Arvados - Story #16809

[keep-web] Check V4 signature on S3 requests, don't require sending entire Arvados token as AccessKey

09/08/2020 02:22 PM - Tom Clegg

Status:	Resolved	**Start date:**	09/22/2020
Priority:	Normal	**Due date:**	
Assigned To:	Tom Clegg	**% Done:**	100%
Category:	Keep	**Estimated time:**	0.00 hour
Target version:	2020-10-07 Sprint		

Description

If a client has Arvados V2 token "v2/zzzzz-gj3su-077z32aux8dg2s1/3kg6k6lzmp9kj5cpkcoxie963cmvjahbt2fod9zru30k1jqdmi" it should be able to use the S3 API with

- AccessKey: zzzzz-gj3su-077z32aux8dg2s1
- SecretKey: 3kg6k6lzmp9kj5cpkcoxie963cmvjahbt2fod9zru30k1jqdmi

This avoids revealing the secret key to a remote host in case of a misconfigured endpoint, etc., and aligns better with S3 clients' expectation that AccessKey does not need to be protected as a sensitive secret.

(Currently, the client has to send the secret part of the token as AccessKey.)

Subtasks:

Task # 16824: Review 16809-s3-v4-signature	**Resolved**

Related issues:

Related to Arvados Epics - Story #16360: Keep-web supports S3 compatible inte...	**Resolved**	**07/01/2020**	**04/30/2021**

Associated revisions

Revision 3facf89b - 10/06/2020 05:56 PM - Tom Clegg

Merge branch '16809-s3-v4-signature'

closes #16809

Arvados-DCO-1.1-Signed-off-by: Tom Clegg <tom@tomclegg.ca>

History

#1 - 09/08/2020 02:22 PM - Tom Clegg

- Related to Story #16360: Keep-web supports S3 compatible interface added

#2 - 09/09/2020 04:45 PM - Peter Amstutz

- Target version changed from Arvados Future Sprints to 2020-09-23 Sprint

#3 - 09/09/2020 04:45 PM - Peter Amstutz

- Assigned To set to Tom Clegg

#4 - 09/21/2020 02:48 PM - Tom Clegg

- Status changed from New to In Progress

#5 - 09/22/2020 01:40 PM - Tom Clegg

16809-s3-v4-signature @ b15d39ec33dde9639f09bd1aff22fde7806aa24a -- https://ci.arvados.org/view/Developer/job/developer-run-tests/2111/

#6 - 09/22/2020 06:41 PM - Peter Amstutz

As mentioned in chat: the intended integration will use an OIDC token to authorize with the Arvados S3 API, which means we need to have a way of accepting that. The proposed solution is to allow (the same) valid token to appear in both the AccessKey and SecretKey fields.

#7 - 09/22/2020 07:01 PM - Tom Clegg

16809-s3-v4-signature @ 580d77ef4d6b244971bc26c649e017e912ca8737 -- https://ci.arvados.org/view/Developer/job/developer-run-tests/2113/

#8 - 09/23/2020 01:26 PM - Lucas Di Pentima

I've tried connecting with Cyberduck, the client that previously worked by using the secret v2 token part on both accesskey & secretkey but now it doesn't work even if I try that. The logs at keep-web are:

```
2020-09-23_13:16:04.73728 {"RequestID":"req-1j71jt8h8t9w81ls154b","level":"info","msg":"request","remoteAddr":"127.0.0.1:54204","reqBytes":0,"reqForwardedFor":"10.1.1.2","reqHost":"10.1.1.7","reqMethod":"GET","reqPath":"xyt5u-j7d0g-rf3t04o4zul5lmk/","reqQuery":"versioning","time":"2020-09-23T13:16:04.737176196Z"}
2020-09-23_13:16:04.75945 {"RequestID":"req-1j71jt8h8t9w81ls154b","level":"info","msg":"response","remoteAddr":"127.0.0.1:54204","reqBytes":0,"reqForwardedFor":"10.1.1.2","reqHost":"10.1.1.7","reqMethod":"GET","reqPath":"xyt5u-j7d0g-rf3t04o4zul5lmk/","reqQuery":"versioning","respBody":"signature verification failed: signature does not match\n","respBytes":56,"respStatus":"Forbidden","respStatusCode":403,"time":"2020-09-23T13:16:04.759337471Z","timeToStatus":0.022144,"timeTotal":0.022153,"timeWriteBody":0.000008}
2020-09-23_13:16:04.76302 {"RequestID":"req-17004400sn44410ic55k","level":"info","msg":"request","remoteAddr":"127.0.0.1:54208","reqBytes":0,"reqForwardedFor":"10.1.1.2","reqHost":"10.1.1.7","reqMethod":"GET","reqPath":"xyt5u-j7d0g-rf3t04o4zul5lmk/","reqQuery":"encoding-type=url\u0026max-keys=1000\u0026prefix\u0026delimiter=%2F","time":"2020-09-23T13:16:04.762891998Z"}
2020-09-23_13:16:04.78465 {"RequestID":"req-17004400sn44410ic55k","level":"info","msg":"response","remoteAddr":"127.0.0.1:54208","reqBytes":0,"reqForwardedFor":"10.1.1.2","reqHost":"10.1.1.7","reqMethod":"GET","reqPath":"xyt5u-j7d0g-rf3t04o4zul5lmk/","reqQuery":"encoding-type=url\u0026max-keys=1000\u0026prefix\u0026delimiter=%2F","respBody":"signature verification failed: signature does not match\n","respBytes":56,"respStatus":"Forbidden","respStatusCode":403,"time":"2020-09-23T13:16:04.784578564Z","timeToStatus":0.021677,"timeTotal":0.021683,"timeWriteBody":0.000006}
```

