Re: NLB Cluster − Ping fails or long time to reply from outside local subnet − SOLVED

Source:
http://www.tech−archive.net/Archive/Windows/microsoft.public.windows.server.clustering/2007−09/msg00019.html

- *From*: "Chuck [MSFT]" <ctimon@xxxxxxxxxxxxxxxxxxx>
- *Date*: Wed, 5 Sep 2007 07:06:43 −0400

Yeah, ARP is defined by RFC. So, networking devices, by RFC, are supposed to only update their ARP tables when a packet is received...and that's it. Of course, there exists the capability place 'static' entries in the ARP tables which needs to be done for routers when using Multicast with NLB. I did run into a similar scenario years ago, and it took getting a CCIE in onsite late a night to finally figure out it was an upstream Layer 3 switch that was causing the problem. The way he did it was to simultaneously monitor the ARP table updates for several switches until he found the culprit.

Good hunting.

−−
Chuck Timon, Jr.
Microsoft Corporation
Windows Server 2008 Readiness Team
This posting is provided 'AS IS" with no warranties, and confers no rights.
"Bookham Measures" <bookham_measures_no_spam@xxxxxxxxx> wrote in message
news:OLphBz57HHA.5360@xxxxxxxxxxxxxxxxxxxxxx

> Thanks Chuck.
>
> I thought that the litmus test was that the router functions fine when no NLB is installed, but when it is, things start going screwy. The NLB functions fine on the local network before going through any routers as well. The servers are a couple of switches away from the router so I would have thought that any duplicate MAC info. or similar would have been come irrelavent/concealed at the router.
>
> The command you mentioned reports convergence as desired.
>
> I guess I need to get on the wire and compare a standard ARP query reply and that being given by the adapter when clustering has been configured.
>
> Thanks again
>
> David
>
> "Chuck [MSFT]" <ctimon@xxxxxxxxxxxxxxxxxxx> wrote in message
> news:%23VBs3q07HHA.4436@xxxxxxxxxxxxxxxxxxxxxx

Well, the 'litmus' test, if you will, to tell if NLB is working is to see if all the nodes are converged. If you run the command 'wlbs query' and the node it is run on says it is converged with all of your nodes in the cluster, that basically says the NLB configuration is correct and the nodes are talking to each other. Beyond that, it is something 'exterior' to the NLB cluster itself.

−−
Chuck Timon, Jr.
Microsoft Corporation
Windows Server 2008 Readiness Team
This posting is provided 'AS IS" with no warranties, and confers no rights
.
"David Morgan" <microsoft_newsgroups@xxxxxxxxxxxxxxxxxxxxxx>
wrote in message news:evrIog07HHA.5424@xxxxxxxxxxxxxxxxxxxxxx

Hi Chuck

Thanks for your support on this one.

I think our set−up is pretty standard, if not a little luxurious.

2 servers, each with:

1 NIC for the Cluster
1 NIC for administration (remote desktop, FTP etc.)
1 NIC for communication to the database server

We're trying to future proof ourselves a bit. At the moment we have one webserver that is sending out at over 25Mbps during business hours and via a separate adapter about 10Mbps to the database server.

There should be no reason for the router to discard/ignore/forget ARP replies from the cluster NIC unless it is given a reason to do so.

In some W2K documentation I read that the admin and cluster interfaces should be on different subnets although it did not define what it meant as a subnet. Does it mean different IP range, or is being split on two different VLANs sufficient? I have seen many posts here where people have all adapters on the same IP subnet/range, presumably without any problems.

That having been said, I disabled the admin interface and still had these problems.

I don't like having these ARP mappings manually configured as I lose the flexibility of being able to remove the cluster and still access the servers. Now of course when the permanent MAC is reinstated the router will get confused

over the IP/MAC combination of the dedicated cluster NIC.

All of the switches and routers are from the Foundry "Iron" range, be they EdgeIron's or BigIrons. These are enterprise/datacenter grade products.

I know Microsoft will be tempted to say, it's not our product that is having a problem, but it is their implementation of ARP within the NLB IP stack that is replying to the routers ARP queries. To date the router has had no problems when dealing with ARP replies form non−NLB Windows machines. NLB must be responding with something strange that is causing the router not to add the route.

