

## JobSub - Bug #18608

### jobsub\_fetchlog --list not contacting all jobsub servers behind haproxy

12/19/2017 12:01 PM - Dennis Box

<b>Status:</b>	Closed	<b>Start date:</b>	12/19/2017
<b>Priority:</b>	Normal	<b>Due date:</b>	
<b>Assignee:</b>	Dennis Box	<b>% Done:</b>	0%
<b>Category:</b>		<b>Estimated time:</b>	0.00 hour
<b>Target version:</b>	v1.2.7	<b>Spent time:</b>	1.00 hour
<b>First Occurred:</b>		<b>Stakeholders:</b>	
<b>Occurs In:</b>			
<b>Description</b>			
See INC000000911154			

### History

#### #1 - 12/19/2017 04:51 PM - Shreyas Bhat

- Status changed from New to Work in progress

Fetchlog does indeed default to 'https://fifebatch.fnal.gov:8443' if nothing is set.

#### #2 - 02/09/2018 02:55 PM - Shreyas Bhat

Also see INC000000925695

#### #3 - 02/14/2018 04:45 PM - Dennis Box

- Status changed from Work in progress to Resolved

#### #4 - 02/15/2018 07:08 AM - Arthur Kreymer

I still see the problem in jobsub\_client v1\_2\_6

```
ssh minospro@minos-data.fnal.gov
```

```
export JOBSUB_GROUP=minos
export X509_USER_PROXY=/opt/minospro/minospro.Production.proxy
setup jobsub_client v1_2_6
```

```
jobsub_fetchlog --list
JobsubJobID CreationDate for user minospro in Accounting Group minos
2992128.0@jobsub02.fnal.gov Mon Jan 15 14:01:10 2018
...
17 jobs; 0 completed, 0 removed, 0 idle, 0 running, 0 held, 0 suspended
```

```
jobsub_fetchlog --list --jobsub-server=jobsub01.fnal.gov
JobsubJobID CreationDate for user minospro in Accounting Group minos
2872219.0@jobsub01.fnal.gov Tue Jan 16 03:04:15 2018
...
3798722.0@jobsub01.fnal.gov Thu Feb 15 06:12:16 2018
20 jobs; 0 completed, 0 removed, 0 idle, 0 running, 0 held, 0 suspended
```

```
jobsub_fetchlog --list --jobsub-server=jobsub02.fnal.gov
JobsubJobID CreationDate for user minospro in Accounting Group minos
2992128.0@jobsub02.fnal.gov Mon Jan 15 14:01:10 2018
...
17 jobs; 0 completed, 0 removed, 0 idle, 0 running, 0 held, 0 suspended
```

**#5 - 02/27/2018 11:21 AM - Shreyas Bhat**

- Status changed from Resolved to Work in progress

Testing now with the new RC.

UPDATE: Testing failed. I see what Art sees:

```
-bash-4.1$ jobsub_q -G nova --jobsub-server=jobsub-dev.fnal.gov
JOBSUBJOBID      OWNER      SUBMITTED      RUN_TIME      ST PRI  SIZE CMD
8020.0@htcjsdev01.fnal.gov  sbhat      02/27 11:21    0+00:01:04 R  0   0.0 probe_20180227_1121
21_3207292_0_1_wrap.sh
8020.1@htcjsdev01.fnal.gov  sbhat      02/27 11:21    0+00:01:04 R  0   0.0 probe_20180227_1121
21_3207292_0_1_wrap.sh
8020.2@htcjsdev01.fnal.gov  sbhat      02/27 11:21    0+00:01:04 R  0   0.0 probe_20180227_1121
21_3207292_0_1_wrap.sh
8020.3@htcjsdev01.fnal.gov  sbhat      02/27 11:21    0+00:01:04 R  0   0.0 probe_20180227_1121
21_3207292_0_1_wrap.sh
6616.0@htcjsdev02.fnal.gov  sbhat      02/14 15:45    0+00:01:29 H  0   0.0 probe_20180214_1545
25_1948569_0_1_wrap.sh
```

5 jobs; 0 completed, 0 removed, 0 idle, 4 running, 1 held, 0 suspended

```
-bash-4.1$ jobsub_q -G nova --jobsub-server=jobsub-dev.fnal.gov
JOBSUBJOBID      OWNER      SUBMITTED      RUN_TIME      ST PRI  SIZE CMD
6616.0@htcjsdev02.fnal.gov  sbhat      02/14 15:45    0+00:01:29 H  0   0.0 probe_20180214_1545
25_1948569_0_1_wrap.sh
```

```
-bash-4.1$ jobsub_fetchlog -G nova --list --jobsub-server jobsub-dev.fnal.gov
JobsSubJobID CreationDate for user sbhat in Accounting Group nova
1890.0@htcjsdev02.fnal.gov  Mon Feb  5 13:42:16 2018
1891.0@htcjsdev02.fnal.gov  Mon Feb  5 13:51:43 2018
6616.0@htcjsdev02.fnal.gov  Wed Feb 14 15:45:25 2018
3 jobs; 0 completed, 0 removed, 0 idle, 0 running, 0 held, 0 suspended
```

So I'm afraid we can't call this done yet.

