An extension to RELOAD to support Direct Response and Relay Peer routing
draft-jiang-p2psip-relay-01.txt

Status of this Memo

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

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt.

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html.

This Internet-Draft will expire on June 30, 2009.

Copyright Notice

Copyright (c) 2008 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust’s Legal Provisions Relating to IETF Documents (http://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document.

Abstract

This document proposes an extension to RELOAD to support direct
response and relay peer routing modes. The document starts with the problem statement and addresses concerns about these two routing modes mentioned in RELOAD. Then methods about how to detect whether a node is publicly reachable in P2PSIP situations are discussed. The last part proposes extensions to RELOAD for supporting these additional routing modes.

Table of Contents

1. Introduction .............................................. 3
2. Terminology ............................................. 3
3. Problem Statement ....................................... 4
   3.1. Symmetric Route Stability .......................... 4
   3.2. Application Scenarios ............................... 5
   3.3. Advantages from DRR and RPR ....................... 5
4. Concerns From RELOAD ..................................... 6
   4.1. Concerns About DRR ................................. 6
   4.2. Concerns About RPR ................................. 6
5. Determining If Node Is Publicly Reachable? ............... 7
   5.1. Getting Addresses To Be Used As Candidates for DRR . 7
   5.2. Test On Being Publicly Reachable Or Not .......... 8
6. Discovery Of Relay Peer .................................... 9
7. Relationship Between SRR And DRR/RPR ....................... 9
8. Extensions To RELOAD ..................................... 10
   8.1. Steps For Running DRR/RPR ......................... 10
   8.1.1. Procedure For Running DRR ...................... 10
   8.1.2. Procedure For Running RPR ...................... 11
   8.2. Modification To RELOAD Message Structure ........ 11
   8.3. Request And Response Processing ................... 13
       8.3.1. Sending Peer: Sending Request And Receiving Response .............. 13
       8.3.2. Destination Peer: Receiving Request And Sending Response .......... 13
       8.3.3. What Relay Peer Does? ......................... 14
9. An Example ............................................... 14
10. Security Considerations ................................. 15
11. IANA Considerations .................................... 15
   11.1. A new RELOAD Forwarding Option .................. 15
12. Acknowledgements ...................................... 16
13. References .............................................. 16
   13.1. Normative References .............................. 16
   13.2. Informative References ............................ 16
Authors’ Addresses .......................................... 17
1. Introduction

RELOAD [I-D.ietf-p2psip-base] recommends symmetric recursive routing for routing messages and describes the extensions that would be required to support additional routing algorithms. This document proposes an extension to RELOAD to support two additional routing options: direct response and relay peer which are also discussed in Appendix B in [I-D.ietf-p2psip-base]. However, symmetric recursive routing will not work well in some cases. For example, overlay churn makes the reverse path of the request unstable, therefore fails the transaction. In these cases, direct response or relay peer would be helpful to improve the reliability and efficiency of the P2PSIP transaction. So it is useful to make these three modes work together to provide better service to the upper layer applications on top of the P2PSIP overlay. We analyze the problem statement first and try to address the concerns in RELOAD about these two routing options. Then, some methods are presented to show that it is feasible for a peer to learn whether it is publicly reachable under RELOAD architecture. In section 7 the document discusses the methods to make direct response and relay peer routing mode work together with symmetric routing mode. In the end, an extension to base RELOAD is proposed to make them work.

2. Terminology

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].

We use the terminology and definitions from the Concepts and Terminology for Peer to Peer SIP [I-D.ietf-p2psip-concepts] draft extensively in this document. We also use terms defined in NAT behavior discovery [I-D.ietf-behave-nat-behavior-discovery]. Other terms used in this document are defined inline when used and are also defined below for reference.

There are two types of roles in RELOAD architecture, peer and client. Node is used when describing both peer and client. In specific descriptions to peer or client, peer or client is used instead.

