapsulation method dedicated for Softwire service (including both mesh and hub-spoke modes).

IPv6 flow label has been proposed as an entropy field for load balancing in IPv6 network environment [RFC6438]. However, as stated in [RFC6936], the end-to-end use of flow labels for load balancing is a long-term solution and therefore the use of load balancing using the transport header fields would continue until any widespread deployment is finally achieved. As such, IP-in-UDP encapsulation would still have a practical application value in the IPv6 networks during this transition timeframe.

Similarly, the IP-in-UDP encapsulation format defined in this document by itself cannot ensure the integrity and privacy of data packets being transported through the IP-in-UDP tunnels and cannot enable the tunnel decapsulators to authenticate the tunnel encapsulator. Therefore, in the case where any of the above security issues is concerned, the IP-in-UDP SHOULD be secured with IPsec [RFC4301] or DTLS [RFC6347]. For more details, please see Section 6 of Security Considerations.

1.1. Conventions

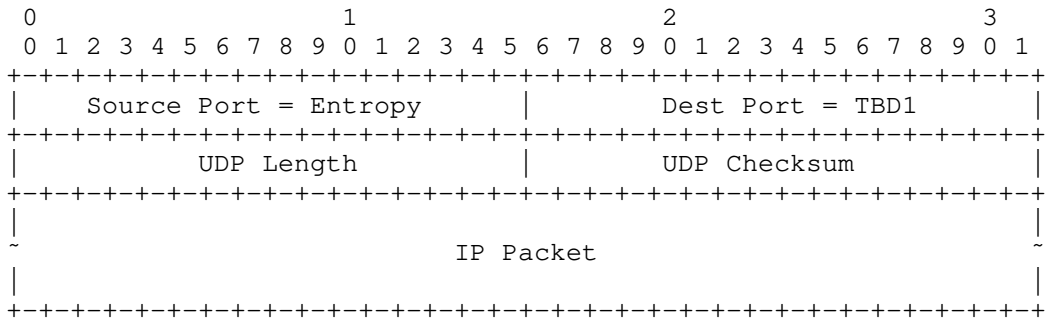
The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

2. Terminology

This memo makes use of the terms defined in [RFC5565].

3. Encapsulation in UDP

IP-in-UDP encapsulation format is shown as follows:



Source Port of UDP

This field contains a 16-bit entropy value that is generated by the encapsulator to uniquely identify a flow. What constitutes a flow is locally determined by the encapsulator and therefore is outside the scope of this document. What algorithm is actually used by the encapsulator to generate an entropy value is outside the scope of this document.

In case the tunnel does not need entropy, this field of all packets belonging to a given flow SHOULD be set to a randomly selected constant value so as to avoid packet reordering.

To ensure that the source port number is always in the range 49152 to 65535 (Note that those ports less than 49152 are reserved by IANA to identify specific applications/protocols) which may be required in some cases, instead of calculating a 16-bit hash, the encapsulator SHOULD calculate a 14-bit hash and use those 14 bits as the least significant bits of the source port field while the most significant two bits SHOULD be set to binary 11. That still conveys 14 bits of entropy information which would be enough as well in practice.

Destination Port of UDP

This field is set to a value (TBD1) allocated by IANA to indicate that the UDP tunnel payload is an IP packet. As for whether the encapsulated IP packet is IPv4 or IPv6, it would be determined according to the Version field in the IP header of the encapsulated IP packet.

UDP Length

The usage of this field is in accordance with the current UDP specification [RFC0768].

UDP Checksum

For IPv4 UDP encapsulation, this field is RECOMMENDED to be set to zero for performance or implementation reasons because the IPv4 header includes a checksum and use of the UDP checksum is optional with IPv4. For IPv6 UDP encapsulation, the IPv6 header does not include a checksum, so this field MUST contain a UDP checksum that MUST be used as specified in [RFC0768] and [RFC2460] unless one of the exceptions that allows use of UDP zero-checksum mode (as specified in [RFC6935]) applies.

