Network Working Group
Internet-Draft
Intended status: Informational
Expires: September 10, 2009

J. Abley
TekSavvy
W. Maton
NRC-CNRC
March 9, 2009

AS112 Nameserver Operations
draft-ietf-dnsop-as112-ops-02

Status of this Memo

Copyright Notice

Abstract

Many sites connected to the Internet make use of IPv4 addresses which are not globally unique. Examples are the addresses designated in RFC1918 for private use within individual sites.

Devices in such environments may occasionally originate reverse DNS queries corresponding to those private-use addresses. Since the addresses concerned have only local significance, it is good practice for site administrators to ensure that they are answered locally. However, it is not uncommon for such queries to follow the normal delegation path in the public DNS instead of being answered within the site.

It is not possible for public DNS servers to give useful answers to such queries. In addition, due to the wide deployment of private-use addresses and the continuing growth of the Internet, the volume of such queries is large and growing. The AS112 project aims to provide a distributed sink for such queries in order to reduce the load on the root and IN-ADDR.ARPA authority servers.

This document describes the steps required to install a new AS112 node, and offers advice relating to such a node's operation.

Table of Contents

1. Introduction

Many sites connected to the Internet make use of IPv4 addresses which are not globally unique. Examples are the addresses designated in [RFC1918] for private use within individual sites.

Devices in such environments may occasionally originate reverse DNS queries [RFC1034] corresponding to those private-use addresses. Since the addresses concerned have only local significance, it is good practice for site administrators to ensure that they are answered locally [I-D.ietf-dnsop-default-local-zones]. However, it is not uncommon for such queries to follow the normal delegation path in the public DNS instead of being answered within the site.

It is not possible for public DNS servers to give useful answers to such queries. In addition, due to the wide deployment of private-use addresses and the continuing growth of the Internet, the volume of such queries is large and growing. The AS112 project aims to provide a distributed sink for such queries in order to reduce the load on the root and IN-ADDR.ARPA authority servers.

The AS112 project encompasses a loosely-coordinated collection of independently-operated nameservers. Each nameserver functions as a single node in an AS112 anycast cloud [RFC4786], and is configured to answer authoritatively for a particular set of nominated zones.

2. AS112 DNS Service

2.1. Zones

AS112 nameservers answer authoritatively for the following zones, corresponding to [RFC1918] private-use netblocks:

- 10.IN-ADDR.ARPA
- 16.172.IN-ADDR.ARPA, 17.172.IN-ADDR.ARPA, ..., 31.172.IN-ADDR.ARPA
- 168.192.IN-ADDR.ARPA

and the following zone, corresponding to the "link local" netblock 169.254.0.0/16 listed in [RFC3330]:

- 254.169.IN-ADDR.ARPA

To aid identification of AS112 anycast nodes, each node also answers authoritatively for the zone HOSTNAME.AS112.NET. See Section 3.5 for more details on the resource records contained within that zone.

It is possible that the IANA might delegate other zones corresponding to private-use infrastructure to AS112 servers in the future. A current list of zones for which AS112 servers answer authoritatively can be found at <http://www.as112.net/>.

2.2. Nameservers

The zones listed in Section 2.1 are delegated to the two nameservers BLACKHOLE-1.IANA.ORG (192.175.48.6) and BLACKHOLE-2.IANA.ORG (192.175.48.6).

Additionally, the server PRISONER.IANA.ORG (192.175.48.1) is listed in the SOA RDATA of zones served by AS112 nameservers, and receives mainly dynamic update queries.

It should be noted that the addresses of all these nameservers are covered by the single prefix 192.175.48.0/24.

3. Installation of a New Node

3.1. Useful Background Knowledge

Installation of an AS112 node is relatively straightforward. However, prior knowledge of or experience in the following general areas may prove useful:

- Inter-domain routing with BGP [RFC4271];
- DNS authority server operations;
- Anycast distribution of DNS services ([ISC-TN-2003-1], [RFC4786]).

3.2. Topological Location

AS112 nodes may be located anywhere on the Internet. For nodes which are intended to provide a public service to the Internet community (as opposed to private use), it may well be advantageous to choose a location that is easily (and cheaply) reachable by multiple providers, such as an Internet exchange point.

