Analysis and Research of Grid Security Based on Knowledge Base*

XIAO Hai-dong, LI Jian-hua, JIANG Xing-hao, CHEN Li-ya
(Institute of Information Security Technology, Shanghai Jiaotong University, Shanghai 200030, China)

Abstract: Knowledge grid is very popular today. A unique method based on knowledge mining is set up to solve the problem on security of grid node. It is very useful to construct the grid security architecture. This architecture is used to provide grid-based security knowledge mining services and the embranchment boundary deciding principle is mentioned to describe the details.

Key words: grid knowledge; security strategy; grid security
CLC number: TP393.08 **Document code**: A **Article ID**: 0529-6579 (2004) S2-0135-04

1 Introduction

Before utilizing a Grid Computing service, all Grid users should ask themselves a number of important questions: Can my computation's data be modified without my knowing? Can someone else impersonate me on a target machine? How do I know that my computation is truly being executed on the resources that I think they it is being executed on? Developers of Grid Computing services should ask a similar collection of questions, both with regard to their new service using other Grid services and also the security of the Grid Service that they are providing[1].

First of all, General concepts and requirements of Grid security should be given as integrity, confidentiality, authentication, authorization, and fundamentals of PKI such as digital signatures, public/private keys, symmetric keys, etc., and then, let's have a overview of the current state of the art in security in Grid computing (X.509, certificate management, SSL, Kerberos, GAA, GSSAPI, GSI, VPNs, CAS, certificate-based authorization, Web Services Security, etc.) and open issues. Special emphasis will be placed on topics relevant to OGSA Security. While the general topic of computer security is too broad to be covered in this tutorial, the goal is to enable Grid users, Grid system administrators, and Grid service developers to operate more securely in Grid environments.

2 Knowledge mining for grid security

The Grid by its nature involves access to computer systems and data outside one's own company or institution. Security is therefore a major element in any Grid infrastructure, as it is necessary to ensure that only authorized access is permitted[2]. The common Grid security model is as follows:

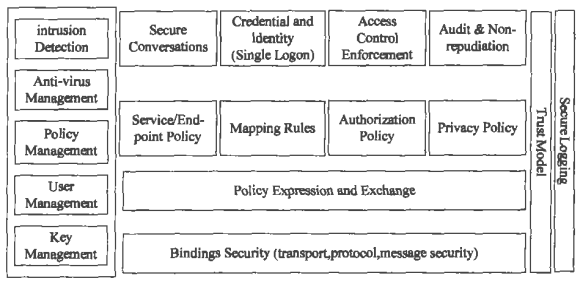

Fig. 1 Grid security model

And this common grid security model associated with the Globus toolkit is supported by a Grid Security Infrastructure (GSI) based on a Public Key Infrastructure where users authenticate to the grid using X509 certificates[3].

Here, the knowledge base is introduced to give a grid security model based on it. In order to assure the grid security, a knowledge base should be constructed to refine the data in the attacks database, as we know, when a common computing package arrived the grid node, some actors of it will be traced, from the common grid security model above, secure logging will trace every computing package, and the useful attacks information will be written to the security database. The purpose is to construct the grid security knowledge base as follows.

2.1 Construct the grid security knowledge base

The function of grid security knowledge base is to

* 收稿日期: 2004-09-10
作者简介: 萧海东 (1975 年生), 男, 工程师; E-mail: xiaohaidong@yahoo.com.cn

manage the knowledge orderly, the base function include store, add, delete, modify, query and doing the verifications of consistency and integrality. Some denotations of knowledge are in follows, one level predication logic, engendering rules, framework theory, etc. The knowledge in this paper will be expressed with engender rules[4]. The knowledge base and database have close relationship, but they are not the same. In the view of knowledge logic denotation, the relation database is a simple knowledge base, every relationship in the ER database is a atom formula, in other words, it is a predication, the meta group is the fact in the denotation of knowledge, so the grid security knowledge base can be constructed base on the ER database, extreme facility will be available in design of grid security knowledge base for the support of powerful functions of ER database.

At the present time, most knowledge construction architecture is divided into three levers include conception-fact-rule, as follows:

Knowledge=Concept + Fact + Rule

As usual, the conception is included into the fact, so the security knowledge in the grid security knowledge base is constructed by two part: the fact and the rule.

The rule is often described as follow modality:

〈rule〉: = (〈rule ID〉, 〈precondition1〉 | 〈precondition2〉 … 〈conclusion〉, 〈rule intension CF〉)

In the grid security research, because of both the precondition and the conclusion are fact, only two predication need to describe the grid security knowledge base model, they are rule and fact; at least two table should be created to construct security knowledge base in order to make use of the ER database. And the rule should be indexed while designing the rule table, the purpose is to make the rule matching smoother, surely the rule ID will be the index. And the fact table in the grid security knowledge base should be indexed too. Other fields mainly describe the fact illumination information. In this way, the grid security domain special knowledge stored in security knowledge base can be put forth by the accumulate relationship rule:

IF X Then (Y, Con) or X⇒(Y, Con)

means "if X come into existence then there will exist the conclusion which believe consideration equal Con". As described by descartes product formula: 〈X, Y, Con〉

2.2 Grid computing packages' security character selection analysis

Computational grids provide computing power by sharing resources across administrative domains. This sharing, coupled with the need to execute untrusted code from arbitrary users, so lots of security hazards occur, it sounds like a threat to the globe grid system, but in fact, it provide a good platform for training the grid security knowledge model too[5]. According to the embranchment boundary deciding principle, we analysis the grid security knowledge base model as follows:

In the example below, four dimensions tentatively weight are selected. "The embranchment boundary deciding principle" arithmetic is a typical superincumbent method in analysis of security character selection, which can avoid some unnecessary computing and get the best result.

