10.05.2006

Comments on the LCG JSPG User-level Accounting Data Policy

Holger Marten
Forschungszentrum Karlsruhe

Abstract: During the LCG GDB on April 5, 2006, the LCG Joint Security Policy Group (JSPG) provided a draft document for the User-level Accounting Data Policy at http://agenda.cern.ch/askArchive.php?base=agenda&categ=a057704&id=a057704s1t7%2Ftransparencies%2FKelsey5Apr06.pdf and GDB members were asked for feedback.

This document provides feedback from Forschungszentrum Karlsruhe after an internal discussion with the data privacy commissioner of the Centre. It is neither a statement "of Germany" in the sense of a "representation of Germany in the GDB" nor a collection of feedback from the German centres participating in LCG, EGEE and other grid projects. However, since German and European laws for data privacy and protection apply to all German / European centres we believe that most of the statements in this document should be acceptable by and applicable to other centres as well.

We would like to thank the JSPG for the first draft on user-level accounting data policy. The document shows that the data privacy issues within a world wide grid context are understood and taken seriously. The idea to encrypt the user information (DN) of the accounting records, where only a small group of experts owns the key for decryption, is especially acknowledged. We can give a general "go ahead" for the further development of the final policy document.

Comments on the User-level Accounting Data Policy

1. **Two types of accounting data.** The current policy for user-level accounting as described by the JSPG does not involve any *sensitive* user information (dates of birth, addresses, phone numbers etc.), thus allowing for more light weight data handling and protection. Still, it contains two different kinds of data sets that can or have to be treated in a different way:

    a. *Anonymous, statistical* accounting information of type "aggregated consumed CPU time per VO / site / month".

       This type of data is mostly uncritical and may be provided in a "world readable" way to scientific boards and communities. (In this document we always refer to LHC and other VOs dealing with purely unclassified, non-commercial and scientific applications).

    b. *User-related* accounting information.

       This type of data can potentially be used to record and control the work of individual persons, and allows conclusions about his/her working methods, results, performance etc. This **must** be prohibited.

2. **Grid AUP.** The grid AUP should (in the end) contain a link to the user-level accounting data policy such that each individual user is informed and accepts this policy.

3. **Accounting requirements.** The policy must contain a list of strong arguments for the necessity of collecting accounting data and the purpose thereof. We suggest to use keywords like: … necessary for the world wide allocation of technical (compute) resources, … guarantee technical security and availability, etc.

   This list of arguments should be (as) *complete* (as possible) and the policy should contain a statement that other, non-documented use of the collected data is prohibited.

The policy must also state which data are stored, for how long and who has access. – This is already taken into account in the JSPG draft.

4. **Access to the (user-level) data.** The current policy envisions to provide data access for two different entities: authorized GOC and VO managers.

   It must be ensured that these two groups of persons belong to states that accept the European laws for data privacy (i.e., that user-level data are only exchanged within such states). Currently, these are the 25 EU member states, the EEA member states and several other states (including Canada and Switzerland) that have been proven to provide an adequate level of privacy protection. Exchange of private data with organisations in the U.S. is documented in a special treaty called "safe harbour privacy principles", and the organisation to receive / work with private data should verify to accept these principles.

   The policy should contain a special paragraph for those people *handling* user-level data (currently the above mentioned authorized GOC and VO managers), that they sign the policy i.e. especially use the data exclusively for the described purposes and don't provide them to non-authorized persons (i.e., don't misuse them).

5. **Access to the (user-level) data by the user.** Any user has the right to access his/her own accounting records and the mechanism for this must be implemented and described in the policy – already taken into account in the JSPG draft document.

   The user should even have the right to get the data changed if he/she can justify / prove that the stored data are incorrect (this seems to be a rather academic case in a world wide grid context, but who knows…?)

6. **Routine usage of the data.** The policy should describe the routine usage of the data as completely as possible. It should especially describe the detailed procedures to use the data for resource allocation by the VO managers (as they are in a very strong position compared to the "normal" users or even to sites).

7. **Misuse of and suspicion on wrong accounting data.** The policy should describe a procedure to be followed in such cases. E.g.: is there a body or panel that arbitrates in case of conflicts?

Special answers to questions from the JSPG

Q (slide 3): Timely accounting (monitoring) useful?
A: Yes. VO managers or computing coordinators should get the possibility for resource planning (and allocation?). Insofar, a daily or at least weekly "report" seems to be a necessity.

Q (slide 6): accounting records transferred to GOC DB; how frequent?
A: We suggest once a day.

Q (slide 6): do we allow/forbid VOs to do their own accounting or bookkeeping?
A: The collection of private data as well as the compliance with data privacy regulations are serious tasks that have to be carefully documented and justified (see comments above). Processes for one central collection and storage are much easier to define and control. So, no "individual" (VO) activities should be accepted.

Q (slide 8): data aggregation per month or per week?
A: Not that important for us as a site. VO managers might be more interested in weekly statistics.