AS112 nodes may advertise their service prefix to BGP peers for local use (analogous to a conventional peering relationship between two providers) or for global use (analogous to a customer relationship with one or more providers).

It is good operational practice to notify the community of users which may fall within the reach of a new AS112 node before it is installed. At an Internet Exchange, local mailing lists usually exist to facilitate such announcements. For nodes which are intended to be globally reachable, coordination with other AS112 operators is highly recommended. See also Section 5.

3.3. Operating System and Host Considerations

The use of a UNIX or UNIX-like operating system (e.g. FreeBSD, GNU Linux) is recommended for the construction of AS112 nodes, primarily due to the cumulative experience of using such platforms for this purpose. Examples in this document will assume use of such an operating system.

The chosen platform should include support for either cloned loopback interfaces, or the capability to bind multiple addresses to a single loopback interface. The addresses of the nameservers listed in Section 2.2 will be configured on these interfaces in order that the DNS software can respond to queries properly.

A host which is configured to act as an AS112 anycast node should be dedicated to that purpose, and should not be used to simultaneously provide other services.

System startup scripts should be arranged such that the various AS112-related components start automatically following a system reboot. The order in which interfaces are configured and software components started should be arranged such that routing software startup follows DNS software startup, and DNS software startup follows loopback interface configuration.

Wrapper scripts or other arrangements should be employed to ensure that the anycast service prefix for AS112 is not advertised while either the anycast addresses are unconfigured, or while the DNS software is not running.

3.4. Routing Software

AS112 nodes signal the availability of AS112 nameservers to the Internet using BGP [RFC4271]: each AS112 node is a BGP speaker, and announces the prefix 192.175.48.0/24 to the Internet with origin AS 112 (see also Section 2.2).

Suitable choices of free software to allow hosts to act as BGP speakers include, but are not limited to:

- OpenBGPD [1]
- The Quagga Routing Suite [2]
- GNU Zebra [3]

The examples in this document are based on Quagga.

The "bgpd.conf" file is used by Quagga's bgpd daemon, which provides BGP protocol support. The router id in this case is 198.32.149.123; the AS112 node peers with external peers 198.32.149.1 and 198.32.149.2, which are route servers at an exchange point. Note the local AS number 112, and the origination of the prefix 192.175.48.0/24.

```
! bgpd.conf
!
hostname as112-bgpd
password <something>
enable password <supersomething>
!
router bgp 112
 bgp router-id 198.32.149.123
 network 192.175.48.0
 neighbor 198.32.149.1 remote-as 2884
 neighbor 198.32.149.1 next-hop-self
 neighbor 198.32.149.2 remote-as 2884
 neighbor 198.32.149.2 next-hop-self
```

The "zebra.conf" file is required to provide integration between protocol daemons (bgpd, in this case) and the kernel.

```
! zebra.conf
!
hostname as112
password <something>
enable password <supersomething>
!
interface lo
!
interface eth0
!
```

3.5. DNS Software

Although the queries received by AS112 nodes are definitively misdirected, it is important that they be answered in a manner which is accurate and consistent. For this reason AS112 nodes operate as fully-functional and standards-compliant DNS authority servers [RFC1034], and hence require DNS software.

Suitable choices of free DNS software for AS112 nodes include, but are not limited to:

- ISC BIND9 [4]
- NLnet Labs' NSD [5]

Examples in this document are based on ISC BIND9.

The following is a sample BIND9 "named.conf" file for a dedicated AS112 server. Note that the nameserver is configured to act as an authority-only server (i.e. recursion is disabled). The nameserver