Thanks and regards

David

"Chuck [MSFT]" <ctimon@xxxxxxxxxxxxxxxxxxxx> wrote in message news:up%23fPDy7HHA.3900@xxxxxxxxxxxxxxxxxxxxxxx

> They are not.....but then again most implementations are not setup this way. The only time static ARP entries are required is when using Multicast and not Unicast. The mapping that is required is the Multi−cast MAC mapped to the Unicast IP
>
> http://support.microsoft.com/default.aspx?scid=kb;en−us;197862
>
> −−
> Chuck Timon, Jr.
> Microsoft Corporation
> Windows Server 2008 Readiness Team
> This posting is provided 'AS IS" with no warranties, and confers no rights.
> "Bookham Measures" <bookham_measures_no_spam@xxxxxxxxx> wrote in message news:%23RS0Dvu7HHA.1184@xxxxxxxxxxxxxxxxxxxxxxx
>
> > My ISP has defined static ARP mappings on the first upstream router. This now allows me to ping the cluster and dedicated IP from a remote location.

From their analysis it seemed that the router was forgetting the MAC address of the relevant IPs and/or getting empty responses when doing an ARP query.

Of course, now if I remove the cluster configuration, I will not be able to connect as the permanent MAC address will be assigned back to the dedicated adapter's IP address in the ARP tables.

So − the question is, why are these static mappings required? I have read about multicast and the requirements of that, but we're using unicast. Devices on the local network seem to be ok, so why would the router discard or not accept the arp replies from the cluster NIC?

"Chuck [MSFT]" <ctimon@xxxxxxxxxxxxxxxxxxx> wrote in message news:ebw51no7HHA.1164@xxxxxxxxxxxxxxxxxxxxxx

> OK, go with that...do a trace route and determine where along the network path it drops off.
>
> −−
> Chuck Timon, Jr.
> Microsoft Corporation
> Windows Server 2008

Readiness
Team
This posting
is provided
'AS IS"
with no
warranties,
and confers
no rights.

"David
Morgan"
<microsoft_newsgroups@xxxxxxxxxxxxxxxxxxxxxx>
wrote in
message
news:On5kQOo7HHA.5136@xxxxxxxxxxxxxxxxxxxxxx

Using
Network
Monitor
I
see
the
pings
being
received
and
replies
being
sent
WHEN
ping
at
my
computer
reports
success.
All
other
times
I
see
nothing
related
to
my
continuing
ping.

This
must

point
to
some
other
router/switch
on
the
network,
not
forwarding
packets
when
required.

"David
Morgan"
<microsoft_newsgroups@xxxxxxxxxxxxxxxxxxxxxx>
wrote
in
message
news:%23X6d98n7HHA.5424@xxxxxxxxxxxxxxxxxxxxxx

Well
I
have
configured
the
cluster
on
NIC3,
a
standalone
DLink
card,
and
still
have
the
same
problem.

Only
this
adapter
has
a
default
gateway
defined.

I

have confirmed that the VLAN (and switch) are Layer 2 only.

I have disabled one of the Broadcom NICs, (NIC1 which had an IP from the same subnet). No difference.

So, the problem persists, from outside the local network pings either fail or respond with very high latency.

Installing a hub wouldn't make any difference as I am trying to get this working with only one cluster host. If I had more then I can see why a hub could help.

Why are the packets being created by the NLB driver not being routed properly... ? In fact,

I've just noticed pattern in the reply times. It looks like you get a long reply then a short one. Very strange.

C:\>ping −t x.x.16.125

Pinging x.x.16.125 with 32 bytes of data:

Request timed out.
Reply from x.x.16.125: bytes=32 time=1009ms TTL=118
Reply from x.x.16.125: bytes=32 time=28ms TTL=118
Request

timed
out.
Request
timed
out.
Request
timed
out.
Reply
from
x.x.16.125:
bytes=32
time=1509ms
TTL=118
Reply
from
x.x.16.125:
bytes=32
time=60ms
TTL=118
Request
timed
out.
Reply
from
x.x.16.125:
bytes=32
time=1509ms
TTL=118
Reply
from
x.x.16.125:
bytes=32
time=37ms
TTL=118
Request
timed
out.
Request
timed
out.
Reply
from
x.x.16.125:
bytes=32
time=1008ms
TTL=118
Reply
from
x.x.16.125:
bytes=32

time=29ms
TTL=118
Reply
from
x.x.16.125:
bytes=32
time=999ms
TTL=118
Reply
from
x.x.16.125:
bytes=32
time=53ms
TTL=118
Reply
from
x.x.16.125:
bytes=32
time=997ms
TTL=118
Reply
from
x.x.16.125:
bytes=32
time=40ms
TTL=118
Request
timed
out.
Request
timed
out.
Reply
from
x.x.16.125:
bytes=32
time=1009ms
TTL=118
Reply
from
x.x.16.125:
bytes=32
time=31ms
TTL=118
Request
timed
out.
Request
timed
out.
Request

timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Reply
from
x.x.16.125:
bytes=32
time=1008ms
TTL=118
Reply
from
x.x.16.125:
bytes=32
time=33ms
TTL=118
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request

timed
out.
Reply
from
x.x.16.125:
bytes=32
time=1008ms
TTL=118
Reply
from
x.x.16.125:
bytes=32
time=32ms
TTL=118
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Reply
from
x.x.16.125:
bytes=32
time=1008ms
TTL=118
Reply
from
x.x.16.125:
bytes=32
time=57ms
TTL=118
Request
timed
out.
Request
timed
out.
Request

timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Reply
from
x.x.16.125:
bytes=32
time=1508ms
TTL=118
Reply
from
x.x.16.125:
bytes=32
time=30ms
TTL=118
Reply

from
x.x.16.125:
bytes=32
time=998ms
TTL=118
Reply
from
x.x.16.125:
bytes=32
time=37ms
TTL=118
Reply
from
x.x.16.125:
bytes=32
time=996ms
TTL=118
Reply
from
x.x.16.125:
bytes=32
time=38ms
TTL=118
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request

timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Reply
from
x.x.16.125:
bytes=32
time=1508ms
TTL=118
Reply
from
x.x.16.125:
bytes=32
time=37ms
TTL=118
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Reply
from
x.x.16.125:
bytes=32
time=1007ms
TTL=118
Reply
from
x.x.16.125:
bytes=32
time=49ms
TTL=118
Request
timed
out.
Request

timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Reply
from
x.x.16.125:
bytes=32
time=1007ms
TTL=118
Reply
from
x.x.16.125:
bytes=32
time=30ms
TTL=118
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request

timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request

timed
out.
Reply
from
x.x.16.125:
bytes=32
time=1007ms
TTL=118
Reply
from
x.x.16.125:
bytes=32
time=29ms
TTL=118
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Reply
from
x.x.16.125:
bytes=32
time=1007ms
TTL=118
Reply
from
x.x.16.125:
bytes=32
time=34ms
TTL=118
Request

timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Reply
from
x.x.16.125:
bytes=32
time=1006ms
TTL=118
Reply
from
x.x.16.125:
bytes=32

time=32ms
TTL=118
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Reply
from
x.x.16.125:
bytes=32
time=1005ms
TTL=118
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Request
timed
out.
Reply
from
x.x.16.125:
bytes=32
time=1505ms
TTL=118
Reply
from
x.x.16.125:
bytes=32
time=28ms
TTL=118

Again,
I
should
re−iterate
that
pings
respond
perfectly
when
performed
from
another
host
on
the
same
subnet/x.x.16.0
network.

Many
thanks.

David

"Chuck
[MSFT]"
<ctimon@xxxxxxxxxxxxxxxxxxx>
wrote
in
message
news:eRnIbWJ7HHA.980@xxxxxxxxxxxxxxxxxxx

> Something
> to
> keep
> in
> mind....if
> the
> switch
> is
> Layer
> 3,
> the
> ports
> the
> NLB
> nodes
> plug
> into

must function at layer 2 or it won't work properly.

I'm still interested in the behavior with only one NIC in the picture.

With NLB, it is pretty common practice to break the problem down into its simplest components and see if it works, if it works then build

up
from
there.

−−
Chuck
Timon,
Jr.
Microsoft
Corporation
Windows
Server
2008
Readiness
Team
This
posting
is
provided
'AS
IS"
with
no
warranties,
and
confers
no
rights.

