

RADOS - Bug #47951

MonClient: mon_host with DNS Round Robin results in 'unable to parse addrs'

10/22/2020 04:34 PM - Wido den Hollander

Status:	Resolved	% Done:	0%
Priority:	Urgent	Spent time:	0.00 hour
Assignee:	Patrick Donnelly		
Category:			
Target version:	v16.0.0		
Source:	Community (user)	Affected Versions:	v14.2.12
Tags:	ipv6,dns,round robin,mon_host,client	ceph-qa-suite:	
Backport:	octopus,nautilus	Component(RADOS):	MonClient
Regression:	Yes	Pull request ID:	37758
Severity:	3 - minor	Crash signature (v1):	
Reviewed:		Crash signature (v2):	

Description

I performed a test upgrade to 14.2.12 today on a cluster using IPv6 with Round Robin DNS for mon_host

```
[global]
auth_client_required = cephx
auth_cluster_required = cephx
auth_service_required = cephx
fsid = 0d56dd8f-7ae0-4447-b51b-f8b818749307
mon_host = mon.objects.xxxx
ms_bind_ipv6 = true
```

Running 'ceph -s' now fails:

```
root@wido-standard-benchmark:~# ceph -s
unable to parse addrs in 'mon.objects.xxx.xxxx.xxxx'
[errno 22] error connecting to the cluster
root@wido-standard-benchmark:~#
```

The hostname is a Round Robin DNS entry pointing to IPv6 addresses:

```
root@wido-standard-benchmark:~# host mon.objects.ams02.cldin.net
mon.objects.xx.xx.net has IPv6 address 2a05:yy:xx:d:84b5:85ff:zzzz:33bf
mon.objects.xx.xx.net has IPv6 address 2a05:yy:xx:d:645f:97ff:zzzz:2b2a
mon.objects.xx.xx.net has IPv6 address 2a05:yy:xx:d:3416:d5ff:zzzz:18db
root@wido-standard-benchmark:~#
```

I took a look with strace and I found this:

```
14980 socket(AF_INET6, SOCK_DGRAM|SOCK_CLOEXEC, IPPROTO_IP) = 3
14980 connect(3, {sa_family=AF_INET6, sin6_port=htons(0), inet_pton(AF_INET6, "2a05:xxx:xxx:d:84b5:85ff:fe40:33bf", &sin6_addr), sin6_flowinfo=htonl(0), sin6_scope_id=0}, 28) = 0
14980 getsockname(3, {sa_family=AF_INET6, sin6_port=htons(52258), inet_pton(AF_INET6, "2a05:xxx:xxx:0:1c00:16ff:fe00:60", &sin6_addr), sin6_flowinfo=htonl(0), sin6_scope_id=0}, [28]) = 0
14980 connect(3, {sa_family=AF_UNSPEC, sa_data="\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0"}, 16) = 0
```

```
14980 connect(3, {sa_family=AF_INET6, sin6_port=htons(0), inet_pton(AF_INET6, "2a05:xxx:xxx:d:645f:97ff:fe7f:2b2a", &sin6_addr), sin6_flowinfo=htonl(0), sin6_scope_id=0}, 28) = 0
14980 getsockname(3, {sa_family=AF_INET6, sin6_port=htons(52850), inet_pton(AF_INET6, "2a05:xxx:xxx:0:1c00:16ff:fe00:60", &sin6_addr), sin6_flowinfo=htonl(0), sin6_scope_id=0}, [28]) = 0
14980 connect(3, {sa_family=AF_UNSPEC, sa_data="\0\0\0\0\0\0\0\0\0\0\0\0"}, 16) = 0
14980 connect(3, {sa_family=AF_INET6, sin6_port=htons(0), inet_pton(AF_INET6, "2a05:xxx:xxx:d:3416:d5ff:fe92:18db", &sin6_addr), sin6_flowinfo=htonl(0), sin6_scope_id=0}, 28) = 0
14980 getsockname(3, {sa_family=AF_INET6, sin6_port=htons(35119), inet_pton(AF_INET6, "2a05:xxx:702:0:1c00:16ff:fe00:60", &sin6_addr), sin6_flowinfo=htonl(0), sin6_scope_id=0}, [28]) = 0
14980 close(3) = 0
14980 write(2, "unable to parse addrs in '", 26) = 26
14980 write(2, "mon.objects.xxx.xxx.net", 27) = 27
14980 write(2, "'", 1) = 1
14980 write(2, "\n", 1)
```

It performs the DNS lookup, but then it doesn't know what to do with it it seems.

