P2PSIP WG
Internet Draft
Intended status: Informational
Expires: September 2007

E. Cooper
P. Matthews
Avaya

March 4, 2007

The Effect of NATs on P2PSIP Overlay Architecture
draft-matthews-p2psip-nats-and-overlays-01.txt

Status of this Memo

Copyright Notice

Abstract

This paper explains the problems created by Network Address Translators (NATs) in Peer-to-Peer (P2P) overlays and recommends some NAT traversal techniques appropriate for P2PSIP networks. Two P2PSIP overlay architectures that accommodate the presence of NATs are described and analyzed. The first is the super-peer scheme used in a

number of p2p file-sharing systems today. The second is a scheme where all peers play an equal role in the overlay.

Table of Contents

1. Introduction

P2P overlay networks have emerged as a popular, scalable and efficient mechanism for sharing music and video files amongst millions of computers. Early P2P file-sharing systems used a combination of highly distributed content storage and a centralized index of the available content to provide easy access to vast amounts of data. Napster was one such system. Due to the legal implications of its unauthorized content distribution, Napster was ordered to cease operations. As soon as the centralized content index was disabled, the Napster network was effectively shut down.

Of course, Napster's demise did nothing to squelch the demand for easy access to vast amounts of content. New P2P file-sharing systems emerged to replace Napster and all of them were designed without a centralized content index. Various schemes were used to locate

content in these new networks. Some transmitted content queries randomly amongst nodes, while others would flood the queries throughout the network. Neither of these techniques was particularly effective or efficient at locating content. Then P2P overlays adopted a distributed hash table (DHT) approach for indexing their content. As their name suggests, DHTs distribute the content index across many nodes. DHTs do provide an effective and relatively efficient indexing mechanism. The drawback is that as a result of this distribution, a query into a DHT must be processed by a number of nodes before the result can be determined.

As a result of their evolution, current P2P file-sharing overlays use both distributed content storage and distributed context indexing. By eliminating all centralization in the system, these P2P overlays have gained a number of interesting benefits. They are self-organizing, highly scalable, highly reliable and require very little administration.

From a structural perspective, today's SIP networks have a lot in common with the Napster file-sharing network. The 'content' in a SIP system is the real-time data that flows directly between the nodes. Although it doesn't need to be stored anywhere, the source addresses of the 'content' must be collected into an index that can be easily accessed. This index is analogous to the contact binding information that is stored by Registrars. Of course, SIP networks do not face the same legal difficulties as P2P file-sharing systems. They do, however, have scalability, reliability and administrative concerns.

The P2PSIP working group is investigating how P2P techniques might be used in conjunction with (or as an alternative to) the DNS-based lookup and routing mechanisms of [RFC3261] and [RFC3263]. This paper explores the problems introduced by the presence of NATs in the network topology and discusses their effects on the P2PSIP overlay architecture.

Comments on this draft are solicited and should be addressed either to the authors or to the P2P-SIP mailing list at p2psip@ietf.org (see https://www1.ietf.org/mailman/listinfo/p2psip).

2. IP Network Structure

Overlay networks create a set of virtual links on top of the routes of the underlying IP network. These virtual links are usually structured to optimize the overlay for a particular purpose, such as content indexing.

