Port Control Protocol (PCP) Rapid Recovery allows PCP clients to repair failed mappings within seconds, rather than the minutes or hours it might take if they relied solely on waiting for the next routine renewal of the mapping. Mapping failures may occur when a NAT gateway is rebooted and loses its mapping state, or when a NAT gateway has its external IP address changed so that its current mapping state becomes invalid.

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 December 8, 2011.

Copyright Notice

Copyright (c) 2011 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. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of
1. 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 "Key words for use in RFCs to Indicate Requirement Levels" [RFC2119].

2. Introduction

Port Control Protocol [PCP] allows a host to control how incoming IPv6 or IPv4 packets are translated and forwarded by a network address translator (NAT) or simple firewall to an IPv6 or IPv4 host, and also allows a host to optimize its outgoing NAT keepalive messages.

PCP Rapid Recovery allows PCP clients to repair failed mappings within seconds, rather than the minutes or hours it might take if they relied solely on waiting for the next routine renewal of the mapping. Mapping failures may occur when a NAT gateway is rebooted and loses its mapping state, or when a NAT gateway has its external IP address changed so that its current mapping state becomes invalid.

3. PCP Restart Announcement

When a PCP-enabled NAT gateway [PCP] that implements PCP Rapid Recovery reboots, restarts its NAT engine, or otherwise enters a state where it may have lost some or all of its previous mapping state (or doesn’t know whether it may have had prior mapping state that it lost) it MUST inform PCP clients of this fact by multicasting the UDP packet shown below to 224.0.0.1:5350 on all multicast-capable interfaces on which it accepts PCP requests. To accommodate packet loss, the PCP-enabled NAT MAY transmit such packets up to ten times (with an appropriate Epoch value in each to reflect the passage of time between transmissions) provided that the interval between the first two notifications is at least 250ms, and the interval between subsequent notification at least doubles.
Figure 1: PCP Restart Announcement Packet

A PCP client that implements PCP Rapid Recovery MUST listen on UDP 224.0.0.1:5350 on all multicast-capable interfaces on which it has sent PCP requests, to receive these PCP Restart Announcements. (The SO_REUSEPORT socket option or equivalent should be used for the multicast UDP port, if required by the host OS to permit multiple independent listeners on the same multicast UDP port.) Upon receiving a PCP Restart Announcement a PCP client MUST (as it does with all received PCP response packets) inspect the Announcement’s source IP address, and if the Epoch value indicates that the NAT gateway has begun a new epoch since the last time the PCP client received a PCP response message from that PCP server address, then for all PCP mappings it made at that address the client should issue new PCP requests to recreate any lost mapping state. The use of the Suggested External IP Address and Suggested External Port fields in the client’s renewal request allows the client to remind the restarted NAT gateway of what mappings the client had previously been given, so that in many cases the prior state can be recreated. For NAT gateways that reboot relatively quickly it is usually possible to reconstruct lost mapping state fast enough that existing TCP connections and UDP communications do not time out and continue without failure.

The PCP Rapid Recovery capability enables users to, for example, connect to remote machines using ssh, and then reboot the NAT gateway (or even replace it with completely new hardware) without losing their established ssh connections.

Use of PCP Rapid Recovery is a performance optimization. Without it, PCP clients will still recreate their correct state when they next renew their mappings, but this routine self-healing process may take hours rather than seconds, and will probably not happen fast enough to prevent active TCP connections from timing out.
4. PCP Mapping Update

If a PCP-enabled NAT gateway has not forgotten its mapping state, but for some other reason has determined that some or all of its mappings have become unusable (e.g. when a home gateway is assigned a different external IPv4 address by the upstream DHCP server) then the NAT gateway MAY chose to remedy this situation by automatically repairing its mappings and notifying its clients.

For PCP MAP mappings, for each one the NAT gateway should update the External IP Address and External Port to appropriate available values, and then send unicast PCP MAP responses to inform the PCP client of the new External IP Address and External Port. Such MAP responses are identical to the MAP responses normally returned in response to client MAP requests, except they may be viewed as a long-delayed response to an earlier MAP request, containing newly updated External IP Address and External Port values.

To accommodate packet loss, the PCP-enabled NAT MAY transmit such packets up to ten times (with an appropriate Epoch value in each to reflect the passage of time between transmissions) provided that the interval between the first two notifications is at least 250ms, and the interval between subsequent notification at least doubles.

Upon receipt of such long-delayed MAP responses, a PCP client MUST to use the information in them to update its DNS records, or other address and port information recorded with some kind of application-specific rendezvous server. Existing TCP connections will be lost, but promptly updating the DNS or rendezvous server with the new data will allow new connections to be made.

For PCP PEER mappings there is no general way to recover them (the remote host doesn’t know the new External IP Address and External Port) so existing connections will be lost. Accordingly, a PCP-enabled NAT gateway is not required to take any specific action for PEER mappings. It MAY delete all PEER mappings immediately (and let application-layer timeouts detect the failure) or it MAY choose to retain them for some time in case another change in the external environment (e.g. a lost DHCP-assigned external address is re-assigned after a few seconds) results in the mappings becoming usable again.
5. Security Considerations

Forged PCP Restart Announcements could be used to cause high load on a PCP server.

Forged MAP responses could be used to mislead a PCP client about what External IP Address and External Port is has been allocated.

6. IANA Considerations

IANA is requested to record that UDP port 5350 is now formally reallocated from "NAT-PMP Restart Announcement" to "PCP Restart Announcement".

7. Normative References

[PCP] Wing, D., "Port Control Protocol (PCP)",
draft-ietf-pcp-base-07 (work in progress), March 2011.


Author’s Address

Stuart Cheshire
Apple Inc.
1 Infinite Loop
Cupertino, California  95014
USA

Phone: +1 408 974 3207
Email: cheshire@apple.com