- Direct Response Routing (DRR): refers to a routing mode in which responses to P2PSIP requests are returned to the sending peer directly from the destination peer based on the sending node’s own local transport addresses. For simplicity, the abbreviation DRR is used instead in the following text.
o Relay Peer Routing (RPR): refers to a routing mode in which responses to P2PSIP requests are sent by the destination peer to a relay peer who will forward the responses towards the sending node. For simplicity, the abbreviation RPR is used instead in the following text.

o Symmetric Recursive Routing (SRR): refers to a routing mode in which responses follow the same request path in the reverse order to get back to the sending node. For simplicity, the SRR is used instead in the following text.

o publicly reachable: A node is publicly reachable if it can receive unsolicited messages from any other node in the same overlay. Note: "publicly" does not mean that the nodes MUST be on the public Internet, because RELOAD protocol may be used in a closed system.

o relay peer: a type of publicly reachable peers that can receive unsolicited messages from all other nodes in the overlay and forward the responses from destination peers towards the request sender.

3. Problem Statement

P2PSIP WG intends to make RELOAD work under a number of application scenarios. In reality, the situations where RELOAD is to be deployed differ greatly. For instance, some systems run in an overlay composed of only ten to hundred stable and powerful nodes. On the other hand, overlays may experience heavy churn or comprise of over 1 million nodes. It is obvious that SRR currently adopted in RELOAD works well in the former case. But for the latter case, SRR may not work well.

3.1. Symmetric Route Stability

As discussed in Appendix B.5 in RELAOD [I-D.ietf-p2psip-base], one concern for the motivation to add DRR and RPR is, if there is a heavy churn in overlay, requests fail first. That is, transactions failed before requests reach the destination peer. In this case, DRR and RPR won’t bring any benefit since the transactions in that these two routing modes are used to avoid the churn influence on the response path. RELOAD mentions that retries of the failed requests will reduce the influence of churn. By using optional DRR and RPR the probability of failure due to churn can also be reduced.

In some literatures, for example [ChurnDHT], one way to handle churn in DHT is to get a more accurate timeout value between neighbors. When the timeout fires, the preferred way for the upstream peer to handle the failure is to try the request through another downstream peer, for there are a number of paths between a sending node and a
destination peer. Therefore, the probability for a request to reach the destination peer successfully under churn will improve greatly. This still leave open the issue of the response failure due to churn.

Through the above analysis, we can see that churn in request path could be solved to great extent by using existing mechanisms. DRR and RPR routing mode are still helpful for the responses to be resilient to the churn.

3.2. Application Scenarios

Overlay churn causes nodes or connections between neighbor nodes to be unstable. If SRR is used for routing responses, any failure of nodes or connections in the reverse path will fail the whole P2PSIP transaction. So a transaction with longer response latency will be vulnerable to churn. Below are two examples of overlays which produce slow response.

- P2P technology has been applied successfully in some global public VoIP services, such as Skype. The number of the simultaneous online users for this type of system is often over 1 million. For a typical overlay network, the lookup operation in P2P systems will end up with the hop counts within \([0, \log(N)]\). Here, \(N\) is the number of nodes in the system. For a system where \(N\) is over 1 million, the maximum hop counts would be about 20. For such a long path, the churn of the overlay will break the reverse path with high probability, for the failure of any node in the path would mean the failure of the reverse path, hence the failure of the transaction.
- If links of the overlay network are wireless link or some low speed links, the latency between links would be significantly long. So messages will take longer to traverse the path back and forth and the nodes in the path are more easily subject to failure during the message transportation duration because of churn.

3.3. Advantages from DRR and RPR

DRR and RPR try to shorten the response path to a fixed number of hops and thus reducing the response time. If DRR is used, the response path is only 1 hop; In the case of RPR, the response path is 2 hops. According to the suggestions in this draft, before the peer starts to use the DRR or RPR, a test that decides whether DRR or RPR could work should be performed. In this way, it could mitigate the concerns about influence from the failure of DRR and RPR. The main benefit from adding the option to use DRR and RPR with SRR fall-back is that it will improve the success rate and reliability of the P2PSIP transaction, especially for the scenarios described in section 3.2.
Another advantage of DRR/RPR is that the via-list grows as requests are forwarded around the overlay, so it is highly possible that the size of the requests is larger than MTU and fragmentation will be unavoidable, either IP level fragmentation or RELOAD level fragmentation. DRR and RPR can work without the support of via-list, making the size of messages more predictable and making it easy to handle the fragmentation.

