Sicurezza dei dati in EGRID

Riccardo Murri

riccardo.murri@ictp.trieste.it

The Abdus Salam ICTP

Cosa intendiamo per *sicurezza dei dati?*

Sicurezza dei dati è la possibilità di decidere *chi* legge *quali* dati e chi li scrive, e di attuare le proprie decisioni nel sistema di griglia.

Esempi

- I dati di una ricerca non ancora conclusa potrebbero essere *privati*: si darà l'accesso al pubblico solo a ricerca conclusa.
- I dati acquistati con i fondi di un gruppo potrebbero non essere accessibili ai membri di un altro gruppo.

Classificazione dei dati

Fonti —

- di largo utilizzo;
- sottoposti a vincoli d'accesso secondo il contratto di acquisto;
- solo chi ha i media di distribuzione può fare l'upload

Dati personali —

- il proprietario decide tutti i permessi di accesso
- possibilmente *segreti*

Dati di progetto —

- condivisi da un gruppo di ricercatori
- lettura/scrittura viene decisa da chi "guida" il gruppo
- l'appartenenza al gruppo è dinamica

I requisiti di sicurezza

- I diversi gruppi di ricerca hanno accesso a diversi file di dati, a seconda dei contratti che hanno sottoscritto – permessi separati di lettura
- All'interno di uno stesso gruppo, solo alcune persone hanno diritto di scrittura sui file delle fonti dei dati – permessi separati di scrittura all'interno dello stesso gruppo.
- Ciascuno deve avere uno spazio personale dove poter leggere e scrivere i proprî dati senza restrizioni.
- Devono esistere directory di *condivisione* dove poter tenere dati comuni a più ricercatori – formazione più flessibile e dinamica di gruppi di condivisione.

La struttura delle directory

/fonti —
- contiene i dati acquistati sotto contratto
- per ogni contratto:
  - una directory
  - un gruppo di "uploaders"
  - un gruppo di "downloaders"

/utenti —
- contiene i dati *personali*
- una directory '`private/`' per i dati leggibili solo all'utente proprietario
- fuori di quella, il contenuto è *leggibile* a *tutti*

/progetti —
- contiene i dati *di progetto*
- chi crea la directory può decidere chi fa parte del gruppo che vi può leggere e scrivere

La struttura degli SE

Struttura a stella:

- Storage Element *centrale* a Padova:
  - accesso a *tutti* gli utenti EGRID;
  - utenti e gruppi di accesso gestiti da EGRID Trieste;
  - gerarchia di directory stabilita
- Storage Element *periferici* a Firenze, Roma, Palermo:
  - accesso solo agli utenti locali;
  - utenti e gruppi di accesso gestiti dal sistemista;
  - gerarchia di directory dinamica

Sui nodi periferici, i sistemisti decidono le politiche di accesso e protezione dei dati!

Tardivo sommario

- Autenticazione ed autorizzazione nel middleware EDG
- Com’è implementata la sicurezza dei dati sullo SE *centrale*
- Gli strumenti per replicare la struttura dati sugli SE periferici

I termini

Autenticazione Accertare l'identità di chi richiede una certa risorsa

Autorizzazione Concedere il permesso di utilizzare una risorsa

In EDG/Globus GT2 si utilizzano:

Autenticazione certifi cati X.509 / infrastruttura PKI

Autorizzazione permessi UNIX

Tre passi da autenticazione ad autorizzazione

- Chi necessita di accesso alla griglia attiva un certifi cato proxy con cui si *identifica* nell'accesso alle funzioni di griglia. – *Autenticazione*
- All'interno di uno stesso computing site, un certifi cato proxy viene mappato in un account UNIX, secondo politiche *locali* (grid-map file)
- L'accesso ad un file specifi co avviene secondo le normali regole del filesystem UNIX – *Autorizzazione*

Autorizzazioni UNIX

Tre gruppi di autorizzazioni:

- *utente proprietario*,
- *gruppo proprietario*,
- *altri*

Tipi di permessi:

	file	**directory**
r	lettura	lista dei contenuti
w	scrittura del contenuto	creazione e cancellazione di file
x	esecuzione	attraversamento
s		i file sono creati con lo stesso gruppo della directory
t		solo il proprietario può cancellare un file

