

Arvados - Feature #14199

[keepstore] copy block from remote keepstore to local keepstore

09/12/2018 02:49 PM - Peter Amstutz

Status:	Resolved	Start date:	10/04/2018
Priority:	Normal	Due date:	
Assigned To:	Tom Clegg	% Done:	100%
Category:		Estimated time:	0.00 hour
Target version:	2018-10-17 sprint		
Description			
API to instruct keepstore to fetch a remote block, save it, and return signed locator (variation on "proxy a block read request" with the added feature of saving the block).			
Proposed API:			
Request header "X-Keep-Signature: local, time=T" (where T is an RFC3339 timestamp) implies the block should be retained by keepstore.			
The response includes the header "X-Keep-Locator: abc+123+Aabc"			
The client will do a HEAD request (equivalent to the GET request) on remote signed locator and look for a response header that has a new locator with local block signature.			
All keepstore GET and HEAD responses for block requests should support this new header/response			
When the "X-Keep-Signature" request header is provided, the timestamp on the underlying block must be updated before a new signature can be returned.			
Responses with an X-Keep-Locator header must also include a "Vary: X-Keep-Signature" header.			
Subtasks:			
Task # 14290: Review 14199-copy-from-remote			Resolved
Related issues:			
Related to Arvados - Feature #14259: [SDK] Python collection class uses copy ...		Resolved	10/30/2018
Related to Arvados - Feature #14406: [SDK] Go collection uses copy remote bl...		Resolved	11/05/2018

Associated revisions

Revision 282536bb - 10/09/2018 08:07 PM - Tom Clegg

Merge branch '14199-copy-from-remote'

refs #14199

Arvados-DCO-1.1-Signed-off-by: Tom Clegg <tclegg@veritasgenetics.com>

History

#1 - 09/17/2018 08:25 PM - Peter Amstutz

- Description updated

#2 - 09/26/2018 02:07 PM - Peter Amstutz

Federated block queries: currently do GET with a signed locator, returns content

We want to retrieve a federated block, save it, and returned the signed locator

Do a HEAD request (equivalent to the GET request) and look for a header that has a new locator with local block signature.

#3 - 09/26/2018 03:42 PM - Peter Amstutz

- Subject changed from [controller] Convert +R to +A in collection POST and PUT operations to [keepstore] copy remote block to local keepstore

- Description updated

#4 - 09/26/2018 03:43 PM - Peter Amstutz

- Related to Feature #14259: [SDK] Python collection class uses copy remote block to local keepstore added

#5 - 09/26/2018 04:15 PM - Peter Amstutz

- Description updated

#6 - 09/26/2018 05:17 PM - Peter Amstutz

- Subject changed from [keepstore] copy remote block to local keepstore to [keepstore] copy block from remote keepstore to local keepstore

#7 - 09/26/2018 05:32 PM - Peter Amstutz

- Description updated

#8 - 09/26/2018 05:40 PM - Peter Amstutz

- Description updated

#9 - 09/26/2018 05:42 PM - Peter Amstutz

- Description updated

#10 - 09/26/2018 05:45 PM - Peter Amstutz

- Description updated

#11 - 09/26/2018 05:59 PM - Tom Clegg

- Description updated

#12 - 09/26/2018 06:00 PM - Peter Amstutz

- Story points set to 2.0

#13 - 10/03/2018 03:28 PM - Ward Vandewege

- Target version changed from To Be Groomed to 2018-10-17 sprint

#14 - 10/03/2018 03:28 PM - Tom Clegg

- Assigned To set to Tom Clegg

#16 - 10/03/2018 07:56 PM - Tom Clegg

- Status changed from New to In Progress

#17 - 10/04/2018 02:29 PM - Tom Clegg

14199-copy-from-remote @ [fde27ce0e46521db9828c228e7fb531e003724a8https://ci.curoverse.com/view/Developer/job/developer-run-tests/918/](https://ci.curoverse.com/view/Developer/job/developer-run-tests/918/)

All keepstore GET and HEAD responses for block requests should support this new header/response

I haven't done this part: so far, the X-Keep-Signature header is ignored in GET requests that aren't proxied with the +R remote signature. IIRC this requirement was supposed to help long-running clients keep their signatures alive longer than blob-signature-ttl without creating a collection. It turns out not to be trivial to implement, and I'm now doubting [a] updating the block's timestamp on the storage backend is a reasonable way to do that, especially now that auto-expiring temp collections are supported, and [b] (even if so) it deserves to be rolled into this pull-from-remote story.

Responses with an X-Keep-Locator header must also include a "Vary: X-Keep-Signature" header.

I think the correct requirement is "Responses that can vary depending on the X-Keep-Signature request header," i.e., requests with +R signatures and no +A signatures in the path, regardless of whether the X-Keep-Signature/X-Keep-Locator headers appear. For example, if a client does "GET without x-k-s", then "GET with x-k-s", the first response has to include the Vary header to prevent a proxy from reusing it for the second request. Implemented accordingly.

#18 - 10/04/2018 06:15 PM - Tom Clegg

14199-copy-from-remote @ [005951a5e62a55894eace6e8fb3fe91d5c4ba84c](https://ci.curoverse.com/view/Developer/job/developer-run-tests/918/)

Fixes occasional test suite deadlock/timeout. From commit message:

```
If a KeepClient has been created using an API server address that is no longer reachable, calling keepclient.RefreshServiceDiscovery() puts
```

the poll() goroutine into an endless retry loop, and a second call never returns because ent.clear is never ready to receive.

Work around this in the pull worker tests by only refreshing services from the API server actually being used, not on additional ones referenced by previous test cases.

#19 - 10/05/2018 03:59 PM - Peter Amstutz

- It checks for the presence of X-Keep-Signature but doesn't check the value. Maybe there should be some validation, even if we don't do anything with it?

```
func (rrc *remoteResponseCacher) Flush(ctx context.Context) {
```

- This doesn't match the http.Flusher interface. Should it? Or should it be renamed? (maybe Finish())

Rest LGTM.

#20 - 10/09/2018 03:39 PM - Tom Clegg

14199-copy-from-remote @ [77ee37c567ba73a0c33455ecf8d5c8200cf69d72](https://github.com/77ee37c567ba73a0c33455ecf8d5c8200cf69d72)

- invoke local cache/signature on X-Keep-Signature: local[, ...]
- rename Flush to Close (its semantics are much, um, closer to io.Closer than http.Flusher so this seems less confusing)

#21 - 10/09/2018 07:51 PM - Peter Amstutz

Tom Clegg wrote:

14199-copy-from-remote @ [77ee37c567ba73a0c33455ecf8d5c8200cf69d72](https://github.com/77ee37c567ba73a0c33455ecf8d5c8200cf69d72)

- invoke local cache/signature on X-Keep-Signature: local[, ...]
- rename Flush to Close (its semantics are much, um, closer to io.Closer than http.Flusher so this seems less confusing)

LGTM.

#22 - 10/10/2018 03:36 AM - Tom Clegg

- Status changed from In Progress to Resolved

#23 - 11/01/2018 08:30 PM - Tom Clegg

- Related to Feature #14406: [SDK] Go collection uses copy remote block to local keepstore added

#24 - 11/13/2018 08:51 PM - Tom Morris

- Release set to 14