4. Concerns From RELOAD

For DRR and RPR routing modes, RELOAD authors has expressed several concerns in Appendix B.3 and Appendix B.4 in [I-D.ietf-p2psip-base]. This section attempts to address these concerns.

4.1. Concerns About DRR

  o Concern1: Due to NAT existence, DRR is subject to failure. So it requires the destination peer to detect the failure and then fallback to SRR.
    * Answer: The concern is on how DRR works together with SRR. In this draft we suggest tests in which a node can learn whether it is publicly reachable. With this information in mind, the success rate of DRR will improve greatly. On the other hand, a node can decide which routing mode it will try first based on past success rate. After several attempts, the node can transition to other routing mode as a fallback.
  o Concern2: Due to NAT filtering behavior, the NAT would drop the response if it is not a p2p-friendly NAT.
    * Answer: This draft introduces tests that a node should perform before DRR is offered in order to learn whether a node is publicly reachable, so that DRR will succeed with high probability.
  o Concern3: The response happens to be received by another peer who is not the sending node and hence it adds to the transaction latency.
    * Answer: We think this is a rare case since this draft recommends a method to test whether a node is publicly reachable.

4.2. Concerns About RPR

  o Concern1: RPR requires the help from relay peers of the sending node. It is not easy to identify which category a node is in.
    * Answer: RPR does not intend to replace SRR. This draft proposes RPR and DRR as an enhancement to SRR. As for the difficulty in identifying a node’s role, the draft suggests tests for learning whether a node is publicly reachable. Using
eligible peers as relays is also common in p2p network where peers can declare themselves as supernodes if they have a publicly reachable address.

5. Determining If Node Is Publicly Reachable?

For DRR and RPR to function correctly, a node should be able to determine whether it is publicly reachable. If it is not, RPR should be chosen to route the response with the help from relay peers; otherwise, DRR may be offered. NATs and firewalls are two major contributors preventing DRR and RPR from functioning properly.

There are a number of techniques with which a node can get its reflexive address which is on the public side of the NAT. After getting the reflexive address, a peer can perform further tests to learn whether the reflexive address is publicly reachable. If the address proves publicly reachable, the nodes to which the address belongs can use DRR for responses and also can be a candidate for relay peer. Nodes being publicly unreachable use RPR to shorten response path with the help from relay peers.

On the other hand, some conditions are unique in P2PSIP architecture which could be leveraged to facilitate the tests. In P2P overlay network, every node only has partial view of the whole network and then knows a few nodes in the overlay. P2P routing algorithm can easily deliver the request from a sending node to a peer with whom the sending node has no direct connection. So it is easy for a node to get other nodes to send unsolicited message to him.

In this section, we first introduce several ways for a node to get the addresses for the further test. Then a test for learning whether a peer is publicly reachable is proposed.

5.1. Getting Addresses To Be Used As Candidates for DRR

In order to test whether a peer is publicly reachable, the node should first get one or more addresses which will be used by other nodes to send messages directly. This address is either a local address of a node or a translated address which is assigned by NAT to the node.

STUN is used to get a reflexive address on the public side of NAT with the help of STUN servers. There is also a STUN usage [I-D.ietf-behave-nat-behavior-discovery] to discover NAT behavior. Under RELOAD architecture, a few infrastructure servers can be leveraged for this usage, such as enrollment servers, diagnostic servers, bootstrap servers, etc.
The node can use STUN Binding request to one of STUN servers to trigger a STUN Binding response which returns the reflexive address from the server’s perspective. If the reflexive transport address is the same as the source address of Binding request, the node can determine that there is no NAT between him and the chosen infrastructure server. (Certainly, in some rare cases, the allocated address happens to be the same as the source address. Further tests will detect this case and rule it out in the end.). Usually, these infrastructure servers are publicly reachable in the overlay, so the node can be considered as being publicly reachable. On the other hand, with the techniques in [I-D.ietf-behave-nat-behavior-discovery], a node can also decide whether it is behind NAT with endpoint-independent mapping behavior. If the node is behind NAT with endpoint-independent mapping behavior, the reflexive address should also be a candidate for further tests.