is also configured to listen on the various AS112 anycast nameserver addresses, as well as its local addresses.

```
// named.conf

// global options

options {
  listen-on {
    127.0.0.1;         // localhost
    198.32.149.252;    // local address (globally-unique, unicast)
    192.175.48.1;      // prisoner.iana.org (anycast)
    192.175.48.6;      // blackhole-1.iana.org (anycast)
    192.175.48.42;     // blackhole-2.iana.org (anycast)
  };
  directory "/var/named";
  recursion no;        // authority-only server
  query-source address *;
};

// log queries,  so that when people call us about unexpected
// answers to queries they didn't realise they had sent, we
// have something to talk about.  Note that activating this
// has the potential to create high CPU and take enormous
// amounts of disk space.

logging {
  channel "querylog" {
    file "/var/log/query.log" versions 2 size 500m;
    print-time yes;
  };
  category queries { querylog; };
};

// RFC 1918

zone "10.in-addr.arpa" { type master; file "db.empty"; };
zone "16.172.in-addr.arpa" { type master; file "db.empty"; };
zone "17.172.in-addr.arpa" { type master; file "db.empty"; };
zone "18.172.in-addr.arpa" { type master; file "db.empty"; };
zone "19.172.in-addr.arpa" { type master; file "db.empty"; };
zone "20.172.in-addr.arpa" { type master; file "db.empty"; };
zone "21.172.in-addr.arpa" { type master; file "db.empty"; };
zone "22.172.in-addr.arpa" { type master; file "db.empty"; };
zone "23.172.in-addr.arpa" { type master; file "db.empty"; };
zone "24.172.in-addr.arpa" { type master; file "db.empty"; };
zone "25.172.in-addr.arpa" { type master; file "db.empty"; };
zone "26.172.in-addr.arpa" { type master; file "db.empty"; };
```

```
zone "27.172.in-addr.arpa" { type master; file "db.empty"; };
zone "28.172.in-addr.arpa" { type master; file "db.empty"; };
zone "29.172.in-addr.arpa" { type master; file "db.empty"; };
zone "30.172.in-addr.arpa" { type master; file "db.empty"; };
zone "31.172.in-addr.arpa" { type master; file "db.empty"; };
zone "254.169.in-addr.arpa" { type master; file "db.empty"; };
zone "168.192.in-addr.arpa" { type master; file "db.empty"; };

// also answer authoritatively for the HOSTNAME.AS112.NET zone,
// which contains data of operational relevance

zone "hostname.as112.net" { type master;
  file "db.hostname.as112.net"; };
```

The "db.empty" file follows, below. This is the source data used to populate all the zones listed in Section 2.1.

```
; db.empty
;
; Empty zone for AS112 server.
;
$TTL    1W
@  IN  SOA  prisoner.iana.org. hostmaster.root-servers.org. (
                                1       ; serial number
                                1W      ; refresh
                                1M      ; retry
                                1W      ; expire
                                1W )    ; negative caching TTL
;
       NS     blackhole-1.iana.org.
       NS     blackhole-2.iana.org.
;
; There should be no other resource records included in this zone.
;
; Records which relate to RFC1918-numbered resources within the
; site hosting this AS112 node should not be hosted on this
; nameserver.
```

The "db.hostname.as112.net" file follows, below. This zone contains various resource records which provide operational data to users for troubleshooting or measurement purposes, and should be edited to suit local circumstances. Note that the response to the query "HOSTNAME.AS112.NET IN TXT" should fit within a 512 octet DNS/UDP datagram: i.e. it should be available over UDP transport without requiring EDNS0 support.

The LOC record [RFC1876] included in the zone apex provides information about the geospacial location of the node.

```
; db.hostname.as112.net
;
$TTL    1W
@       SOA     flo.gigafed.net. dns.ryouko.imsb.nrc.ca. (
                        1               ; serial number
                        1W              ; refresh
                        1M              ; retry
                        1W              ; expire
                        1W )            ; negative caching TTL
;
        NS      blackhole-2.iana.org.
        NS      blackhole-1.iana.org.
;
        TXT     "Federal GigaPOP" "Ottawa, Canada"
        TXT     "See http://as112.net/ for more information."
;
        LOC     45 25 0.000 N 75 42 0.000 W 80.00m 1m 10000m 10m
```

3.6. Testing a Newly-Installed Node

The BIND9 tool "dig" can be used to retrieve the TXT resource records associated with the domain "HOSTNAME.AS112.NET", directed at one of the AS112 anycast nameserver addresses. Continuing the example from above, the response received should indicate the identity of the AS112 node which responded to the query. See Section 3.5 for more details about the resource records associated with "HOSTNAME.AS112.NET".

```
% dig @prisoner.iana.org hostname.as112.net txt +short +norec
"Federal GigaPOP" "Ottawa, Canada"
"See http://www.as112.net/ for more information."
%
```

If the response received indicates a different node is being used, then there is probably a routing problem to solve. If there is no response received at all, there might be host or nameserver problem. Judicious use of tools such as traceroute, and consultation of BGP looking glasses might be useful in troubleshooting.