**#6 - 02/28/2018 08:53 AM - Arthur Kreymer**

The issue persists in jobsub\_client v1\_2\_6\_2\_rc1

#### #7 - 03/07/2018 11:48 AM - Dennis Box

- Target version changed from v1.2.6 to v1.2.6.2

#### #8 - 03/07/2018 04:02 PM - Shreyas Bhat

Alright - finally got around to looking at this.

The issue occurs because of two things:

### get\_jobsub\_server\_aliases doesn't work the way we think it should

In client/jobsub\_fetchlog, the function that performs jobsub\_fetchlog --list, list\_sandboxes(options), tries to get all the aliases behind HAProxy by using a function called get\_jobsub\_server\_aliases, and calling it on the DNS alias (e.g. fifebatch.fnal.gov:8443). The first problem is that this doesn't work the way we intend. It should be returning something like ['https://jobsub01.fnal.gov:8443', 'https://jobsub02.fnal.gov:8443'], but instead, when I isolated it and tried running get\_jobsub\_server\_aliases(fifebatch.fnal.gov):

```
>>> get_jobsub_server_aliases("fifebatch.fnal.gov")
['https://fifebatch.fnal.gov:8443']
```

The code I used to test it was the following:

```
#!/usr/bin/python

import socket

def is_port_open(server, port):
    is_open = False
    server = server.strip().replace('https://', '')
    sp = server.split(':')
    server = sp[0]
    if len(sp) == 2 and not port:
        port = sp[1]

    try:
        serverIP = socket.gethostbyname(server)
        s = socket.socket(socket.AF_INET, socket.SOCK_STREAM)
        result = s.connect_ex((serverIP, int(port)))
        s.close()
        if result == 0:
            is_open = True
    except:
        pass
    return is_open

def get_jobsub_server_aliases(server):
    # Set of hosts in the HA mode
    aliases = []

    host_port = server.replace('https://', '')
    host_port = host_port.replace('/', '')
    tokens = host_port.split(':')
    if tokens and (len(tokens) <= 2):
        host = tokens[0]
        if len(tokens) == 2:
            port = tokens[1]
        else:
            port = 8443
        # Filter by TCP ports (5th arg = 6 below)
        addr_info = socket.getaddrinfo(host, port, 0, 0, 6)
        for info in addr_info:
            # Each info is of the form (2, 1, 6, '', ('131.225.67.139', 8443))
            ip, p = info[4]
            js_s = 'https://%s:%s' % (
                socket.gethostbyaddr(ip)[0], p)
            if is_port_open(socket.gethostbyaddr(ip)[0], p):
                aliases.append(js_s)

    if not aliases:
```

```
# Just return the default one
aliases.append(server)
if len(aliases) > 1:
    random.shuffle(aliases)

return aliases
```

## Because of this, `jobsub_fetchlog` only runs on the server it happens to contact (that fifebatch.fnal.gov's HAProxy load balances sends its query to)

`jobsub_fetchlog --list` uses the endpoint '`<server>/jobsub/acctgroups/<experiment>/jobs/<user>/sandbox'`. The behavior of this endpoint is defined in `server/webapp/sandboxes.py`. Here, we can see that there is no code that tries to look at other schedds (jobsub servers). We simply run a `find` in the correct dir, find the sandboxes, and report on them.

## `jobsub_q` gets around this by using `condor_q -schedd` to get a list of the schedds

Speaking of this, `jobsub_q` doesn't run into this issue because the client sends `jobsub_q` requests to endpoints that look like `"/jobsub/jobs/..."` or `"/jobs/users/..."`. The actions at these endpoints are defined in `server/webapp/job.py` and `server/webapp/users_jobs.py`. Both of these use the `server/webapp/condor_commands.py` module to run `ui_condor_q`. This function calls the function `schedd_list()` in the same module, which in turn runs

```
condor_status -schedd -af name
```

That, run on any of the production jobsub servers, returns:

```
-bash-4.1$ condor_status -schedd -af name
gpce03.fnal.gov
gpce04.fnal.gov
jobsub01.fnal.gov
jobsub02.fnal.gov
-bash-4.1$ hostname
jobsub01.fnal.gov
```

This list is then used in the `condor_q` command that `ui_condor_q` runs. This is why this issue isn't present for `jobsub_q` commands.

**#9 - 03/19/2018 07:27 PM - Dennis Box**

- Target version changed from v1.2.6.2 to v1.2.7

**#10 - 06/15/2018 03:16 PM - Dennis Box**

- Status changed from Work in progress to Resolved

- Assignee changed from Shreyas Bhat to Dennis Box

I figured this one out while investigating [#20181](#), will check in fix shortly

**#11 - 07/11/2018 04:12 PM - Dennis Box**

- Status changed from Resolved to Closed