UPnP-IGD is a mechanism that a node can use to get the assigned address by its residential gateway and after getting this address and communicating it with other nodes, the node can receive unsolicited messages from outside, even though it is behind a NAT. So the address obtained through the UPnP mechanism should also be used for further tests.

Another way that a node behind NAT can use to learn its assigned address by NAT is NAT-PMP. Like in UPnP-IGD, the learned address should also be tested further.

The above techniques are not exhaustive. For example, MIDCOM SIMCO can also serve the same purpose as UPnP-IGD. These techniques can be used to get candidate transport addresses for further tests.

5.2. Test On Being Publicly Reachable Or Not

Using the transport addresses obtained by means of the above techniques, a node can start a test to learn whether the transport address in question is publicly reachable. The basic idea for the test is for a node to send a request and expect another node in the overlay to send back a response. If the response is received by the sending node successfully and also the node giving the response has no direct connection with the sending node, the sending node can determine that the address is probably publicly reachable and hence the node may be publicly reachable at the tested transport address.

In P2P overlay, a request is routed through the overlay and finally a destination peer will terminate the request and give the response. It is high probability that the destination peer has no direct connection with the sending node. Especially in RELOAD architecture, every node maintains a connection table. So it is easier for a node
to check whether it has direct connection with another node.

Note: Currently, no existing message in base RELOAD can achieve the test. In our opinion, this kind of test is within diagnostic scope, so authors hope WG can define a new diagnostic message to do that. We don't plan to define the message in this document, for the objective of this draft is to propose an extension to support DRR and RPR. The following texts is informative.

If a node wants to test whether its transport address is publicly reachable, it can send a request to the overlay. The routing for the test message will be a little different from other kinds of requests because it is not for storing/fetching something to/from the overlay or locating a specific node, instead it is to get a peer who can deliver the sending node an unsolicited response and must have no direct connection with him. So each intermediate peer receiving the request first checks whether it has direct connections with the sending peer. If there is any, keep routing the request. Otherwise, the intermediate peer terminates the request and sends the response back directly to the sending node with the transport address under test.

After performing the test, if the peer determines that it is publicly reachable, it can try DRR in the later transaction. And it should advertise that it is a candidate for relay peers.

6. Discovery Of Relay Peer

In order to support RPR, a node behind p2p-unfriendly NAT must have one or more relay peers to help him receive response from destination peers. The major requirement for relay peers is that they should be publicly reachable. By performing the test proposed in section 5, a peer could determine whether it is publicly reachable.

There are several means to get the information about relay peers. For example, as proposed in base RELOAD [I-D.ietf-p2psip-base], service discovery for TURN server can also be used to discover relay peers. In some other cases, the service provider may provide a set of relay peers to improve the efficiency of the overlay.

7. Relationship Between SRR And DRR/RPR

The major contribution in this draft is to provide available tools to use DRR/RPR in different situations. In order to make DRR/RPR work well with SRR, we give some suggestions on how to transition between them in a P2PSIP transaction.
Editor's Note: it is not required that the implementation should use the same strategy. The specific implementation can use its own strategy to achieve better performance.

Basically, a peer can collect statistics data on different routing modes based on previous transactions or develop a dedicated component to initiate transactions and collect the related statistics data. For example, a node may have a diagnostic component which sends PING message with different routing mode, either DRR/RPR or SRR. After collecting enough data, the peer will have a clearer view about the success rate of different routing modes. Other than the success rate, the peer can also get data of fine granularity, for example, the number of retransmission the peer needs to get a desirable success rate.