Note that an appropriate set of tests for a new server will include queries sent from many different places within the expected service area of the node, using both UDP and TCP transport, and exercising all three AS112 anycast nameserver addresses.

4. Operations

4.1. Monitoring

AS112 nodes should be monitored to ensure they are functioning correctly, just as with any other production service. An AS112 node which stops answering queries correctly can cause failures and timeouts in unexpected places, and can lead to failures in dependent systems which can be difficult to troubleshoot.

4.2. Downtime

An AS112 node which needs to go off-line (e.g. for planned maintenance, or as part of the diagnosis of some problem) should stop advertising the AS112 service prefix to its BGP peers. This can be done by shutting down the routing software on the node altogether, or by causing the routing system to withdraw the route.

Withdrawing the service prefix is important in order to avoid blackholing query traffic in the event that the DNS software on the node is not functioning normally.

4.3. Statistics and Measurement

Use of the AS112 node should be measured in order to track long-term trends, identify anomalous conditions and to ensure that the configuration of the AS112 node is sufficient to handle the query load.

Examples of free monitoring tools which might be useful to operators of AS112 nodes include, but are not limited to:

- bindgraph [6]
- dnstop [7]
- DSC [8]

5. Communications

It is good operational practice to notify the community of users which may fall within the reach of a new AS112 node before it is installed. At Internet Exchanges, local mailing lists usually exist to facilitate such announcements.

For nodes which are intended to be globally reachable, coordination with other AS112 operators is especially recommended.

Operational notices relating to all AS112 nodes may be sent to <mailto:112@root-servers.org>. Information pertinent to AS112 operations is maintained at <http://www.as112.net/>.

Information about an AS112 node should also be published within the DNS, within the "HOSTNAME.AS112.NET" zone. See Section 3.5 for more details.

6. Future Usefulness of AS112 Nodes

It is recommended practice for the operators of recursive nameservers to answer queries for zones served by AS112 nodes locally, such that queries never have an opportunity to reach AS112 servers [I-D.ietf-dnsop-default-local-zones]. Operational experience with AS112 nodes does not currently indicate an observable trend towards compliance with those recommendations, however.

It is expected that some DNS software vendors will include default configuration which will implement measures such as those described in [I-D.ietf-dnsop-default-local-zones]. If such software is widely deployed, it is reasonable to assume that the query load received by AS112 nodes will decrease; however, it is safe to assume that the query load will not decrease to zero, and consequently that AS112 nodes will continue to provide a useful service for the foreseeable future.

There may be a requirement in the future for AS112 nodes to answer for their current set of zones over IPv6 transport. Such a requirement would necessitate the assignment of a corresponding IPv6 netblock for use as an anycast service prefix.

There may be a requirement in the future for AS112 nodes to serve additional zones, or to stop serving particular zones that are currently served. Such changes would be widely announced in operational forums, and published at <http://www.as112.net/>.

7. IANA Considerations

The AS112 nameservers are all named under the domain IANA.ORG (see Section 2.2). The IANA is the organisation responsible for the coordination of many technical aspects of the Internet's basic infrastructure. The AS112 project nameservers provide a public service to the Internet which is sanctioned by and operated in coordination with the IANA.

This document does not require any IANA actions.

8. Security Considerations

Hosts should never normally send queries to AS112 servers; queries relating to private-use addresses should be answered locally within a site. Hosts which send queries to AS112 servers may well leak information relating to private infrastructure to the public network, which could represent a security risk. This risk is orthogonal to the presence or absence of authority servers for these zones in the public DNS infrastructure, however.

Requests which are answered by AS112 servers are usually unintentional; it follows that the responses from AS112 servers are usually unexpected. Unexpected inbound traffic can trigger intrusion detection systems or alerts by firewalls. Operators of AS112 servers should be prepared to be contacted by operators of remote infrastructure who believe their security has been violated.

