Infrastructure - Story #2922

review and fix exposure of individual CNs on nodelist

2012-06-12 19:14 - Matthew Jones

Status:	Closed	**Start date:**	2012-06-12
Priority:	Normal	**Due date:**	2013-04-27
Assignee:	Dave Vieglais	**% Done:**	100%
Category:	Documentation	**Estimated time:**	0.00 hour
Target version:	2013.16-Block.2.4		
Story Points:			

Description

Design of the DataONE CN system was predicated on the existence of a single, virtual CN node that would reside at one address (cn.dataone.org) and handle all requests and operations via delegation to hidden backend processes (sub-CNs, like cn-ucsb-1, cn-unm-1, and cn-orc-1) that can be dynamically added and removed without impact on the system. This was designed to provide failover capabilities, rolling updates of CNs, and eventually load balancing and scaling (through the addition of more CN nodes). The current implementation actually registers these individual CN subnodes as if they were full and alternate CNs, and so MNs could try to contact these CNs directly, which would compromise our design goals above. I think we need to review how we have implemented the registration of CNs, and change to a system in which only the "urn:node:CN" is registered, and the other CNs are invisible to clients (but might have their subjects registered as valid subjects for that CN). What we don't want is for MNs and clients to be depending upon the existence of a given sub-CN, and certainly not caching the hostnames, baseURLs, or IP addresses for those subCNs. We had an IRC discussion about this topic, and its recorded below for reference.

IRC conversation regarding CN registration and subjects

matt: I was just talking with the folks from CDL -- they said they want to use IP filtering to restrict access to the CNs
[10:37am] benMac: it's a big red panic! button, so easy to brush up against it
[10:37am] matt: but in my opinion people hould not be relying on the IP numbers of CNs
[10:37am] benMac: *access to them from the CNs*
[10:37am] matt: right
[10:37am] matt: they shouldn't even be relying on the baseurls of CNs
[10:38am] benMac: can they filter using hostnames that they then look up IPs for?
[10:38am] matt: really, the only reliable thing identifying a CN is its subject in a client cert
[10:38am] matt: the hostnames might change too
[10:38am] vieglais: as long as the rules can be defined in apache config then all is good for hostname vs ip
[10:38am] matt: the onbly fixed hostname is cn.dataone.org
[10:38am] matt: the others behind RR might change
[10:38am] matt: ie, we might add in cn-unm-2 sometime
*[10:38am] vieglais: heads up - NASA currently *requires* at least email address for access to all download content*
[10:39am] vieglais: so, Mercury will be heading to tier 2 quickly
[10:39am] matt: so they will need to set acls that require user verification
[10:39am] vieglais: quickly = within the next month or so
[10:39am] vieglais: yep, no public access to any DAAC content
[10:39am] vieglais: well, not anonymous access
[10:40am] matt: our system handles that ! (assuming we can get verification up and running
[10:40am] vieglais: yep
[10:40am] matt: back to CDL -- so I was looking at the node list
[10:41am] matt: I was wondering why the CN node didn't linst CNUNM1, CNUCSB1, and CNORC1 as valid subjects
[10:41am] shaun: hi benMac, matt -- just saw your message from friday about KNB -- all the GlobalMarine stuff is off of KNB, so don't worry about us
[10:41am] matt: and, I don't think I ever created a node certificiate for urn:node:CN, so how does that work
[10:42am] matt: shaun -- thanks!
[10:43am] benMac: https://cn.dataone.org/cn/v1/node
[10:43am] benMac: the RR entry is fictional and will never be made to make a call
[10:44am] matt: really? the MNs will call it routinely
[10:44am] benMac: fictional is not the correct word
[10:44am] matt: it is the only node the MNs will actually call
[10:44am] benMac: you're muddying things up, matt
[10:44am] matt: sorry
[10:44am] matt: just confused
[10:44am] benMac: MNs don't look at the node list to call a specific node
[10:45am] matt: ok, then how do they know the baseurl for cn.dataone.org?