So a typical strategy for the transition is as follows. A node chooses the routing mode with higher success rate first, for example, SRR. And he will try for several times based on the statistics data, i.e., the number of retransmission for the routing mode. After the suggested number of retransmissions, the node can transition to DRR or RPR based on the peer is publicly reachable or not.

If the overlay is run within a private network, and all nodes in the system can reach each other directly, nodes try most of the transactions with DRR. On the other hand, if the relay peer is provided by the service provider, nodes prefer RPR over SRR.

8. Extensions To RELOAD

Adding the support of DRR and RPR requires extensions to the current RELOAD. In this section, first we show the steps taken by a node in order to use DRR and RPR. Then based on the results, we could point out the difference from the current RELOAD and define the required extensions.

8.1. Steps For Running DRR/RPR

DRR and RPR have the similar effect on P2PSIP transaction, i.e. reduce the response path and improve the success rate. There are still minor differences between them. In this section, we will give procedure for running DRR and RPR respectively.

8.1.1. Procedure For Running DRR

After running the tests suggested in section 5, a node can determine whether it is publicly reachable, either reachable at its own local public address or a reflexive address assigned by NAT. If a node is
publicly reachable, it can use DRR for future transaction.

If a node decides to use DRR for a transaction, it MUST include in the request its node-id and associated transport address which has been tested as being publicly reachable. So the destination peer can use the publicly reachable transport address to send back the response to the sending node.

8.1.2. Procedure For Running RPR

If a node has no publicly reachable transport address, it chooses RPR instead of DRR. In order to use RPR, the node MUST get some relay peers first and then establish direct connection with at least one of the relay peers before initiating transaction.

As discussed in section 6, overlay has existing mechanisms to discover relay peers. In RELOAD, it is also easy for a node to use Attach to create a direct connection with relay peer. Relay peer may reject requests based on its current load. (Note: Do we need to add a new error code to describe this kind of error: the node can not accommodate new connections?) So the node will try other relay peers till getting a direct connection with relay peer successfully.

When the node decides to use RPR for a transaction, it MUST include in the request the relay peer’s node-id and its associated publicly reachable transport address. The sending node’s node-id also MUST be included in the request. Upon receiving a request demanding RPR, the destination peer will send the response directly to the relay peer with the relay peer’s publicly reachable transport address. Then the relay peer will forward the response to the sending peer through the existing connection between the relay peer and sending peer.

8.2. Modification To RELOAD Message Structure

RELOAD provides an extensible framework to accommodate future extensions. This proposal defines a new forwarding option in the ForwardingHeader structure to allow DRR and RPR. In this section, we introduce this new forwarding option. The normative procedures which the peers have to follow for the new option are proposed in the following section.

RELOAD supports state-keeping style to realize SRR. If DRR or RPR is to be used, the response will not follow the reverse path so that the state in the intermediate peers won’t be cleared until states expire. In order to address this issue, the proposal intends to define a new flag, state-keeping flag, in the message header to indicate whether the state will be allowed to maintain in the intermediate peers.
Editor’s note: this flag will be useful for other cases so the authors suggest adding it in the base draft.

The new forwarding option is named Extensive_Routing_Mode and conforms to the structure defined in section 5.2.2.3 in [I-D.ietf-p2psip-base]".

type : 0x1 extensive_routing_mode

flag: 0x02 DESTINATION_CRITICAL

The option value will be illustrated in the following figure.

```c
enum { 0x0, 0x01 (DRR), 0x02(RPR), 255} RouteMode;

struct {
    RouteMode routemode;
    Transport transport;
    Destination destination<1..2>;
    IpAddressPort ipaddressport;
} Extensive_Routing_Mode;
```

The above structure reuses: Transport, Destination and IpAddressPort structure defined in section 5.2.1.1 and 5.2.2.1 in [I-D.ietf-p2psip-base].

Route mode: refers to which kind of routing mode is indicated to the destination peer. Currently, only DRR and RPR are defined.