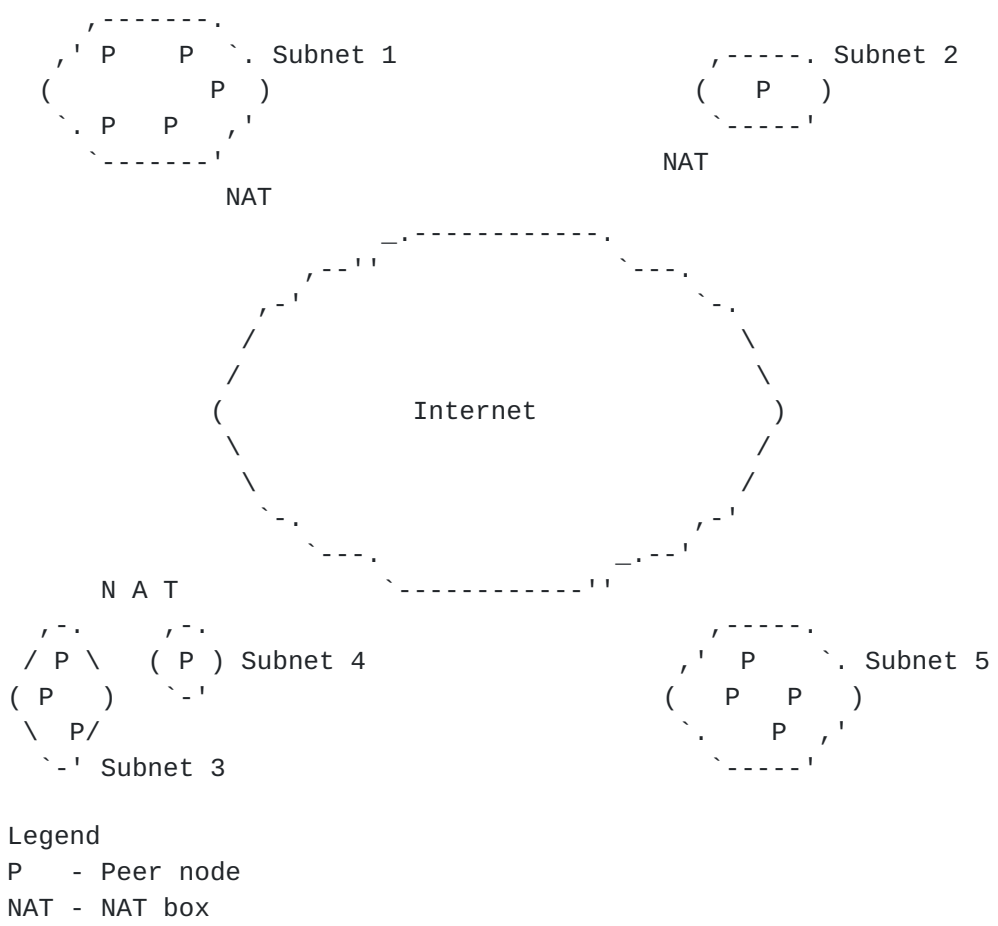

Figure 1 Example IP Network Topology

Figure 1 shows a set of peers (denoted by Ps) that want to create an overlay on top of an IP network. In this figure we see six clouds. Five represent IP subnets containing peers and one represents the Internet. One of the subnets uses public IP addresses, while the other subnets have NATs between them and the Internet and thus use private addresses. Two of the subnets are sitting behind the same NAT. More complex network topologies are not depicted, but it would not be uncommon for an overlay network to include peers that were separated from the Internet by two or more NATs.

2.1. NAT-Induced Problems in Overlay Networks

Some P2P overlays assume that all participating nodes are linked by unimpeded IP connectivity. Unfortunately, the use of NATs is very common, which means that a great many IP networks span multiple addressing spaces.

Straightforward deployment of P2P overlays on IP networks involving NATs would cause the overlay's mechanisms to fail because:

- The private IP addresses of some peers would be considered un-deliverable by the routers in the public Internet. This would cause messages to be discarded.

- The IP addresses of some peers could conflict if the overlay included multiple private address spaces. This would cause messages to be delivered to the wrong peer.

- NATs perform mapping and filtering functions at the borders between two addressing realms, and will frequently discard packets they consider "unsolicited". From a NAT's perspective, a message must be sent from a peer on its "private" side to a peer on its "public" side before a message can travel in the opposite direction.

All these mis-directed and dropped messages will cause overlay services to fail and may prevent the participating peers from constructing or maintaining overlay correctly.

2.2. NAT Traversal Techniques for P2PSIP Overlays

NAT traversal is a well-known problem for SIP networks and much effort has been devoted to solving it. [p2p-comm] discusses some popular mechanisms for P2P systems and recommends a combination of "NAT hole-punching" and "relay" techniques to establish communications between peers in NATed networks.