Account e gruppi

- Ogni utente di EGRID ha il *suo proprio account*
- Ogni account di un utente ha un suo *gruppo privato*
- Il gruppo `egridusr` raccoglie tutti gli account di utenti EGRID
- *Due* gruppi per ogni contratto/fonte: gruppo degli amministratori dei dati e gruppo dei fruitori dei dati
- Gruppi defi niti dagli utenti

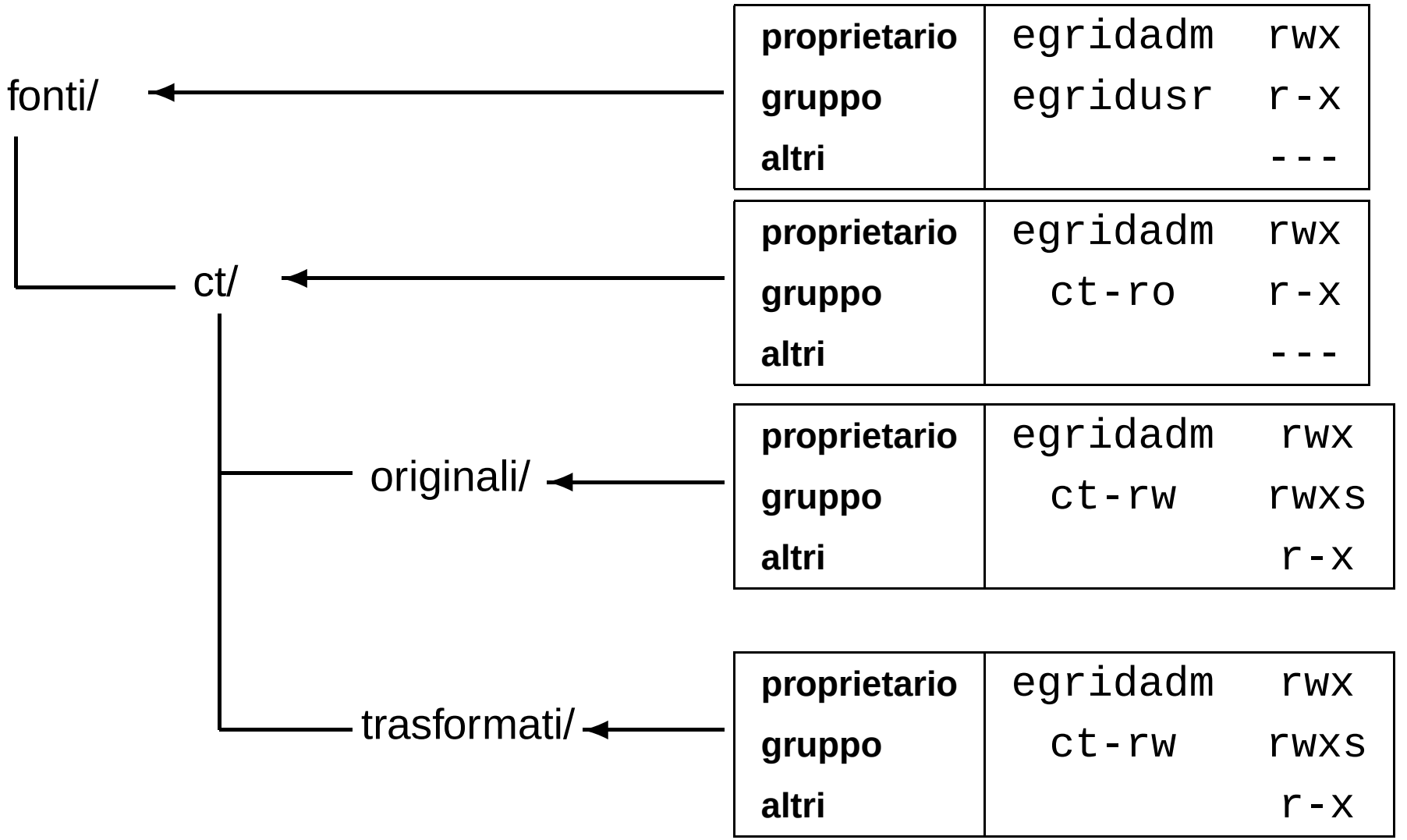
`/fonti` — gruppi

Per ogni fonte/contratto `ct`, esistono:

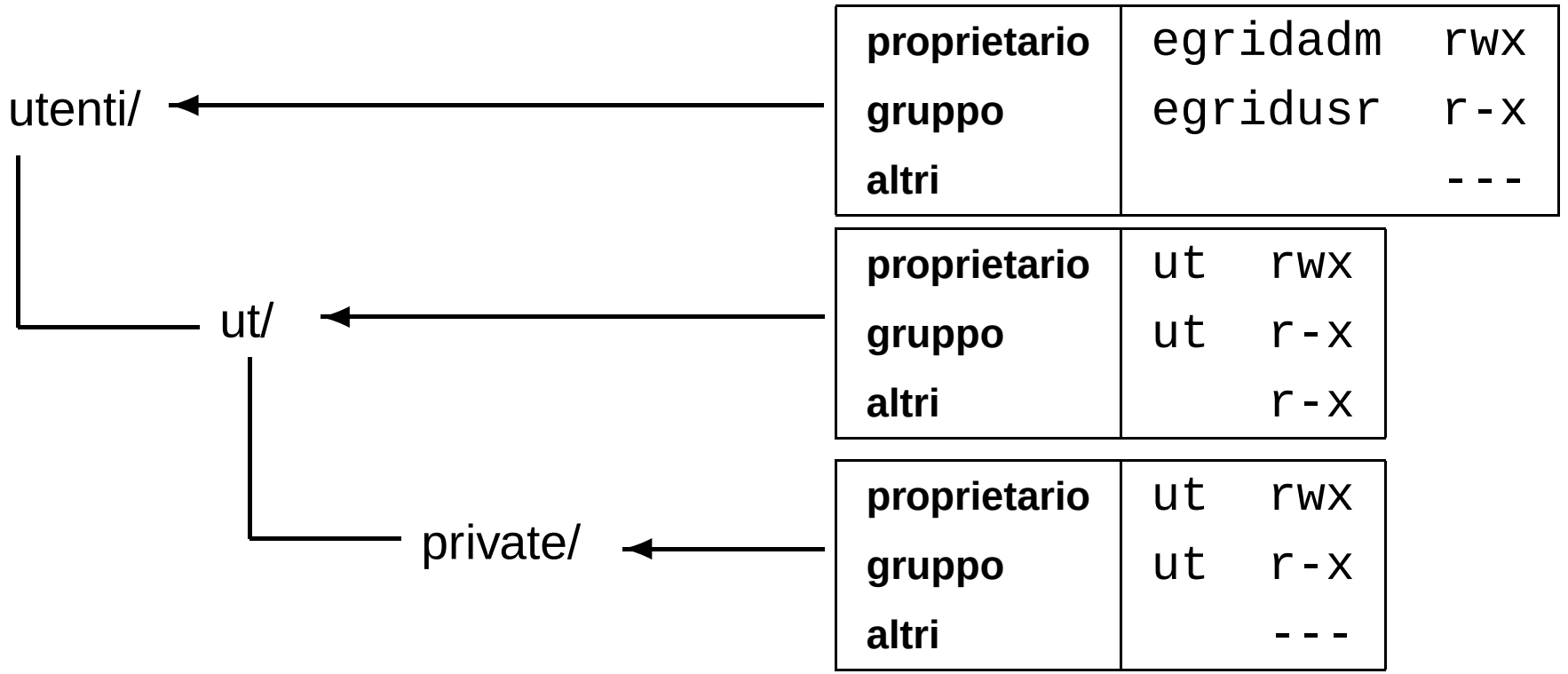
- una directory `/fonti/ct/`
- un gruppo `ct-ro` di *fruitori* – utenti che possono *solo leggere* i dati di `ct`
- un gruppo `ct-rw` di *amministratori dei dati* – utenti che possono *leggere e scrivere* i dati di `ct`

Il gruppo `ct-rw` è un *sottoinsieme* di `ct-ro`.