IP Packet

This field contains one IP packet.

4. Processing Procedures

This IP-in-UDP encapsulation causes E-IP[RFC5565] packets to be forwarded across an I-IP [RFC5565] transit core via "UDP tunnels". While performing IP-in-UDP encapsulation, an ingress AFBR (e.g. PE router) would generate an entropy value and encode it in the Source Port field of the UDP header. The Destination Port field is set to a value (TBD1) allocated by IANA to indicate that the UDP tunnel payload is an IP packet. Transit routers, upon receiving these UDP encapsulated IP packets, could balance these packets based on the hash of the five-tuple of UDP packets. Egress AFBRs receiving these UDP encapsulated IP packets MUST decapsulate these packets by removing the UDP header and then forward them accordingly (assuming that the Destination Port was set to the reserved value pertaining to IP).

Similar to all other Softwire tunneling technologies, IP-in-UDP encapsualtion introduces overheads and reduces the effective Maximum Transmision Unit (MTU) size. IP-in-UDP encapsulation may also impact Time-to-Live (TTL) or Hop Count (HC) and Differentiated Services (DSCP). Hence, IP-in-UDP MUST follow the corresponding procedures defined in [RFC2003].

Ingress AFBRs MUST NOT fragment I-IP packets (i.e., UDP encapsulated IP packets), and when the outer IP header is IPv4, ingress AFBRs MUST set the DF bit in the outer IPv4 header. It is strongly RECOMMENDED that I-IP transit core be configured to carry an MTU at least large enough to accommodate the added encapsulation headers. Meanwhile, it is strongly RECOMMENDED that Path MTU Discovery [RFC1191] [RFC1981] is used to prevent or minimize fragmentation. Once an ingress AFBR needs to perform fragmentation on an E-IP packet before encapsulating, it MUST use the same source UDP port for all fragmented packets so as to ensures these fragmented packets are

always forwarded on the same path. In a word, IP-in-UDP is just applicable in those Softwire network environments where fragmentation on the tunnel layer is not needed.

5. Congestion Considerations

Section 3.1.3 of [RFC5405] discussed the congestion implications of UDP tunnels. As discussed in [RFC5405], because other flows can share the path with one or more UDP tunnels, congestion control [RFC2914] needs to be considered. As specified in [RFC5405]:

> "IP-based traffic is generally assumed to be congestion-controlled, i.e., it is assumed that the transport protocols generating IP-based traffic at the sender already employ mechanisms that are sufficient to address congestion on the path. Consequently, a tunnel carrying IP-based traffic should already interact appropriately with other traffic sharing the path, and specific congestion control mechanisms for the tunnel are not necessary".

Since IP-in-UDP is only used to carry IP traffic which is generally assumed to be congestion controlled by the transport layer, it generally does not need additional congestion control mechanisms.

6. Security Considerations

The security problems faced with the IP-in-UDP tunnel are exactly the same as those faced with IP-in-IP [RFC2003] and IP-in-GRE tunnels [RFC2784]. In other words, the IP-in-UDP tunnel as defined in this document by itself cannot ensure the integrity and privacy of data packets being transported through the IP-in-UDP tunnel and cannot enable the tunnel decapsulator to authenticate the tunnel encapsulator. In the case where any of the above security issues is concerned, the IP-in-UDP tunnel SHOULD be secured with IPsec or DTLS. IPsec was designed as a network security mechanism and therefore it resides at the network layer. As such, if the tunnel is secured with IPsec, the UDP header would not be visible to intermediate routers anymore in either IPsec tunnel or transport mode. As a result, the meaning of adopting the IP-in-UDP tunnel as an alternative to the IP-in-GRE or IP-in-IP tunnel is lost. By comparison, DTLS is better suited for application security and can better preserve network and transport layer protocol information. Specifically, if DTLS is used, the destination port of the UDP header will be filled with a value (TBD2) indicating IP with DTLS and the source port can still be used as an entropy field for load-sharing purposes.

