International Conference
AUTOMATICS AND INFORMATICS'11
3-7 October 2011, Sofia, Bulgaria

AN APPROACH TO SECURITY ANALYSIS OF DISTRIBUTED COMPUTING GRID ENVIRONMENT

V. Dimitrov, Z. Minchev, D. Todorov, H. Turlakov

Institute of Information and Communication Technologies, Bulgarian Academy of Sciences,
Sofia 1113, Acad Georgi Bonchev Str., Bl.25A, tel. +359 2 979 66 15, +359 2 979 66 31
E-mails: vgd@acad.bg, zlatogor@acad.bg, dttod@acad.bg, tour@acad.bg

Abstract: The paper describes an approach for analysis of a distributed system's components influence to its security as a whole. The approach is illustrated on a Grid structure base. A generic model of the Grid Middleware gLite technology and threats analysis obtained as a result of Entity-Relationship complex system machine interpretation in I-SCIP v 2.1 software environment combined with experts' knowledge are presented.

Key words: distributed systems, security, grid, cloud, complex systems modeling, threats analysis

INTRODUCTION

Nowadays systems security is getting more and more an emerging multidimensional field of research and development. The widespread usage of distributed computing environments couldn't reach its real level due to the presence of still opened and unsolved security problems. Most probably, this fact is a consequence of the high complexity of distributed systems and their great number of potential points for attacks. That gives to the intruder many potential options for malicious influence over the whole distributed system which in many cases is made additionally easier because of the well-known functionality of the existing internal connections.

The paper tries to present an approach for the analysis of the influence of every single component of a distributed system to its security as a whole. The influence level of every component relies on experts' assessment of existing functional connections and should be perceived as a start of discussion. The approach implementation is illustrated on a real Grid distributed environment which is realized, fully functional and in production for several years. Our future intentions are this approach to be implemented (after some amendments) for security analysis of other existing distributed systems including Cloud computing environments.

GRID MIDDLEWARE GLITE OVERVIEW

The gLite [1] is a middleware that makes a set of independent server clusters with a high speed network connections suitable for virtual use as a unified resource in a way widely known as a Grid Computing [2]. This middleware is currently deployed on several hundreds of server clusters known as Grid-sites. It enables a global e-science approach to a number of disciplines. There are tens of thousands of end-users and thousands of support and management staff using the Grid infrastructure with this middleware. The evolution of the Grid infrastructure and further development of gLite continues with EGI-InSPIRE FP7 project (Integrated Sustainable Pan-European Infrastructure for Researchers in Europe, 2010-2014) [3].

Analysing the potential threats in a Grid infrastructure running gLite middleware is important due to following facts:

- The environment provides high performance computing possibilities which could be used for harmful purposes;
- The environment provides huge storage space which could be used for illegal contents;
- Some Grid applications are related to critical infrastructures;
- Some databases stored on Grid contain sensitive or confidential information;

The gLite middleware has a Service Oriented Architecture (SOA). The services are integrated in the Grid infrastructure and their summarised structure is given on **Figure 1**.

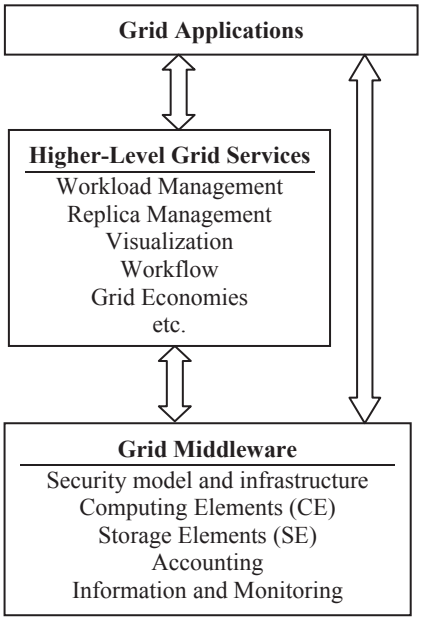

Figure 1. gLite middleware services main structure.