`/fonti` — permessi



`/utenti` — permessi



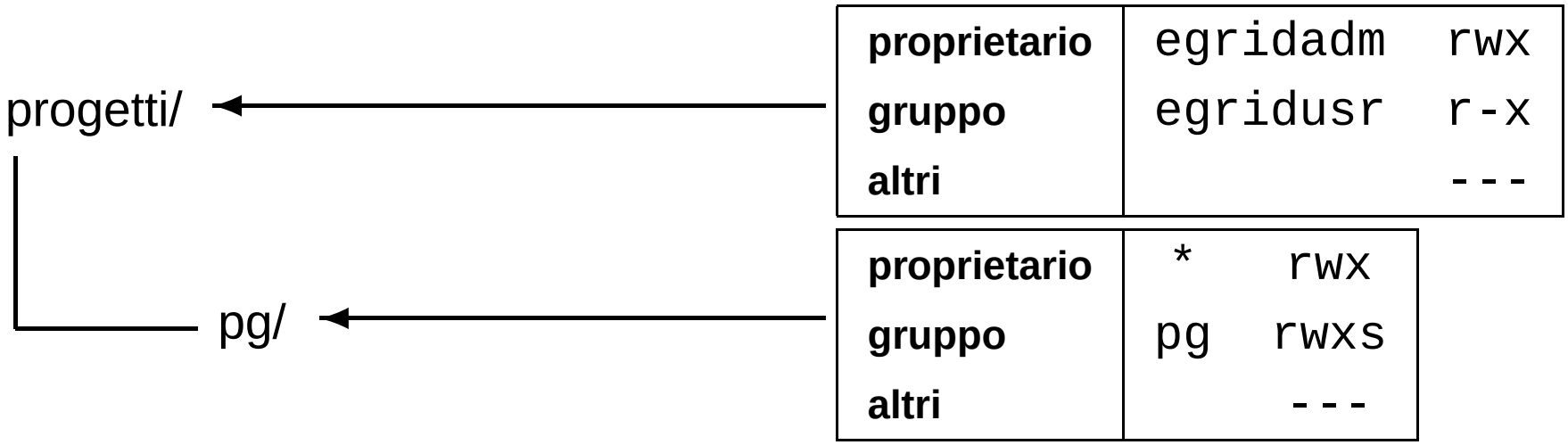
`/progetti` — creazione e gruppi

Le sottodirectory di `/progetti` sono create *su richiesta*, insieme ad un gruppo che le può *leggere e scrivere*.
Per creare la directory del progetto `pg`:

1. creare il gruppo `pg` col comando `egrid-groupm`
2. creare la sottodirectory `pg` con il comando `egrid-groupdir`

La directory ed il gruppo *devono* avere lo stesso nome!

/progetti — permessi



I permessi su `pg` e le sue sottodirectory si possono cambiare con il comando `egrid-chmod`.

Gestione dei gruppi

I gruppi utente si gestiscono con `egrid-groupm`:

- Ogni gruppo ha *uno ed un solo* proprietario (solitamente, chi ha creato il gruppo).
- Il proprietario è l'unico autorizzato ad aggiungere o rimuovere altri utenti dal gruppo.
- Non si possono aggiungere altri utenti al proprio gruppo privato.

I gruppi di fonti/contratti sono creati da EGRID Trieste, che ne assegna la proprietà all'amministratore del contratto.

Il problema

L'amministratore dello SE periferico decide:

- quali utenti di griglia sono autorizzati ad usare il nodo, tramite il grid-mapfile
- utenti e gruppi locali
- organizzazione della storage area

Idea: replicare la gerarchia delle directory, con gli stessi permessi e gruppi, ma modifi care *localmente* il contenuto dei gruppi.

Replica dei gruppi

- server LDAP distribuisce i gruppi da Trieste
- `egrid-ldap2users` permette di trasformare le mappe LDAP nel formato di `/etc/group`
- `egrid-groupchg` permette di operare sui gruppi, aggiungendo e togliendo utenti
- Rimane al sistemista la libertà di decidere chi sia membro dei gruppi, e di mantenere gruppi locali in `/etc/group`

Replica delle directory

- Modello client/server
- `egrid-print-dirhier` server su SE centrale
- `egrid-copy-dirhier` client su SE periferico:
  - attivato da riga di comando o periodicamente come cron job
  - replica gerarchia delle directory, permessi, proprietà di gruppo e di account
  - se un gruppo o un account non esistono in locale, usa un gruppo o account predefi nito (confi gurabile)
  - permette di escludere parte della gerarchia (p.es. le home degli utenti non locali)