If the tunnel is not secured with IPsec or DTLS, some other method should be used to ensure that packets are decapsulated and forwarded

by the tunnel tail only if those packets were encapsulated by the tunnel head. If the tunnel lies entirely within a single administrative domain, address filtering at the boundaries can be used to ensure that no packet with the IP source address of a tunnel endpoint or with the IP destination address of a tunnel endpoint can enter the domain from outside. However, when the tunnel head and the tunnel tail are not in the same administrative domain, this may become difficult, and filtering based on the destination address can even become impossible if the packets must traverse the public Internet. Sometimes only source address filtering (but not destination address filtering) is done at the boundaries of an administrative domain. If this is the case, the filtering does not provide effective protection at all unless the decapsulator of an IP-in-UDP validates the IP source address of the packet.

7. IANA Considerations

One UDP destination port number indicating IP needs to be allocated by IANA:

Service Name: IP-in-UDP

Transport Protocol(s): UDP

Assignee: IESG <iesg@ietf.org>

Contact: IETF Chair <chair@ietf.org>.

Description: Encapsulate IP packets in UDP tunnels.

Reference: This document.

Port Number: TBD1 -- To be assigned by IANA.

One UDP destination port number indicating IP with DTLS needs to be allocated by IANA:

Service Name: IP-in-UDP-with-DTLS

Transport Protocol(s): UDP

Assignee: IESG <iesg@ietf.org>

Contact: IETF Chair <chair@ietf.org>.

Description: Encapsulate IP packets in UDP tunnels with DTLS.

Reference: This document.

Port Number: TBD2 -- To be assigned by IANA.

8. Acknowledgements

Thanks to Vivek Kumar, Carlos Pignataro and Mark Townsley for their valuable comments on the initial idea of this document. Thanks to Andrew G. Malis for their valuable comments on this document.

9. References

9.1. Normative References

[RFC0768] Postel, J., "User Datagram Protocol", STD 6, RFC 768, DOI 10.17487/RFC0768, August 1980, <http://www.rfc-editor.org/info/rfc768>.

[RFC1191] Mogul, J. and S. Deering, "Path MTU discovery", RFC 1191, DOI 10.17487/RFC1191, November 1990, <http://www.rfc-editor.org/info/rfc1191>.

[RFC1981] McCann, J., Deering, S., and J. Mogul, "Path MTU Discovery for IP version 6", RFC 1981, DOI 10.17487/RFC1981, August 1996, <http://www.rfc-editor.org/info/rfc1981>.

[RFC2003] Perkins, C., "IP Encapsulation within IP", RFC 2003, DOI 10.17487/RFC2003, October 1996, <http://www.rfc-editor.org/info/rfc2003>.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <http://www.rfc-editor.org/info/rfc2119>.

[RFC2460] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", RFC 2460, DOI 10.17487/RFC2460, December 1998, <http://www.rfc-editor.org/info/rfc2460>.

[RFC2784] Farinacci, D., Li, T., Hanks, S., Meyer, D., and P. Traina, "Generic Routing Encapsulation (GRE)", RFC 2784, DOI 10.17487/RFC2784, March 2000, <http://www.rfc-editor.org/info/rfc2784>.

[RFC4301] Kent, S. and K. Seo, "Security Architecture for the Internet Protocol", RFC 4301, DOI 10.17487/RFC4301, December 2005, <http://www.rfc-editor.org/info/rfc4301>.

[RFC5405] Eggert, L. and G. Fairhurst, "Unicast UDP Usage Guidelines for Application Designers", BCP 145, RFC 5405, DOI 10.17487/RFC5405, November 2008, <http://www.rfc-editor.org/info/rfc5405>.

[RFC6347] Rescorla, E. and N. Modadugu, "Datagram Transport Layer Security Version 1.2", RFC 6347, DOI 10.17487/RFC6347, January 2012, <http://www.rfc-editor.org/info/rfc6347>.