Transport: refers to the transport type which is used to deliver responses from the destination peer to the sending peer or the relay peer;

Destination: refers to the relay peer or the sending node itself. If the routing mode is DRR, then the destination only contains the sending node’s node-id; If the routing mode is RPR, then the destination contains two destinations, which are the relay peer’s node-id and the sending node’s node-id;

IpAddressPort: refers to the transport address the destination peer is using to send the response to.
8.3. Request And Response Processing

This section gives normative text on the normal message processing after DRR and RPR are introduced. Here, we only describe the additional procedures for supporting DRR and RPR and please refer to [I-D.ietf-p2psip-base] for RELOAD base procedures.

8.3.1. Sending Peer: Sending Request And Receiving Response

If a peer can offer DRR, it MUST fill the extensive_routing_mode option first defined in section 8.2 before sending a request out. The routing mode MUST be set to 0x1(DRR); The destination carries the sending node node-id and the ipaddressport field MUST carry only one transport address of the sending node which has been tested.

If a peer can offer RPR, it MUST fill extensive_routing_mode option. Routing mode MUST be set to 0x2(RPR). The destination MUST hold one relay peer’s node-id and the sending node’s node-id in a sequential order so that the destination peer can tell who is who from the order of the destinations. The ipaddressport MUST carry one of the transport addresses of the relay peer.

Upon receiving response, the peer follows the rule in [I-D.ietf-p2psip-base].

8.3.2. Destination Peer: Receiving Request And Sending Response

When the destination peer receives a request, it will check the options in the forwarding header. If the destination peer can not understand extensive_routing_mode option in the request, it tries to use SRR to return a error response to the sending peer.

If the routing mode is DRR, then the peer constructs the Destination lists for the response with only one entry, i.e, the sending peer node-id got from the option in the request; if the routing mode is RPR, the destination peer constructs Destination list for the response with two entries, the relay peer node-id and the sending node node-id which is carried in the option of the request. For RPR, if there is only one Destination in the option, the destination peer tries to send a error response to the sending peer using SRR.

After the peer constructs the destination list for the response, it sends the response to the transport address which is indicated in the ipaddressport field in the option using the specific transport mode in the option.
8.3.3. What Relay Peer Does?

Relay peers are designed to forward responses to nodes who are not publicly reachable. For the routing of the response, this draft still uses the destination list. The only difference from SRR is that the destination list is not the reverse of the via-list, instead it is constructed from the forwarding option.

As the ordinary peer, it receives the response, validate the message, re-adjust the destination-list and forward the response to the next hop in the destination list based on the connection table. There is no added requirement for relay peer.

9. An Example

In this section, we give illustration to show the use of DRR or RPR to route messages. In the following illustration, we use SP for sending peer, IP for intermediate peer, DP for destination peer, RP for relay peer.
The above figures show flow diagrams in the cases where RPR or DRR is used to routing message. From the above figures, the response does not follow the reverse path of the request. Responses either go back directly to the sending peer or go through the relay peer, then return to the sending peer.

10. Security Considerations

TBD

11. IANA Considerations

11.1. A new RELOAD Forwarding Option

A new RELOAD Forwarding Option type is add to the Registry.

Type: 0x1 - extensive_routing_mode
12. Acknowledgements

Authors would like to thank Ted Hardie, Narayanan Vidya and Dondeti Lakshminath for their comments. Special thanks go to Bruce Lowekamp for his constructive comments.

13. References

13.1. Normative References

[I-D.ietf-p2psip-base]

[I-D.ietf-p2psip-concepts]


13.2. Informative References

[ChurnDHT]

[I-D.ietf-behave-nat-behavior-discovery]

[I-D.ietf-behave-tcp]

[I-D.lowekamp-mmusic-ice-tcp-framework]

[ChurnDHT]
draft-lowekamp-mmusic-ice-tcp-framework-00 (work in progress), October 2008.


Authors’ Addresses

XingFeng Jiang
Huawei Tech.
Huihong Mansion, No.91 Baixia Rd.
Nanjing, Jiangsu 210001
P.R.China

Phone: +86(25)84565867
Email: jiang.x.f@huawei.com

Roni Even
Gesher Erove
14 David Hamelech
Tel Aviv 64953
Israel

Email: ron.even.tlv@gmail.com