The deployment of AS112 nodes are very loosely coordinated, compared to other services distributed using anycast. The compromise of an AS112 node and subversion of the data served by the node is hence more difficult to detect due to the lack of central management. Since it is conceivable that changing the responses to queries received by AS112 nodes might influence the behaviour of the hosts sending the queries, such a compromise might be used as an attack vector against private infrastructure.

Operators of AS112 should take appropriate measures to ensure that AS112 nodes are appropriately protected from compromise, such as would normally be employed for production nameserver or network infrastructure. The guidance provided for root nameservers in [RFC2870] may be instructive.

9. References

9.1. Normative References

[RFC1034] Mockapetris, P., "Domain names - concepts and facilities", STD 13, RFC 1034, November 1987.

[RFC1918] Rekhter, Y., Moskowitz, R., Karrenberg, D., Groot, G., and E. Lear, "Address Allocation for Private Internets", BCP 5, RFC 1918, February 1996.

[RFC2870] Bush, R., Karrenberg, D., Kosters, M., and R. Plzak, "Root Name Server Operational Requirements", BCP 40, RFC 2870, June 2000.

[RFC4271] Rekhter, Y., Li, T., and S. Hares, "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, January 2006.

[RFC4786] Abley, J. and K. Lindqvist, "Operation of Anycast Services", BCP 126, RFC 4786, December 2006.

9.2. Informative References

[I-D.ietf-dnsop-default-local-zones] Andrews, M., "Locally-served DNS Zones", draft-ietf-dnsop-default-local-zones-08 (work in progress), February 2009.

[ISC-TN-2003-1] Abley, J., "Hierarchical Anycast for Global Service Distribution", <http://www.isc.org/pubs/tn/isc-tn-2003-1.html>.

[RFC1876] Davis, C., Vixie, P., Goodwin, T., and I. Dickinson, "A Means for Expressing Location Information in the Domain Name System", RFC 1876, January 1996.

[RFC3330] IANA, "Special-Use IPv4 Addresses", RFC 3330, September 2002.

URIs

[1] <http://www.openbgpd.org/>

[2] <http://www.quagga.net/>

[3] <http://www.zebra.org/>

[4] <http://www.isc.org/products/BIND/>

[5] <http://www.nlnetlabs.nl/nsd/>

[6] <http://www.linux.it/~md/software/>

[7] <http://dns.measurement-factory.com/tools/dnstop/>

[8] <http://dns.measurement-factory.com/tools/dsc/>

Appendix A. History

Widespread use of the private address blocks listed in [RFC1918] followed that document's publication in 1996.

The idea of off-loading IN-ADDR.ARPA queries relating to [RFC1918] addresses from the root nameservers was first proposed by Bill Manning and John Brown.

The use of anycast for distributing authority service for [RFC1918] IN-ADDR.ARPA zones was subsequently proposed at a private meeting of root server operators.

ARIN provided an IPv4 prefix for the anycast service, and also the autonomous system number 112 for use in originating that prefix. This assignment gave the project its name.

In 2002, the first AS112 anycast nodes were deployed.

The use of anycast nameservers in the AS112 project contributed to the operational experience of anycast DNS services, and can be seen as a precursor to the anycast distribution of other authority servers in subsequent years (e.g. various root servers).

Appendix B. Acknowledgements

The authors wish to acknowledge the assistance of Bill Manning, John Brown, Marco D'Itri, Daniele Arena, Stephane Bortzmeyer, Frank Habicht and Peter Losher in the preparation of this document.

Appendix C. Change History

This section to be removed prior to publication.

00 Initial draft, circulated as draft-jabley-as112-ops-00 and reviewed at the DNSOP working group meeting at IETF 66.

00 Document adoped by the DNSOP working group and renamed accordingly.

01 Input from reviewers of DNSOP and others, some cosmetic tweaks.

02 Version bump as request by working group chairs.

02 Added missing IANA Considerations section.

02 Updated author's addresses.

Authors' Addresses

Joe Abley
TekSavvy Solutions, Inc.
330 Richmond Street, Suite 205
Chatham, ON N7M 1P7
Canada

Phone: +1 519 670 9327
Email: jabley@teksavvy.com

William F. Maton Sotomayor
National Research Council of Canada
1200 Montreal Road
Ottawa, ON K1A 0R6
Canada

Phone: +1 613 993 0880
Email: wmaton@ryouko.imsb.nrc.ca