[10:45am] benMac: they might look at the nodelist to see what valid CN subjects exist, but they will ALWAYS us "cn.dataone.org" as the base url
[10:45am] benMac: all clients (MNs included) must be configured to know the CN url
[10:45am] matt: that's not the baseurl -- its just the hostname
[10:46am] benMac: notices i didn't use the proper noun "baseURL"!
[10:46am] benMac: ammended: they might look at the nodelist to see what valid CN subjects exist, but they will ALWAYS us "cn.dataone.org/cn" as the base url
[10:47am] matt: ok, right
[10:47am] benMac: you can't get he nodelist to find the CN url in order to get the node list....
[10:47am] matt: so, if they make a call to cn.dataone.org/cn, and cn-ucsb-1 responds with its client cert, how does the MN that is valid for urn:node:CN?
[10:48am] benMac: "and cn-ucsb-1 responds with its client cert" != how it works
[10:48am] matt: really? which cert does it use when responding?
[10:48am] benMac: "responding" doesn't make sense
[10:49am] matt: sure it does
[10:49am] matt: http request leads to http response
[10:49am] benMac: it is the *.dataone.org certificate
[10:49am] matt: ah, right
[10:49am] benMac: that is the only thing that identifies the server at that stage
[10:49am] matt: ok, so lets go the other way around
[10:50am] matt: if the CN calls a retricted MN method, how does it identify itself
[10:50am] benMac: when the CN *makes* a call to a MN, a particular DataONE client certificate is used
[10:50am] matt: and how is that linked to urn:node:CN?
[10:50am] benMac: and the particular subject of the CoordNode is used
[10:50am] benMac: it will never be urn:ndoe:CN
[10:51am] matt: so how does the MN know its a legit CN?
[10:51am] benMac: I believe that nodeId is used in replication to signify that the object has been synched to the CN[s]
[10:51am] matt: it "knows" about cn.dataone.org, aka "urn:node:CN"
[10:51am] matt: it doesn't "know" about the others
[10:52am] matt: and they can change over time
[10:52am] matt: come and go
[10:52am] benMac: it "knows" what's in the nodeList
[10:52am] benMac: and trusts the CNs that are registered in the nodelist
[10:52am] matt: ah, there it is
[10:52am] benMac: whenever it gets a call with a CN certificate it checks the nodelist for a match
[10:52am] matt: so type = cn
[10:52am] benMac: yes
[10:53am] matt: critical
[10:53am] matt: do we check that in metacat?
[10:53am] benMac: type?, yes
[10:54am] matt: so, the fact that we've registered cn-ucsb-1, cn-unm-1, and cn-orc-1 is somewhat troubling to me from a failover perspective
[10:54am] robert: what do you mean by registered?
[10:54am] matt: an enterprising MN or client might query the nodelist for type=cn, and get that list
[10:54am] robert: we have the in the nodelist?
[10:54am] benMac: again, they are never directly *CALLED*
[10:54am] matt: then start making calls to a specific CN, rather than the general urn:node:CN
[10:54am] benMac: sure
[10:54am] benMac: anyone could do that
[10:54am] robert: that is a problem
[10:55am] matt: which would break failover if they cache it
[10:55am] benMac: given an IP, can't you figure out what the host is?
[10:55am] matt: in my mind, the only kowledge of the CNs hsould be the urn:node:CN entry
[10:55am] matt: all the others are just load-balanced instances of that server
[10:55am] benMac: that's not how it's designed, sorry
[10:55am] vieglais: so apache deny / allow using host name requires reverse DNS lookup
[10:56am] matt: that's not how it has evolved - it was how it was designed
[10:56am] vieglais: which we don't have, so IP address restriction is the only option for protecting getLogRecords
[10:56am] benMac: we could share the same Client cert on each CN and it would use the urn:node:CN subject
[10:57am] matt: or we could register the 3 client certs as all valid subjects for urn:node:CN
[10:57am] benMac: but maybe chris and robert have a special need for havin individual CN mchines in the nodelist?
[10:57am] matt: and hisde their individual resigtrations
[10:58am] robert: ok, we're not changing this today, right?
[10:58am] matt: no.
[10:58am] matt: I don't see any changes possible today.
[10:58am] robert: or next week, or next month.
[10:58am] robert: (maybe two)