Based upon the arguments for an UNSAF approach presented in [nat-consider], [ice] defines a mechanism for employing these two techniques in SIP networks to route media streams between two NATed endpoints. The use of ICE and other SIP NAT traversal techniques, such as "symmetric signaling response" and "connection re-use", is encouraged by [nat-scenarios].

Since NATs create similar (and perhaps more severe) problems for P2PSIP overlays, all of these mechanisms will need to be adopted for P2PSIP overlay signaling protocols. Other SIP techniques, such as [outbound], may also prove useful for P2PSIP systems.

The applicability of some other NAT traversal techniques is discussed in APPENDIX A:

3. Super-Peer Overlay Networks

One way to organize a P2P overlay is to create an overlay topology in which the publicly addressable peers (dubbed "super-peers") act as relay points for the NATed peers. This structure essentially creates a hierarchy in the overlay network. The super-peers (at the top level) supply the content indexing function and provide message routing to/from NATed peers. NATed peers are at a lower level. They may advertise their content and query for content via their associated super-peer, but do not process any queries or store any of the content index data.

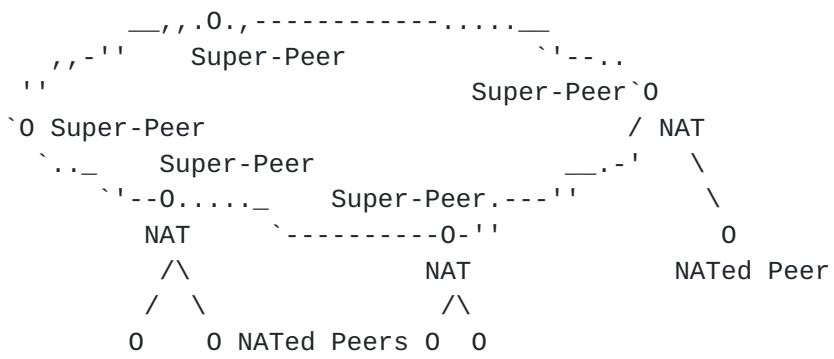

Figure 2 P2P Overlay with Super-Peers

3.1. NAT-Induced Problems in P2PSIP Super-Peer Overlays

The super-peer overlay network organization provides a simple and efficient model for NAT traversal in P2PSIP networks. Its routing structure has the advantage that a message sent from one peer to another traverses at most 3 hops.

The main drawback of this approach is that it requires a sufficient number of publicly addressable nodes to act as super-peers. In addition, the super-peers must bear the entire load associated with message routing.

Several P2PSIP use case scenarios are described in [use-cases]. Referring back to Figure 1, one example of a P2PSIP "Managed Private Network" scenario could include peers from subnets 1-4, but no peers with public addresses. In such a network, a super-peer routing topology is simply not possible.

Other use cases, including the "Public P2P VoIP Service Providers" and "Open Global P2P VoIP Network" scenarios do include peers from all five subnets. These overlay networks will have some percentage of publicly addressable peers. One measurement has found that 74% of

web-browsing clients are behind NATs [illuminati]. If a similar percentage of P2PSIP peers are NATed, a super-peer overlay topology will be able to utilize only 1/4 of the resources available.

The bandwidth available at the super-peers is of particular concern. Since every message in the overlay must traverse the IP links to the super-peers, it's possible that super-peers with low-bandwidth links will be overwhelmed, while high-bandwidth links to NATed peers will be almost completely unused.

Further, the use of a super-peer routing structure requires that each NATed peer must establish a long-lived association to a super-peer. [behave-udp] and [behave-tcp] require the use of periodic "keep-alive" traffic to ensure connectivity across the intervening NAT. It follows that the amount of keep-alive traffic arriving at a given super-peer will be proportional to the number of NATed peers it serves. Thus, in super-peer overlays, it is important to assign NATed peers to super-peers such that only a reasonably small fraction of the super-peer's bandwidth is consumed by keep-alive traffic. In other words, the routing structure should be constructed such that the super-peer's bandwidth is not overwhelmed. A mechanism for distributing the load across super-peers will need to be created.

Another consequence of the super-peer routing structure is that the amount of keep-alive traffic crossing a given NAT will be proportional to the number of peers behind that NAT (regardless of how those peers are distributed across super-peers).

