EPJ Web of Conferences **214**, 04003 (2019) https://doi.org/10.1051/epjconf/201921404003
CHEP 2018

Dataclient: a simple interface for scientific data transfers hiding x.509 complexities

Vincenzo Ciaschini[1,*], *Lucia* Morganti[1,**], *Matteo* Tenti[1,***], and *Carmelo* Pellegrino[1,****]

[1]INFN CNAF, V.le Berti Pichat 6/2, 40127 Bologna, Italy

Abstract. Since the current data infrastructure of the HEP experiments is based on GridFTP, most computing centres have adapted and based their own access to the data on the X.509. This is an issue for smaller experiments that do not have the resources to train their researchers in the complexities of X.509 certificates and that would prefer an approach based on username/password.
On the other hand, asking computing centres to support different access strategies is not so straightforward, as this would require a significant investment of effort and manpower.
At CNAF-INFN Tier1 we tackled this problem by creating a layer on top of the gridftp client/server, that completely hides the X.509 infrastructure under an authentication/authorization process based on the Kerberos realm of our centre, and therefore based on username/password. We called this Dataclient.
In this article we will describe both the principles that drove its design and its general architecture, together with the measures undertaken to simplify the user experience and maintenance burden.

1 Introduction

Many users have difficulties in correctly managing personal user certificates and the whole X.509 infrastructure necessary to properly use the storage infrastructure at CNAF-INFN Tier1, where remote access is primarily based on GridFTP.

Dataclient is a tool that intends to address this difficulty by completely hiding it from the user, presenting him with a username/password based interface, while handling all of the X.509 complexities (certificates, proxies, VOMS[1]) itself in a way that is invisible to the user.

2 Operation

Dataclient operation is quite straightforward. The tool does not require any kind of installation, and can be used by simply copying it in a home directory. When first invoked, it sets up the whole infrastructure (X.509 CA certificates, Globus, downloading the latest version from the official repositories) and then logs the user in with his INFN-CNAF credentials. After a

*e-mail: vincenzo.ciaschini@cnaf.infn.it
**e-mail: lucia.morganti@cnaf.infn.it
***e-mail: matteo.tenti@cnaf.infn.it
****e-mail: carmelo.pellegrino@cnaf.infn.it

successful login, a proxy is created based on the user experiment, and from that moment on Dataclient can be used as a simple drop-in replacement for globus-url-copy, supporting the same options, without the user ever seeing a certificate or being required to reauthenticate unless:

- the user login password at CNAF has changed (note that CNAF policies require at least one password change each year) or,
- the user membership has been revoked.

In case the password has changed, the next time the command is run, the user will be asked to reauthenticate, otherwise he will be informed of his membership revocation and he will have to contact CNAF directly in order to have it reinstated.

3 Architecture

Dataclient has two main components: a client that will be run by user and a server that is at CNAF.

The server is contacted by the client in a TLS-protected session during the initial setup to download the locations of the CA/Globus distribution, and once more during user authentication.

In the latter case, it receives the user's username/password data and checks them with Linux Pluggable Authentication Modules (PAM)[2] against CNAF kerberos (please see Security Considerations below to read how the risk is handled).

PAM is the standard framework that Linux computers use to authenticate users when they login and for many other uses. Using it to authenticate the user's username/password data makes dataclient completely able to rely on whatever mechanism a site already has in place to authenticate users, wether that be local user, Kerberos, LDAP, or anything else.

If the user authenticates successfully, a proxy is created containing a full set of VOMS credentials for the user experiment, and sent back to the user. This proxy will be used transparently during all file transfer operations. This means that from CNAF point of view, its GridFTP infrastructure can simply be configured to handle one more VO, just like any other VO already handled.

The server is contacted two additional times: immediately before a file transfer to receive a list of files that will be transferred, and after a successful transfer to mark these files as successfully received. This allows it both to act as a catalog and to be able to warn the user in case of an unsuccessful transfer.

The transfer of all files is verified as successful by using globus-url-copy MD5 hash checking.

The client is also failure resistant in the sense that if it dies for any reason, this will be detected the next time the client is run and the user will be warned. In this case, the transfer operation that was interrupted will be marked as failed.

4 Adopters

Dataclient has been successfully operated by users belonging to several experiments. Among them Cuore, Cupid, and KM3NeT, that have shown to have special requirements for scientific data transfers. Indeed, Dataclient greatly improves and simplifies automation of continuous file transfers by substantially removing the need for user interaction.

In particular, these experiments need to continuously copy scientific data from their acquisition sites, LNGS for Cuore and Cupid and both Capopassero and Toulon for KM3NeT, to the permanent storage located at CNAF.

Dataclient has also been proposed for adoption to the Borexino, Limadou, and CosmoWNext collaborations, who are currently testing it.

5 Security considerations

Transmitting a username/password pair over the net has undeniable security implications, and therefore Dataclient takes special steps to protect them.

The username/password couple is protected in transit not only by the TLS protocol, but also by being additionally ciphered before the transfer with an additional public key, whose corresponding private key is held by the server.

The server itself takes care of never saving or logging the password anywhere.

6 Thanks

The authors would like to thank Enrico Fattibene for his help.

References

[1] R. Alfieri, R. Cecchini, V. Ciaschini, L. dell'Agnello, A. Frohner, K. Lorentey ,F. Spataro, *From gridmap-file to VOMS: managing authorization in a Grid Environment, Future Generation Computer Systems*, Vol. 21 issue 4, Pages 549-558 (2005).

[2] *Linux-PAM* http://www.linux-pam.org