"David
Morgan"
<microsoft_newsgroups@xxxxxxxxxxxxxx
wrote
in
message
news:ejPPAjI7HHA.5772@xxxxxxxxxxxxx

> Yes,
> NIC1
> and
> NIC2
> (in
> each
> machine)
> are
> dual
> port
> Broadcom
> cards.
> They
> are

definitely not being teamed and I have tested the locally administered address is being set correctly on the cluster adapter. ARP reports everything as expected.

I will set−up the cluster using the DLink card, NIC3 which was added afterwards.

Regardless, none of this would suggest why it works locally

but not from outside the LAN. It is as if the router is intermittently not accepting the packets to/from the LAA address or something.

"Chuck [MSFT]" <ctimon@xxxxxxxxxxxxxxxxxxx> wrote in message news:ewVUlnC7HHA.1900@xxxxx

> Let's simplify this.....disable all but one NIC in each member of the NLB cluster, setup NLB on

it,
plug
the
NICs
into
a
hub,
remember
those?
:{>
then
uplink
the
hub
to
a
switch
port
and
test.

If
you
have
NLB
in
your
environment....keep
a
spare
dumb
hub
around....never
know
when
you
might
need
it.

––
Chuck
Timon,
Jr.
Microsoft
Corporation
Windows
Server
2008
Readiness
Team

This posting is provided 'AS IS" with no warranties, and confers no rights.

"Bookham Measures" <bookham_measures_no_spa wrote in message news:OlGc7c96HHA.5404@

Hello

We have set−up NLB cluster with two servers, ultimately for an IIS application.

Everything functions perfectly when the clients are on the same subnet as defined

by their IP address.

Hosts outside the subnet, i.e. come via a router, cannot get reliable ping responses. They get a mixture of "request timed out" or a delayed reply, between 54ms to 1510ms. The ping summary always reports more than 60% packet loss. This applies to the cluster IP

address and the dedicated IP on the cluster NIC in each server. The mixture of timeouts and replies come at different times when running simultaneous pings to the three IP addresses.

All configuration has been performed via the Network Load Balancing Manager.

As we have plenty of adapters the

cluster has been configured in Unicast mode. We have set−up a VLAN on our switch and plugged the dedicated cluster NICs in to those ports. The configuration is as follows.

Server 1
−−−−−−−−−
NIC1: x.x.16.121
DG x.x.16.1
in VLAN
NIC2: x.x.16.123
DG x.x.16.1
NIC3: 192.168.1.21
DG not set.

Server
2
−−−−−−−−−
NIC1:
x.x.16.122
DG
x.x.16.1
in
VLAN
NIC2:
x.x.16.124
DG
x.x.16.1
NIC3:
192.168.1.22
DG
not
set.

Cluster
−−−−−−−
Server
1
NIC1
Priority
1
Server
2
NIC1
Priority
2
Cluster
IP
x.x.16.125
Equal

NIC3
on
each
server
is
numbered
so
as
to
communicate
with
the
database
cluster.
I

have tried removing the default gateway from NIC1/2 to see if I can different results, but I cannot. Interestingly, when one of the servers is offline the problem persists. Remember, this problem only occurs from outside the x.x.16.0 subnet. Hosts on the same subnet (not in the VLAN), have no problems

communicating to the cluster or NIC1 IP addresses.

When the cluster is deleted via the Manager, the IPs on NIC1 in each machine start responding to pings normally with good times.

NIC2 in both servers responds to pings from anywhere satisfactorily, the whole time, (they are not in the

VLAN used by NIC1s).

NIC1 and NIC2 in each machine are a "Broadcom NetXtreme Gigabit Ethernet" dual port adapter and have the latest driver from the IBM website. Both servers are IBM System−X 3850 M2s running Windows 2003 Server R2 SP2. (32 Bit). Quad Xeon 2.5Ghz with 3 Gb RAM.

The switch is a Foundry EdgeIron 48G.

Where can I go next to troubleshoot this problem? The fact that the IPs respond normally from everywhere when there is no cluster configured, must mean that there is something wrong at the NLB driver level.

Many thanks in advance.

David

.