METHODOLOGY

The methodological approach that has been chosen for model creation is based on the utilization of the well-known Berta-

lanffy's General Systems Theory [4], and thus concerning the studied system building elements in-between nonlinear interactions. The reason for that is because the Grid like distributed system could be easily approximated with a complex dynamic system that consists of a lot of objects, time-dependent weighted relations sufficient enough for complicated behavior interpretation.

As far as the model interpretation is not unique, due to experts' opinion usage and different viewpoints, we consider initially to accomplish brainstorming combined with discussion and Delphi based questionnaire filling/filtering [5], [6]. The selected formalism for model machine interpretation is the Entity-Relationship (E-R) one [7], which allows intuitive work and implementation of causality. A more detailed description of the methodology is given in [6].

SOFTWARE IMPLEMENTATION

The software implementation is based on I-SCIP-SA v.2.0 environment [8]. Briefly, I-SCIP-SA allows creation of models using objects (interpreted as rectangles, squares and circles), which are connected with relations (interpreted as headed weighted arrows – uni- and bi- directional). The arrows' weights are marked as yellow labels over the arrows and are expressed in percentages from the interval [0, 1] using the following scale: low [0-30], middle [30-50] and high [50-100]. The building elements' interrelations weights' generalization produce a resulting Sensitivity Diagram (SD) that uses and extends the ideas of Vester's sensitivity model [9], allowing model building elements' zone classification and system sensitivity analysis as follows: Red zone (active elements, Influence/Dependence Maximum Ratio (IDMR) =100/50, SE (South-East) part of SD cube), Blue zone (passive elements IDMR=50/100, NW (North-West) part of SD cube), Yellow zone (critical elements, IDMR=100/100, NE (North-East) part of the SD cube) and Green zone (buffering elements, IDMR=50/50, SW (South-West) part of SD cube). Additionally, the 3D SD gives a possibility for direct sensitivity (z-coordinate, marked with red arrow in **Figure 3**) calculation of a given object from the system as an absolute difference between the influence (x-coordinate, marked with green arrow in **Figure 3**) and dependence (y-coordinate, marked with blue arrow in **Figure 3**) values, concerning a certain object from the system of interest. When this difference is negative the object in SD is classified as passive (producing a decreased system sensitivity in its SD zone) and is colored in light grey, otherwise it is active (producing an increased system sensitivity in its SD zone) and is colorized in white.

Finally, it should be noted that this classification could be both – static or dynamic depending on the available data – a number or an array of numbers provided from a statistical data base or experts in general or for a given time horizon. In the dynamic case the time steps number that is showing the discretes of the transition function are marked within blue labels above the arrows (for visual illustration see the model from Figure 2).

THE MODEL

The gLite Grid conceptual model E-R implementation is based on the information, presented in [10]. Additionally, a group of ten subject matter experts' brainstorming discussion and a questionnaire fill-up for relations weighting (using the following simplified scale: "strong", "middle", and "weak") have been also accomplished. These experts are selected among people with at least 5 years practical experience on different positions in the gLite Grid infrastructure – from end user to Grid site administrator, security officer and manager. The model has been developed in the following typical scenario context:

Typical scenario for using the Grid infrastructure – becoming a Grid user and simple execution of a Grid job.

1. The User applies to the regional Certification Authority (CA) for a Grid certificate.
2. CA signs and publishes the certificate.
3. The user applies for membership in an appropriate Virtual Organization (VO) with this Grid certificate.
4. The VO management approves the membership.
5. The User installs and configures a computer with User Interface (UI) software.
6. The user develops a Grid application (job).
7. The user authenticates and authorizes himself/herself using the valid Grid certificate and VO membership credentials.

