Metacat - Bug #3367

Harvester stores passwords in clear text

06/05/2008 01:18 PM - Chad Berkley

Status:	New	**Start date:**	06/05/2008
Priority:	Immediate	**Due date:**	
Assignee:	Duane Costa	**% Done:**	0%
Category:	harvester	**Estimated time:**	0.00 hour
Target version:	Unspecified	**Spent time:**	0.00 hour
Bugzilla-Id:	3367		

Description

The harvester stores the user's password in clear text in the database. Passwords need to be stored as md5s or use some other secure form of encryption.

History

#1 - 07/25/2008 01:30 PM - Matt Jones

It would be best if we didn't store this password at all. Reassigning to Duane to determine why we are storing redundant password in the database.

#2 - 07/28/2008 09:07 AM - Duane Costa

Matt,

Harvester stores LDAP DN/password information for each site so that it can run an automated login (via the Metacat client) prior to harvesting documents for a site.

From your comment it sounds like you might have been thinking that Harvester stores the metacat database username and password, but this is not the case.

I agree that the LDAP passwords stored in the database should be encoded rather than in clear text.

Thanks,
Duane

#3 - 07/29/2008 09:44 AM - Mark Servilla

I believe the need for the user passwords in the database is for Harvester to interact with Metacat on the user's behalf - that is, to upload the EML document on behalf of the user without an interactive login session by the user. I agree that user passwords should not be stored in clear text. MD5s, however, are hash values and cannot be used in this instance because they cannot be converted back to the original password for subsequent interaction with Metacat. I would suggest that the Harvester support the optional storage of encrypted passwords through the use of a local private key and a preferred encryption algorithm. Two use cases would be affected: 1) the user setup of a harvest account would result in the initial encryption of the password and 2) the harvest process would dynamically decrypt each password for the duration of the session for use with Metacat. In this case, only the local private key would have to be protected. Just my two cents...

Sincerely,
Mark

#4 - 07/29/2008 10:39 AM - Matt Jones

I thought this was the case. We discussed using a private key to encrypt the password, as we're trying to tighten up these issues in other places in metacat. However, it seems to me that we don't really need either in this case. The current process is to:

1) authenticate user upon harvest registration
2) store user password if valid
3) periodically harvest documents and insert, update, delete as that user

However, the key here is that metacat is simply keeping track of who it has previously authenticated, and gains no new information by storing the password. Instead, it seems to me that Metacat could simply store the account name and a 'validated' flag, then do the insertions as that user without further logging in. This would solve two potential problems. First, the password list is vulnerable to perusal if the db were to be compromised. Second, if the user changes their password in LDAP, the harvest will fail. Both of these would be solved by eliminating the stored password in favor of a simple 'authenticated' flag.

I suspect the reason this is done is that MetacatClient is used to do much of the document insertion, etc. So the system may need to be architected somewhat differently, such as by using an IPC-based version of MetacatClient that wouldn't communicate over http. Such a client has been planned since we first created the MetacatClient, but has never been implemented.

#5 - 03/27/2013 02:23 PM - Redmine Admin

Original Bugzilla ID was 3367