Due to its connectionless nature, the bandwidth considerations are considerably more pronounced for UDP than for TCP. However, TCP connections require more state information to be maintained at the super-peer. Both protocols require state information to be maintained at the NAT.

4. Fully Distributed Overlay Networks

As an alternative to the hierarchy created by super-peer overlays, it is also possible to use the techniques of section 2.2. to create a completely flat overlay network in which all peers are equal participants. Such fully-distributed overlays also avoid the problems created by NATs in the search algorithm (discussed in section 2.1. and the problems in the super-peer routing topology (discussed in section 3.1.

This approach does not place special emphasis on nodes with publicly reachable addresses and can be deployed over any IP network topology.

Since all peers participate in the search scheme and in message routing, all of the available resources can be utilized.

[dSIP-nat] is one example of a fully distributed overlay that uses SIP-based messaging to implement a DHT search algorithm. Fully distributed P2PSIP overlay networks could also be built using other protocols and/or other search algorithms.

4.1.1. Aligning the Search and Routing Structures

Many DHTs route queries amongst peers such that any query can reach the appropriate (authoritative for that query) peer in O(log N) hops. As previously mentioned, the problem is that these searching structures do not account for impediments in IP routing. Creating a routing structure that mirrors the search algorithm will preserve the efficiency of the search algorithm as much as possible.

To determine the specifics of the routing structure, we examine the search algorithm in a bit more detail. Chord is used as an example. To process queries, peers participating in a chord overlay maintain tables of other peers that will assist in routing queries to their destinations. These tables form a good starting point for the routing topology.

In Chord, some unique peer attribute is hashed using SHA-1 and the result (called a peer identifier) is used to place peers on a conceptual ring. Each peer then maintains connections to peers located at exponentially increasing locations going clockwise around the ring. For example, a peer P, with ID 0 might have a table consisting of addresses for peers with IDs 2^0, 2^1, 2^2,..., 2^(n/2) (where n=# of output bits for SHA-1).

In this routing structure, a message to peer Q can be addressed to Q's location in the ring, and any intermediate peer R can forward the message by selecting the peer from its connection table that is that is closest to Q without overshooting Q.

Many other connection structures exist. For example, structured routing topologies can be created using the ideas contained in any one of a number of DHT schemes. The important point is that the structure of the routing topology matches the message flow required by the search algorithm.

4.1.1.1. Symmetric Interest

When considering connection topologies, there is a property we have dubbed "symmetric interest". A connection structure exhibits

"symmetric interest" if, when peer P desires a connection to peer Q, then peer Q also desires a connection to peer P.

A routing structure based on peers randomly selecting other peers to connect to does NOT exhibit symmetric interest because peer P can select peer Q without peer Q selecting peer P. Similarly, a Chord-based connection structure (depicted in Figure 3) also does NOT exhibit symmetric interest because a given peer P in the ring desires connections to peers in the clockwise half-circle but not in the counter-clockwise half-circle.

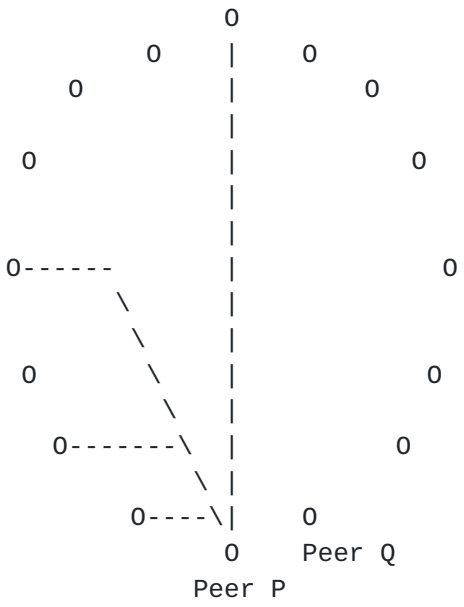

Figure 3 Chord-based Connection Structure 1

Figure 3 depicts a connection topology from the perspective of a single peer P. As described in section 4.1.1. if P's ID were 0, it might have a table consisting of addresses for peers with IDs 2^0, 2^1, 2^2,..., 2^(n/2). Peer P would never include Peer Q in its connection table, since Q's ID is greater 2^(n/2). However, Peer Q's connection table may include Peer P, since P's ID is contained in the clockwise half-circle starting at Q's ID. Figure 4 illustrates the same connection topology from the perspective of Peer Q.