[10:58am] matt: this just arose because I started reviewing the CN in more detail after getting questios from CDL
[10:58am] benMac: why not? adding additional subjects to urn:node:CN is easy
[10:58am] robert: ok, its a very good discussion that we should create a faq from
[10:59am] robert: i use the node list for distributing execution
[10:59am] robert: distributed execution
[10:59am] chris: so, why did you make individual certs for the CNs, then matt? Did you intend those certs to be used for something else?
[10:59am] benMac: well, there's a compelling reason to keep it as is!
[10:59am] robert: to change that would require a bit of work
[11:00am] robert: ya, i would love to redesign it, but that is a big piece of code to write
[11:00am] matt: I made them because we needed to identify individual CN nodes for replication and other places, but I wasn't expecting them to be registered as nodes
[11:01am] matt: I agree with Robert -- let's table this discussion and return to it.
[11:02am] matt: in my mind, the guiding principal is: there is only one CN
[11:02am] matt: everything behind the scenes is implementation details for failover

Related issues:

Related to Infrastructure - Story #3269: The Nodelist should only expose the ...	**Rejected**

History

#1 - 2012-06-14 15:58 - Dave Vieglais

The intent of the node list is to provide a list of nodes participating in the DataONE system, as such it should include a list of all nodes as it currently does.

Interactions with the coordinating nodes should always be through the "virtual" cn as currently expressed through the RR DNS entry for the respective environments, however since each CN is a mirror of each other, there is no harm for a client (MN or otherwise) to interact directly with a specific CN instance, and in fact there may be benefit to do so for some operations (e.g. taking advantage of server side caching and index warming during search).

The more important issue is that clients should never rely on availability of a particular CN (or indeed *any* specific node), and so should always consult the node list "reasonably often" to determine the baseURL of any node for further interaction. This means that the node list *must* always exist at a well known URL, as it is essential for reliable interaction with the DataONE platform.

One could argue that rather than removing the specific CN instances from the node list, the virtual CN entry should instead be removed. It would not change the way clients interact with DataONE other than a tendency to interact mostly with a specific CN for a session. The problem of course is that clients will tend to select the first entry on the list, so keeping the RR DNS entry first in the list will help with balancing load across the CNs.

There should perhaps be further discussion on this, but I am not convinced that any changes are necessary.

#2 - 2012-07-05 22:19 - Rob Nahf

We also cannot overlook / neglect auditing and testing use-cases in our design. Currently, we can get the extent of the nodes in an environment by querying the nodelist, and from there can compare the different CN implementations for consistency (let's say: the number of objects on each system).

Also, /resolve relies on the NodeList to convert a replica into a url. Since the CNs participate in providing replicas, and the virtual round-robin NodeReference is used, the RR CN node needs to remain in the NodeList, otherwise a special mechanism would be needed for getting the baseUrl of the RR CN.

#3 - 2012-07-09 14:05 - Dave Vieglais

- Target version changed from Sprint-2012.25-Block.4.1 to Sprint-2012.33-Block.5.1

#4 - 2012-08-31 14:14 - Dave Vieglais

- Target version changed from Sprint-2012.33-Block.5.1 to Sprint-2012.37-Block.5.3

#5 - 2012-10-08 16:43 - Dave Vieglais

- Milestone changed from CCI-1.1 to None

- Target version changed from Sprint-2012.37-Block.5.3 to Sprint-2012.44-Block.6.2

- translation missing: en.field_remaining_hours set to 0.0

- Due date set to 2012-11-10

this topic requires further discussion.

#6 - 2012-12-12 18:41 - Dave Vieglais

- Target version changed from Sprint-2012.44-Block.6.2 to 2013.2-Block.1.1

- Due date changed from 2012-11-10 to 2013-01-19

#7 - 2013-03-01 19:16 - Chris Jones

- Due date changed from 2013-01-19 to 2013-04-27

- Target version changed from 2013.2-Block.1.1 to 2013.16-Block.2.4

#8 - 2013-08-02 12:18 - Dave Vieglais

- Status changed from New to Closed

With respect to CN authentication (i.e. when a CN calls a method on an MN, and the MN needs to determine if the call is from a valid CN).

a) CNs should use a common client certificate with the identity "urn:node:CN"
b) CN's may use a node specific certificate to identify themselves, however the equivalent identity mapping to the virtual CN must be provided in the certificate, and the primary identity should be of the virtual CN
c) If using (b), then MNs should support examination of alternate identity mappings presented in the certificate

Closing this ticket - may be re-opened at some point in the future for further discussion, but no changes required at this point.