#9 - 09/23/2020 03:54 PM - Tom Clegg

- Target version changed from 2020-09-23 Sprint to 2020-10-07 Sprint

#10 - 09/25/2020 02:25 PM - Tom Clegg

Reproduced (possibly) same issue with s3cmd. We're supposed to use req.Host to get the requested host:port, not req.URL.Host. Turns out req.URL.Host is sometimes populated, though, so this wasn't caught by tests. Maybe it depends on the incoming protocol (http/https/http2)? Anyway, this fixes it for s3cmd.

I added s3cmd to the dev dependencies for arvados-server install -type dev but made the s3cmd-based test optional for now so existing test workers/images don't fail.

16809-s3-v4-signature @ 6a67a1b576bb695e9b274c277b7220590da1a39d -- https://ci.arvados.org/view/Developer/job/developer-run-tests/2120/

#11 - 09/25/2020 10:53 PM - Lucas Di Pentima

- I was able to connect both using the secret token on all fields and using the token uuid on the access key.
- When I tried to upload a file through the S3 Cyberduck client, got this message: Request Error: Header "x-amz-content-sha256" set to the hex-encoded SHA256 hash of the request payload is required for AWS Version 4 request signing, please set this on: PUT https://10.1.1.7:9002/x20qf-j7d0g-r15tr3mjgtgqwun/New%20collection/20200821%20rho%20ophiuchi%20processed.tif HTTP/1.1. Please contact your web hosting service provider for assistance. This is the same as https://dev.arvados.org/issues/16535#note-21, but I'm not sure if that was dismissed for some reason on that other ticket or if it's a regression. This happens with a 112 MB file but doesn't with a 87 MB file.
- Haven't tried it but it seems Cyberduck has a CLI for Linux: https://duck.sh/ — maybe it's worth a shot at using it for additional automated tests?
- s3cmd isn't available in Debian Buster stock, but it is as a backport. There's a s4cmd on buster that's supposed to be a replacement but I'm not sure if its API is 100% compatible.

#12 - 10/05/2020 06:43 PM - Tom Clegg

I'm not sure what to make of the Cyberduck error. Possibly related to https://trac.cyberduck.io/ticket/11038, although that "only affects files with exactly 104857600B in file length." Not sure it's worth adding a 3rd party package repo for the sake of testing cyberduck in the test suite, but I'll see if I can reproduce the problem with the CLI version.

Meanwhile, I've added a test that uses the official AWS client library to upload + download a 100MB file, and removed s3cmd from lib/install so it doesn't fail on Debian buster. Tried s4cmd but it doesn't seem to have plain-http or no-verify options, so it's not exactly convenient for testing.

16809-s3-v4-signature @ cc8cffec8e1c612b6be03f4446ab6beebf479f5b -- https://ci.arvados.org/view/Developer/job/developer-run-tests/2132/

#13 - 10/06/2020 03:13 PM - Tom Clegg

The error message "Request Error: Header "x-amz-content-sha256" set to the hex-encoded SHA256 hash" follows a 405 Method Not Allowed response to a POST request. I suspect it is trying to use the POST Object API and, when that fails with 405, falling back to PUT Object, which fails in the client library (without contacting the server again) because Cyberduck doesn't populate the content hash header [with "UNSIGNED-PAYLOAD"].

```
> POST /ce8i5-4zz18-ykh8x2x89lq6iet/101M?uploads HTTP/1.1
> Date: Tue, 06 Oct 2020 15:06:23 GMT
> Content-Type: application/octet-stream
> x-amz-content-sha256: e3b0c44298fc1c149afbf4c8996fb92427ae41e4649b934ca495991b7852b855
> Host: collections.ce8i5.arvadosapi.com
> x-amz-date: 20201006T150623Z
> Authorization: ********
> Content-Length: 0
> Connection: Keep-Alive
> User-Agent: Cyberduck/7.6.2.33520 (Linux/4.9.0-13-amd64) (amd64)
< HTTP/1.1 405 Method Not Allowed
< Server: nginx/1.14.0 (Ubuntu)
< Date: Tue, 06 Oct 2020 15:06:23 GMT
< Content-Type: text/plain; charset=utf-8
< Content-Length: 19
< Connection: keep-alive
Upload 101M failed. Request Error: Header "x-amz-content-sha256" set to the hex-encoded SHA256 hash of the req
uest payload is required for AWS Version 4 request signing, please set this on: PUT https://collections.ce8i5.
arvadosapi.com:443/ce8i5-4zz18-ykh8x2x89lq6iet/101M HTTP/1.1. Please contact your web hosting service provider
 for assistance.
```

#14 - 10/06/2020 04:12 PM - Lucas Di Pentima

Thanks for the detailed investigation. This LGTM, please merge.

#15 - 10/06/2020 06:03 PM - Anonymous

- % Done changed from 0 to 100

- Status changed from In Progress to Resolved

Applied in changeset arvados|3facf89bf048487ee718fe15d012b489f2d407b7.

#16 - 10/07/2020 02:11 AM - Peter Amstutz

- Release set to 25