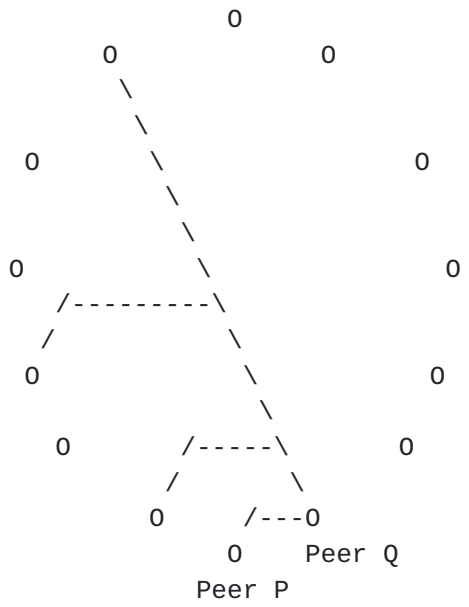

Figure 4 Chord-based Connection Structure 2

One topology that does exhibit symmetric interest has each peer maintaining connections to peers located an exponentially increasing distances going both clockwise AND counter-clockwise around the ring.

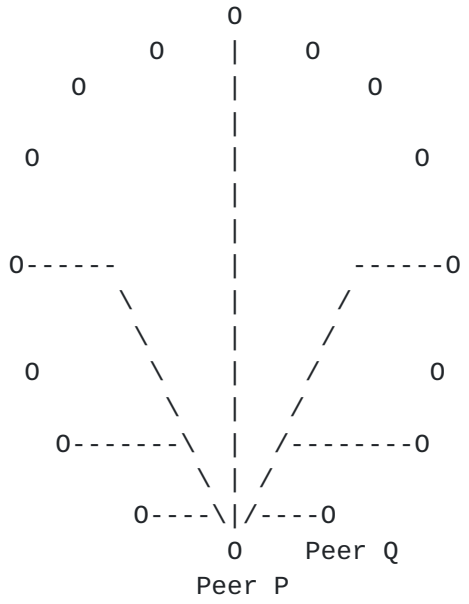

Figure 5 Symmetric Partial Mesh

"Symmetric interest" seems a desirable property for routing topologies because connections through NATs, by their nature, are bi-

directional and because both peers incur the overhead of sending keep-alives to maintain the connection.

4.1.2. NAT-Induced Problems in Fully Distributed Overlays

As mentioned in section 3.1. each routing connection that crosses a NAT must be maintained by P2PSIP peers. This applies to the fully distributed overlay network too.

There is a possibility that the number of viable connections in a peer's table might be constrained by the number of 'pinhole' mapping and filtering entries that can be supported by a peer's local NAT. Unfortunately, NAT behaviour is notoriously variable, so it is difficult to predict the achievable size for a peer's connection table. If the number of entries in this table is reduced below the DHT's prescribed size, the message routing efficiency may be reduced, or fail completely. For example, under a Chord-based routing topology, the connection to the peer's immediate successor is critically important. Without that link, messages may fail to reach their destination. The other connections in a Chord-based structure are used to improve routing efficiency, but some may be removed without jeopardizing routing correctness.

So in a fully distributed overlay, peers may need to reduce the size of their connection tables to accommodate limitations in their local NATs. This can reduce the search algorithm's efficiency.

Interestingly, when large numbers of peers are operating behind the same NAT, DHT-based search algorithms is likely to create many connections that do not need to cross the NAT.

