Arvados - Bug #7120

[Keep] keepproxy should log real IP address

08/24/2015 04:23 PM - Tom Clegg

Status:	New	**Start date:**	08/24/2015
Priority:	Normal	**Due date:**	
Assigned To:		**% Done:**	0%
Category:	Keep	**Estimated time:**	0.00 hour
Target version:	Arvados Future Sprints		

Description

Currently, if keepproxy sees an X-Real-IP header, it logs that value (and the X-Forwarded-For value if provided) instead of the actual remote IP address.

This is weird in several ways:

- Our documented proxy config uses nginx's X-Forwarded-For feature, which always includes the information given in X-Real-IP -- so X-Real-IP is redundant -- except that keepproxy doesn't log X-Forwarded-For unless X-Real-IP is also given. (And in that case it logs both, so the X-Real-IP value is given twice in the log message.)
- If we put multiple nginx proxies in front of one keepproxy, we can't tell which of those nginx proxies made a given request.
- If keepproxy is deployed *without* a proxy, or behind a different proxy that doesn't ensure X-Real-IP is replaced by a trusted value, a client can trivially cause keepproxy to log arbitrary value instead of the real remote IP. If a different upstream proxy is in use, that proxy will probably log the real IP -- but if no proxy is being used at all, the real IP will not be logged anywhere. Even though we currently recommend installing keepproxy behind a TLS proxy which *does* overwrite any X-Real-IP provided by the client, it's easy to imagine someone getting it wrong (e.g., add a header instead of replacing it) when porting the config to another kind of proxy, or thinking "I'll just point clients directly to keepproxy because I don't need to add TLS in front of it".
- Resulting log format makes logs harder to parse when X-Forwarded-For != X-Real-IP.

We should just fix this logging bug rather than letting it elevate our specific nginx config from a recommendation to a security requirement.

The fix seems to be trivial -- just log the provided X-Forwarded-For value, with the real remote IP appended, just like nginx does:

```
if xff := req.Header.Get("X-Forwarded-For"); xff != "" {
  return xff + "," + req.RemoteAddr
} else {
  return req.RemoteAddr
}
```

This will give us "192.168.1.23,1.2.3.4,10.0.0.1", for example, if the client 192.168.1.23 uses its proxy at 1.2.3.4 to connect to our nginx proxy, and our nginx proxy connects to keepproxy from 10.0.0.1. When we read the log files, we can see the trail of proxies and make a reasonable decision about which proxies (if any) we should trust to tell us the "real" IP originating the request.

Once this is done, we can remove the redundant X-Real-IP field from the proxy configuration in the docs.

While we're at it, we should make another easy logging fix: use "%q" instead of "%s" when logging client-provided strings, to ensure incoming newlines, spaces, and quotation marks don't make our log files unparseable.