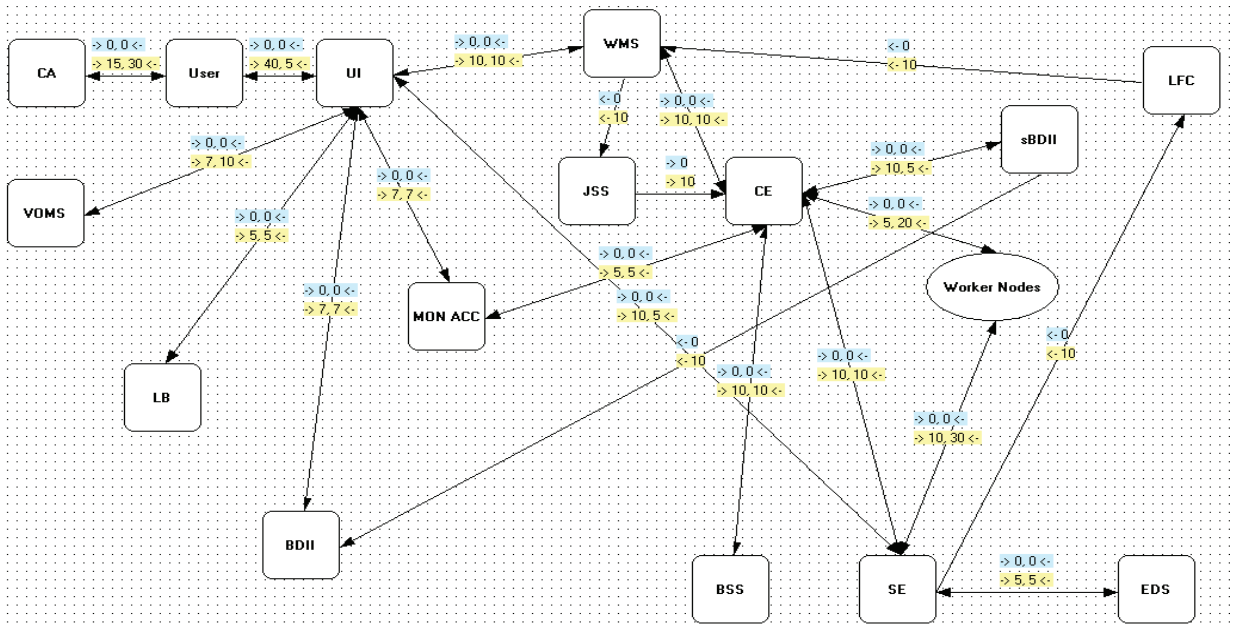

Figure 2. gLite Grid E-R model screen-shot from I-SCIP v.2.1.

8. The User queries Berkeley Database Information Index (BDII) nodes for Grid resources currently available.
9. The User submits the job to appropriate Workload Management System (WMS) node.
10. The WMS manages and monitors the Grid job and submits it to an appropriate Computing Element (CE) node of Grid site for execution.
11. The CE puts the job on a BSS queue.
12. The job is executed in turn on a Worker node(s) belonging to this Grid site.
13. The User monitors the job status using Logging and Bookkeeping (LB) capabilities during steps 10 to 12.
14. When the job is finished, the User retrieves the job output to UI. End of scenario.

In the model illustrated on **Figure 2** the particular gLite functional entities are presented as model objects and their designation and meanings are as follows:

CA – Certification Authority managed and coordinated by EUGridPMA [11]. Issues X.509 certificates for scientific users in order to access the Grid infrastructure.

VOMS – Virtual Organization Management System. Manages Virtual Organization (VO) of which this user is a member. Authentication and Authorization.

UI – User Interface. Provides all of the functionality for Grid access mainly in a form of command line tools.

BDII – Berkeley Database Information Index. The implementation of the Grid Information System. Provides comprehensive information for the Grid infrastructure all over the world.

sBDII – Site BDII. The information system which holds information for this Grid site only.

LFC – LCG File Catalog. Holds actual information for files and directories on the Grid storages.

LB – Logging and Book-keeping. Answers to the queries for the Grid job status in details. Various logging capabilities.

WMS – Workload Management System. A key entry point to high-end services available on a Grid. It provides reliable and efficient distribution and management of end-user requests for Grid tasks.

JSS – Job Submission Service. Submits Grid jobs to a Grid site.

Grid site – A complete set of computer nodes with different Grid roles.

CE – Computing Element. The main node of a Grid site. The Grid site may have more than one CE.

SE – Storage Element. Manages the storage resources assigned to particular Grid sites. The Grid site may have more than one SE.

EDS – Encrypted Data Storage. An additional storage service which provides data encryption functionality for that SE.