5. Comparing Super-Peer and Fully Distributed Overlay Networks

<< [concepts] describes three modes of operation under which P2PSIP peers can register and make calls. These modes are variations on how user contact information in stored, retrieved and used by peers in the overlay network. A future version of this paper will compare the performance of the super-peer and fully distributed overlays under each mode. >>

6. Other Hierarchical Overlay Network Topologies

6.1. Modified Super-Peer Overlays

As discussed in section 3.1. super-peer routing topologies can encounter difficulties when many peers are behind the same NAT. The resources (bandwidth, state-information) required for NAT traversal

could be reduced if a single "Representative" peer were elected to proxy all the traffic between the NATed peers and the super-peer. An overlay utilizing both super-peers and representatives is depicted in Figure 6.

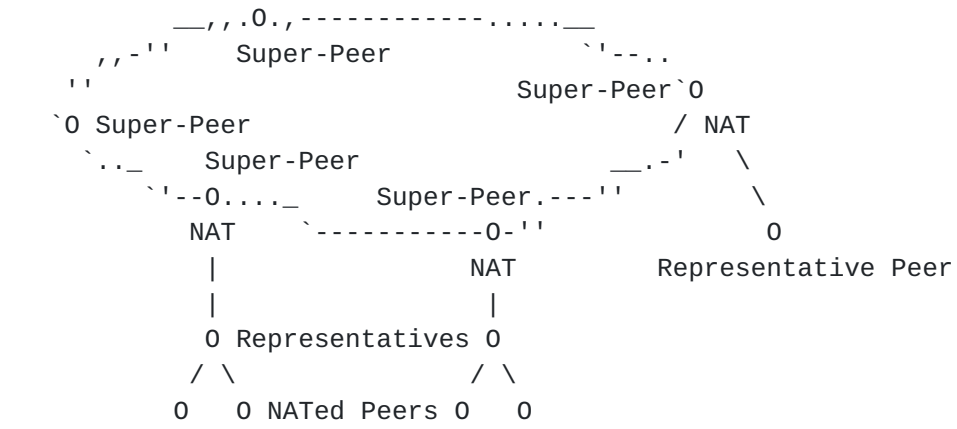

Figure 6 Overlay Network with Super-Peers and Representatives

The network shown in Figure 6 minimizes the amount of effort that needs to be expended for NAT pinhole maintenance, but introduces another level of hierarchy into the overlay and thus increases message hop counts. Further, it requires some new mechanism to allow NATed peers to discover each other and elect a representative.

In this topology, both the super-peer and the representative are assumed to be servicing large numbers of NATed peers, so their performance and availability are a concern.

6.2. Representative Overlays

It's worth noting that the super-peer and representative concepts are independent of each other. It is possible to construct an overlay network in which representative peers (residing behind NATs) use ICE NAT traversal techniques to create connections to other peers in the overlay. No super-peers (publicly addressable peers) need be present in such a network.

This is a similar type of hierarchy to the super-peer hierarchy in that representative peers connected in such a way would have overlay IDs and implement the search algorithm and NATed peers would not. This type of overlay topology would increase the number of connections crossing the NAT above the bare minimum required in Figure 6, but instead of being proportional to the number of super-peers serving nodes behind a NAT, it would instead be related to the number of connections the representative had to other peers.

As with the super-peer overlay, representatives are assumed to be serving large numbers of NATed peers, so performance and availability are concerns.

Introducing hierarchy into an overlay network, either through super-peers or representatives, is a relatively effective NAT traversal technique. However, it requires that super-peers and/or representatives must perform two distinct routing operations: one to direct search queries to other super-peers and another one to allow NATed peers to access the overlay.

The inclusion of a hierarchy also requires the creation of new techniques to distribute the load appropriately and recover from failures. These mechanisms are generally independent from similar mechanisms already present in search algorithms like DHTs.

7. Conclusions

The presence of NATs in IP networks present several challenges for P2PSIP overlays. A super-peer overlay architecture is easy-to-understand and provides effective NAT traversal. However, it concentrates the network load on a small percentage of the participating nodes and cannot be used in networks that have no publicly addressable peers.

Fully distributed overlays traverse NATs equally well and share the load evenly across participating peers, which results in greater performance and scalability. Since they do not require any nodes to have public IP addresses, these architectures can be applied to more IP network topologies.