Setting this one to Urgent as it breaks existing cluster.

Related issues:

Related to CephFS - Backport #47013: nautilus: librados/libcephfs: use latest...	Resolved
Copied to RADOS - Backport #47986: nautilus: MonClient: mon_host with DNS Rou...	Resolved
Copied to RADOS - Backport #47987: octopus: MonClient: mon_host with DNS Roun...	Resolved

History

#1 - 10/22/2020 04:56 PM - Jason Dillaman

- Project changed from Ceph to RADOS
- Category deleted (MonClient)

#2 - 10/22/2020 04:56 PM - Jason Dillaman

- Component(RADOS) MonClient added

#3 - 10/22/2020 05:06 PM - Patrick Donnelly

- Subject changed from nautilus: mon_host with DNS Round Robin results in 'unable to parse addrs' to MonClient: mon_host with DNS Round Robin results in 'unable to parse addrs'
- Status changed from New to In Progress
- Assignee set to Patrick Donnelly
- Target version set to v16.0.0
- Source set to Community (user)
- Backport set to octopus,nautilus

#4 - 10/22/2020 05:10 PM - Patrick Donnelly

- Status changed from In Progress to Fix Under Review
- Pull request ID set to 37758

#5 - 10/22/2020 05:10 PM - Patrick Donnelly

- Related to Backport #47013: nautilus: librados/libcephfs: use latest MonMap when creating from CephContext added

#6 - 10/22/2020 10:39 PM - Kefu Chai

- Regression changed from No to Yes

#7 - 10/23/2020 07:20 AM - Wido den Hollander

Seems like this commit broke this functionality: <https://github.com/ceph/ceph/commit/2f075704073ff80f94c70cf79516028d2754ae4f>

#8 - 10/23/2020 11:30 AM - Jonas Jelten

The fix is probably:

```
diff --git a/src/mon/MonMap.cc b/src/mon/MonMap.cc
index 19092d5326..05c1cfff31 100644
--- a/src/mon/MonMap.cc
+++ b/src/mon/MonMap.cc
@@ -502,7 +502,7 @@ int MonMap::init_with_hosts(const std::string& hostlist,
     return -EINVAL;
     if (addrs.empty())
         return -ENOENT;
-   if (!init_with_addrs(addrs, for_mkfs, prefix)) {
+   if (init_with_addrs(addrs, for_mkfs, prefix)) {
         return -EINVAL;
     }
     calc_legacy_ranks();
```

#9 - 10/23/2020 05:42 PM - Troy Ablan

This appears to break any sort of resolution of IPv6 addresses from hostnames. This affects qemu's usage of rbd, in this case via libvirt, when there were hostnames pointing to IPv6 addresses were specified as monitors, round-robin or not. Substituting IP addresses here instead works around the problem.

```
<source protocol='rbd' name='vm-pool/gcompute1.las-sda'>
  <host name='mon1.example.com' port='6789' />
  <host name='mon2.example.com' port='6789' />
</source>
```

BTW, it's also unfortunate and disappointing that this release is still completely unmentioned on <https://docs.ceph.com/en/latest/releases/nautilus/>. Is this not the authoritative reference for releases?

#10 - 10/24/2020 02:23 AM - Alex Litvak

Will the fix it to it posted soon? I am building ceph in containers from existing releases, is there a tag I can use to either revert the commit that broke the cluster feature or a build that have a fix implemented?

#11 - 10/24/2020 02:23 AM - Alex Litvak

Alex Litvak wrote:

Will the fix to it be posted soon? I am building ceph in containers from existing releases, is there a tag I can use to either revert the commit that broke the cluster feature or a build that have a fix implemented?

#12 - 10/26/2020 07:00 AM - Kefu Chai

- Status changed from *Fix Under Review* to *Pending Backport*

#13 - 10/26/2020 10:32 AM - Nathan Cutler

- Copied to Backport #47986: *nautilus: MonClient: mon_host with DNS Round Robin results in 'unable to parse addrs' added*

#14 - 10/26/2020 10:32 AM - Nathan Cutler

- Copied to Backport #47987: *octopus: MonClient: mon_host with DNS Round Robin results in 'unable to parse addrs' added*

#15 - 11/11/2020 02:19 PM - Nathan Cutler

- Status changed from *Pending Backport* to *Resolved*

While running with `--resolve-parent`, the script "backport-create-issue" noticed that all backports of this issue are in status "Resolved" or "Rejected".