BSS – Batch and Scheduling System. Manages job queues and

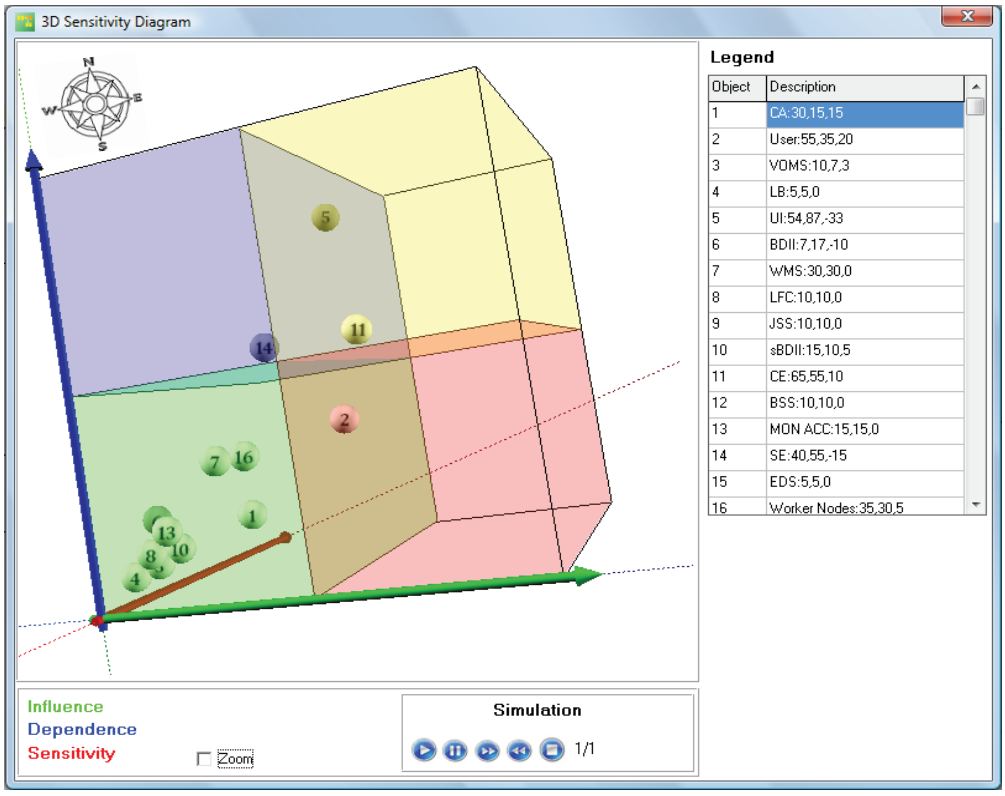

Figure 3. Resulting Sensitivity Diagram of the gLite middleware technology model.

priorities for several different VOs.

MON+ACC – Monitoring and accounting. Various monitoring tools for that Grid sites. Accounting for executed Grid jobs and their details.

Worker nodes – Many high performance computers. The actual job execution happens here by requests and supervision of the CE.

The gLite Grid E-R model is presented in Figure 2 and its objects correspond to nodes and services of gLite Grid environments. The relation between any two model objects represents the influence and dependence between relevant Grid nodes or services. One model object has influence on other model object if the first node or service has a possibility to determine the behaviour of the other relevant node or services. It is obvious that this influence or dependence is very hard to be formalised and assessed. In our opinion the only acceptable way this influence/dependence to be determined is by summarising experts' assessments.

This conceptual model should not be perceived as final comprehensive representation of all functional relations in Grid infrastructure based on gLite. It is quite probable that in this distributed environment there are some not well documented functional relations with a significant role in resulting system security. Our understanding is that the model should have a long lasting continuous development that would increase the reliability of the resulting Sensitivity Diagram.

The resulting Sensitivity Diagram (SD) is presented in and is producing the following classification of the accomplished gLite middleware elements: "passive" (entities with low additive influence compared to their additive dependence in the system, IDMR=50/100): "Storage Element" with negative sensitivity (the entities' sensitivity characteristic is related to the system sensitivity as a whole in every SD zone; the positive one means increased sensitivity in the considered SD zone, whilst the negative one - decreased), "active" (entities producing high additive influence compared to their additive dependence in the system; the opposing of the "passive" ones, IDMR=100/50): "User" with positive sensitivity, critical (entities that have both high additive influence and dependence in the system; most important for system sensitivity control, IDMR=100/100): "User Interface" with negative sensitivity, "Computing Element" with positive sensitivity. Therefore the "User Interface" should be considered as the most dangerous node because its compromising could produce maximal damages to many of the remaining nodes. For example if a "User Interface" node is conquered then the malicious person can gain control over many other Grid nodes. The rest of the model elements have been rated as buffering and are not going to be discussed in more details.