[RFC6935] Eubanks, M., Chimento, P., and M. Westerlund, "IPv6 and UDP Checksums for Tunneled Packets", RFC 6935, DOI 10.17487/RFC6935, April 2013, <http://www.rfc-editor.org/info/rfc6935>.

[RFC6936] Fairhurst, G. and M. Westerlund, "Applicability Statement for the Use of IPv6 UDP Datagrams with Zero Checksums", RFC 6936, DOI 10.17487/RFC6936, April 2013, <http://www.rfc-editor.org/info/rfc6936>.

9.2. Informative References

[RFC2474] Nichols, K., Blake, S., Baker, F., and D. Black, "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers", RFC 2474, DOI 10.17487/RFC2474, December 1998, <http://www.rfc-editor.org/info/rfc2474>.

[RFC2914] Floyd, S., "Congestion Control Principles", BCP 41, RFC 2914, DOI 10.17487/RFC2914, September 2000, <http://www.rfc-editor.org/info/rfc2914>.

[RFC3931] Lau, J., Ed., Townsley, M., Ed., and I. Goyret, Ed., "Layer Two Tunneling Protocol - Version 3 (L2TPv3)", RFC 3931, DOI 10.17487/RFC3931, March 2005, <http://www.rfc-editor.org/info/rfc3931>.

[RFC5565] Wu, J., Cui, Y., Metz, C., and E. Rosen, "Softwire Mesh Framework", RFC 5565, DOI 10.17487/RFC5565, June 2009, <http://www.rfc-editor.org/info/rfc5565>.

[RFC5640] Filsfils, C., Mohapatra, P., and C. Pignataro, "Load-Balancing for Mesh Softwires", RFC 5640, DOI 10.17487/RFC5640, August 2009, <http://www.rfc-editor.org/info/rfc5640>.

[RFC6438] Carpenter, B. and S. Amante, "Using the IPv6 Flow Label for Equal Cost Multipath Routing and Link Aggregation in Tunnels", RFC 6438, DOI 10.17487/RFC6438, November 2011, <http://www.rfc-editor.org/info/rfc6438>.

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

[RFC7348] Mahalingam, M., Dutt, D., Duda, K., Agarwal, P., Kreeger, L., Sridhar, T., Bursell, M., and C. Wright, "Virtual eXtensible Local Area Network (VXLAN): A Framework for Overlaying Virtualized Layer 2 Networks over Layer 3 Networks", RFC 7348, DOI 10.17487/RFC7348, August 2014, <http://www.rfc-editor.org/info/rfc7348>.

[RFC7510] Xu, X., Sheth, N., Yong, L., Callon, R., and D. Black, "Encapsulating MPLS in UDP", RFC 7510, DOI 10.17487/RFC7510, April 2015, <http://www.rfc-editor.org/info/rfc7510>.

Authors' Addresses

Xiaohu Xu
Huawei Technologies
No.156 Beiqing Rd
Beijing 100095
CHINA

Phone: +86-10-60610041
Email: xuxiaohu@huawei.com

Rajiv Asati
Cisco Systems
7200 Kit Creek Road
Research Triangle Park,, NC 27709
USA

Email: rajiva@cisco.com

Tom Herbert
Facebook
1 Hacker Way,
Menlo Park, CA 94052
USA

Email: tom@herbertland.com

Lucy Yong
Huawei USA
5340 Legacy Dr
Plano, TX 75025
USA

Email: Lucy.yong@huawei.com

Yiu Lee
Comcast
One Comcast Center
Philadelphia, PA
USA

Email: Yiu_Lee@Cable.Comcast.com

Yongbing Fan
China Telecom
Guangzhou
CHINA

Email: fanyb@gsta.com

Iljitsch van Beijnum
Institute IMDEA Networks
Avda. del Mar Mediterraneo, 22
Leganes,, Madrid 28918
Spain

Email: iljitsch@muada.com