Fully distributed networks implicitly determine (based upon their search algorithm) how many connections will cross a peer's local NAT. Depending on the search algorithm, it may be possible to adjust the number of connections so no single NAT is overwhelmed by the keep-alive traffic or number of mappings it needs to maintain.

Super-peer overlays have no inherent mechanism for associating NATed peers with super-peers, so one must be created. In creating this mechanism special consideration must be given to the resources available at both the super-peer and in the NAT. Due to their role as routers for overlay messages, super-peers that serve many NATed peers must be highly available and have high-bandwidth Internet links.

In fully distributed networks the connections required for message routing are the same ones used by the search algorithm. Since the routing topology in a super-peer overlay is separate from the

searching mechanism, a super-peer overlay will devote extra resources to NAT traversal.

8. Security Considerations

Security Considerations will be covered in a later version of this paper.

9. IANA Considerations

IANA considerations will be covered in a later version of this paper.

10. Acknowledgments

APPENDIX A: Other NAT Traversal Techniques

In addition to the techniques discussed in section 2.2. other addition to those, some other mechanisms for NAT traversal are: UPnP, ALGs, SBCs, and manual configuration.

Universal Plug-n-Play (UPnP) is a NAT configuration scheme developed by Microsoft. This HTTP/XML based protocol allows applications to dynamically create forwarding rules in the NAT as needed. Many consumer-grade NATs support the UPnP protocol, which makes this seem quite promising for P2P applications targeted only at the consumer market. However, most corporate-grade NATs do not support UPnP. Even in the consumer market space, the user would be required to provide the administrative password for the NAT. Further, even if administrative access to the NAT is possible, UPnP cannot provide a complete solution if there are multiple NATs between the P2PSIP device and the public Internet.

Many NATs contain one or more Application Level Gateways (ALGs). An ALG is special code within the NAT that recognizes packets of a particular application-level protocol and treats the packets specially. ALG support for the File Transfer Protocol (FTP) is almost universal in NATs, and ALG support for SIP is becoming more common. However, ALG support requires that the application protocol not be encrypted end-to-end, and end-to-end encryption of both SIP and P2P messages is likely to be desirable for security reasons.

Session Border Controllers (SBCs) are boxes that are deployed in the network, sometimes by the customer but more commonly by the SIP service provider, to enable NAT traversal for standard client-server SIP. SBCs are becoming more common, but typically sit on the border of a "trusted" SIP service provider network and an "untrusted" network (usually the public Internet). In a distributed network of P2PSIP peers, there is no single boundary where an SBC would be appropriate. Furthermore, SBCs are typically designed to work in client-server deployments, and even then only with the SIP proxy servers of a specific SIP service provider. Thus they are not well suited as a NAT traversal option for P2PSIP networks.

NAT traversal is often much easier if the user can manually configure the NAT. The user can open up pinholes in the NAT and/or modify the NAT's behavior. However, this requires that the user have the knowledge and interest to do the configuration (non-technical users often do not), have a NAT that is configurable (some low-end NATs are not configurable), and have permission to configure the NAT (problematic in corporate environments or when there are NATs in the ISP's network). Like UPnP, manual configuration cannot provide a

complete solution if there are multiple NATs between the P2PSIP device and the public Internet.

11. References

11.1. Normative References

11.2. Informative References

[RFC3261] Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., Handley, M., and E. Schooler, "SIP: Session Initiation Protocol", RFC 3261, http://www.ietf.org/rfc/rfc3261.txt, June 2002.

[RFC3263] Rosenberg, J., and Schulzrinne, H., "Session Initiation Protocol (SIP): Locating SIP Servers", RFC 3263, http://www.ietf.org/rfc/rfc3263.txt, June 2002.

[p2p-comm] Ford, B., Srisuresh, P., and Kegel, D., "State of Peer-to-Peer (P2P) Communication Across Network Address Translators (NATs)", draft-ietf-behave-p2p-state-02 (work in progress), http://www.ietf.org/internet-drafts/draft-ietf-behave-p2p-state-02.txt, February 2007.