DISCUSSION

The paper has given a kind of vulnerability analysis in a real distributed computing Grid environment realized on gLite middleware. The results are based on a pilot group of experts' knowledge and experience processing and generalization. What was noticed during the present study is practically related to the problems of matching the experts' knowledge with I-SCIP software 3D SD, so a small experimental extension of I-SCIP v. 2.0 capabilities in v. 2.1 by means of solving of the reverse classification task in 3D SD recalculating the relations' weights in accordance with a certain element (object) desired location was performed. For that purpose we have accomplished the COTS IBM ILOG CPLEX v.12.3 [12] optimization environment library and quadratic optimization that is not a perfect solution for non-linear optimization but is fast enough and allows very good ad-hoc approximation of the desired location taking into account the work with experts' knowledge as an input with a lot of noise.
The accomplished initial results do not show the detailed nature of the threats but only outline a way how they could be obtained, using exerts' knowledge and the presented system based model approach supported with I-SCIP software. Finally, it should be noted that in our future plans the presented approach will be extended in the context of Cloud environment system security analysis.

ACKNOWLEDGEMENTS

This work was partially supported by the EU FP7 project SysSec – A European Network of Excellence in Managing Threats and Vulnerabilities in the Future Internet: Europe for the World under the agreement n^{o} 257007 and FP7 EGI-InSPIRE project. The authors also express their gratitude to the pilot group of experts for the fruitful discussions and the questionnaires filling-up.

REFERENCES

[1] gLite - Lightweight Middleware for Grid Computing, http://glite.cern.ch
[2] Foster, I., Kesselman, C. The Grid 2, Second Edition: Blueprint for a New Computing Infrastructure (The Elsevier Series in Grid Computing), Morgan Kaufmann, Second Edition, 2003.
[3] EGI-InSPIRE project (Integrated Sustainable Pan-European Infrastructure for Researchers in Europe), http://www.egi.eu/projects/egi-inspire
[4] Bertalanffy, L. General System Theory: Foudation, Development, Applications. New York, 1968.
[5] Nguyen, M. and Dunn M. Some Methods for Scenario Analysis in Defence Strategic Planning, Australian DoD, Joint Operations Division, Defence Science and Technology Organisation, DSTO-TR-2242, 2009.
[6] Minchev, Z. and Shalamanov, V. Scenario Generation and Assessment Framework Solution in Support of the Comprehensive Approach, International Symposium "Analytical Support to Defence Transformation", RTO-MP-SAS-081, Sofia, Boyana, April 26 – 28, 22-1 – 22-16, 2010.
[7] Chen, P. The Entity-Relationship Model-Toward a Unified View of Data, ACM Transactions on Database Systems, 1, no.1, 9-36, 1976.
[8] Minchev, Z. and Petkova, M. Information Processes and Threats in Social Networks: A Case Study, In Conjoint Scientific Seminar "Modelling and Control of Information Processes", Organized by: College of Telecommunications, Institute of ICT - Bulgarian Academy of Sciences and Institute of Mathematics and Informatics – Bulgarian Academy of Sciences, Sofia, Bulgaria, November, 85-93, 2010.
[9] Vester, F. The Art of Interconnected Thinking, Report to the Club of Rome, May, 2002.
[10] S. Burke, Campana S., Lanciotti E., Litmaath M., Lorenzo P. M., Miccio V., Nater C., Santinelli R., Sciaba A. gLite 3.2 User Guide, CERN, 2011, Available at: https://edms.cern.ch/file/722398//gLite-3-UserGuide.pdf
[11] European Policy Management Authority for Grid Authentication, http://www.eugridpma.org
[12] IBM ILOG CPLEX Optimizer, http://www-01.ibm.com/software/integration/optimization/cplex-optimizer