[nat-consider] Rosenberg, J., "Considerations for Selection of Techniques for NAT Traversal", draft-iab-nat-traversal-considerations-00 (expired), http://tools.ietf.org/html/draft-iab-nat-traversal-considerations-00.

[ice] Rosenberg, J., "Interactive Connectivity Establishment (ICE): A Methodology for Network Address Translator (NAT) Traversal for Offer/Answer Protocols," draft-ietf-mmusic-ice-13 (work in progress), http://www.ietf.org/internet-drafts/draft-ietf-mmusic-ice-13.txt.

[nat-scenarios] Boulton, C., Rosenberg, J. and Camarillo, G., "Best Current Practices for NAT Traversal for SIP", draft-ietf-sipping-nat-scenarios-06 (work in progress), http://www.ietf.org/internet-drafts/draft-ietf-sipping-nat-scenarios-06.txt, March 2007.

[outbound] Jennings, C. and Mahy, R., "Managing Client Initiated Connections in the Session Initiation Protocol (SIP)", draft-ietf-sip-outbound-07 (work in progress), http://www.ietf.org/internet-drafts/draft-ietf-sip-outbound-07.txt, January 2007.

[chord] Stoica, I., Morris, R., Liben-Nowell, D., Karger, D., Kaashoek, M., Dabek, F., and H. Balakrishnan, "Chord: A Scalable Peer-to-peer Lookup Service for Internet Applications" article(s) available at http://pdos.csail.mit.edu/chord/pubs.html.

[use-cases] Bryan, D. A., Shim, E. and Lowekamp, B. B., "Use Cases for Peer-to-Peer Session Initiation Protocol (P2P SIP)", draft-bryan-sipping-p2p-usecases-00 (expired), http://tools.ietf.org/id/draft-bryan-sipping-p2p-usecases-00.txt.

[illuminati] Cadaco, M. and Freedman, M., "Illuminati - Opportunistic Network and Web Measurement", http://illuminati.coralcdn.org/stats, February 2007.

[concepts] Bryan, D., Matthews, P., Shim, E. and Willis, D., "Concepts and Terminology for Peer to Peer SIP", draft-willis-p2psip-concepts-03 (work in progress), http://www.ietf.org/internet-drafts/draft-willis-p2psip-concepts-04.txt, March 2007.

[behave-udp] Audet, F. and C. Jennings, "Network Address Translation (NAT) Behavioral Requirements for Unicast UDP", RFC4787/BCP127, http://www.ietf.org/rfc/rfc4787.txt, January 2007.

[behave-tcp] Guha, S., Biswas, K., Ford, B., Sivakumar, S., and Srisuresh, P., "NAT Behavioral Requirements for TCP", draft-ietf-behave-tcp-05 (work in progress), http://www.ietf.org/internet-drafts/draft-ietf-behave-tcp-05.txt, February 2007.

[dSIP-nat] Cooper, E., Matthews, P., Bryan, D. and Lowekamp, B., "NAT Traversal for dSIP", draft-matthews-p2psip-dsip-nat-traversal-00 (work in progress), http://www.ietf.org/internet-drafts/draft-matthews-p2psip-dsip-nat-traversal-00.txt, February 2007.

[stun] Rosenberg, J., Huitema, C., Mahy, R., and Wing, D., "Simple Traversal Underneath Network Address Translators (NAT) (STUN)", draft-ietf-behave-rfc3489bis-05 (work in progress), http://www.ietf.org/internet-drafts/draft-ietf-behave-rfc3489bis-05.txt, October 2006.

Author's Addresses

Eric Cooper
Avaya
1135 Innovation Dr.
Ottawa, Ontario K2K 3G7
Canada

Phone: +1 613 592 4343 x228
Email: ecooper@avaya.com

Philip Matthews
Avaya
1135 Innovation Dr.
Ottawa, Ontario K2K 3G7
Canada

Phone: +1 613 592 4343 x224
Email: philip_matthews@magma.ca

Full Copyright Statement

Intellectual Property

Acknowledgment

Funding for the RFC Editor function is provided by the IETF Administrative Support Activity (IASA).