Suppose that number i security character is denoted by X_i, If $n=4$, we have (X_1, X_2, X_3, X_4). After choosing to lower dimensions, we could get $d=2$. The search tree with branch demarcations algorithm to carry on characteristic choosing is showed in fig 2.

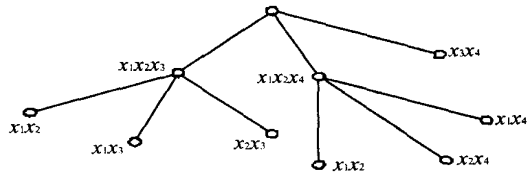

Fig 2 The research tree to carry on characteristic choosing with branch demarcation research

Suppose that at present the computation packet division of security characteristic is in i grade already. This indicates that the abandoned characteristics to successor nodes in $k=0, 1, 2, \cdots, i-1$ grades have been figure out. The abandoned characteristic for q_i successor nodes to the present i grade is expressed as $Q_i=(x_1^{i+1}, x_2^{i+1}, \cdots, x_{qi}^{i+1})$ X_i. X_i as the characteristic group after lower dimension, φ_i as the characteristic group that the i successor node could be abandoned, r_i is the radix for this characteristic group. The algorithm is as follows:

(1) Figure out q_i, $q_i=r_i-(D-d-i-1)$

(2) From,

$$J(X_i-x_1^{i+1}) \leqslant J(X_i-x_2^{i+1}) \leqslant \cdots \leqslant J(X_i-x_{qi}^{i+1}) \leqslant \cdots \leqslant J(X_i-x_{ri}^{i+1}),\ x_i^{i+1} \in \varphi_i$$

figure out

$$Q_i=(x_1^{i+1}, x_2^{i+1}, \cdots, x_{qi}^{i+1})$$

(3) Remove Q_i from φ_i, and amend r_i, that is: $\varphi_{i+1}=\varphi_i-Q_i$; $r_{i+1}=r_i-q_i$;

(4) Compare: if $J(X_i-x_{qi}^{i+1})<B$, then $l=q_i$, and turn to step 5;

If $q_i=0$, turn to step 6;

Under other conditions, calculate $X_{i+1}=X_i-x_{q_i}^{i+1}$, if $i+1=D-d$, turn to step 7, else turn back to step 1;

(5) $\varphi_{i+1}=\varphi_{i+1}+x_{q_i}^{i+1}$, $r_{i+1}=r_{i+1}+1$ at the same time, order, $l=l-1$, $q_i=q_i-1$. If $l=0$, turn to step 4, else loop in this step;

(6) $\varphi_i = \varphi_{i+1}$, $r_i = r_{i+1}$, $i = i-1$. If $i = -1$, then stop; else, $X_i = X_{i+1} - x_{qi}^{i+1}$, order $l = 1$, turn back to step 5;

(7) Change $B = J(\overline{X_{n-d}})$, make $\overline{X_{n-d}}$ as the most effective characteristic group, $l = q_i$, then turn to step 5.

3 Eliminate and classify the computation packets which have security threatens according to computation packets' security knowledge base

The central issue of high lever grid security of cluster server is the performance of grid system. As the service object of grid node, the attribute of the computation packet itself has sizable influence on the performance of the node. When the security threatens computation packets that have reached the node need a large number of calculation resource or store resources, and do harm to the grid node, we should eliminate and classify these invalid computation packets[6]. In another situation, the mistake of procedure participating in operation or hostile code with great threat to the grid will exhaust node resources in a large amount. The prerequisite to resolve problem is to classify computation packets according to the categorized characteristic, and eliminate those packet beyond the limit of categorized characteristic from the computations queue. This has important meanings for automatically examining and automatically correcting error procedure in designing a security grid system.

Now we use Bayes minimum risk decision-making model to realize the classification of the characteristic of the computation packet.

Consider that the characteristic of computation packet is d dimensions; structure a random vector with d dimensions

$$x = [x_1, x_2, \cdots, x_d]^{\mathrm{T}} \tag{1}$$

$x_1, x_2, \cdots, x_d$ are one-dimension random variable. The state spaces Ω to classify characteristics of computation packets are c classes:

$$\Omega = \{\omega_1, \omega_2, \cdots, \omega_c\} \tag{2}$$

The decision space is made up of α_i, $i = 1, 2, \cdots, \alpha$

$$A = \{\alpha_1, \alpha_2, \cdots, \alpha_\alpha\} \tag{3}$$

Here the reason that a is different from c is that to the c class permit other decision besides c kinds of different decisions, such as adopting "refuse" decision, $a = c + 1$, the loss is function $\lambda(\alpha_i, \omega_j)$; $i = 1, 2, \cdots, c$. λ show the loss that when the real state is ω_j but the decision to take is α_i.

The principle of Bayes minimum risk decision-making model is to have the minimum loss when considering the loss that misjudge brought[7]. If we made condition risk minimum before every decision and action, then the expectation risk is minimum while we making decisions to all x. The decision rule is:

If $R(\alpha_k \mid x) = \min\limits_{i=1,\cdots,\alpha} R(\alpha_i \mid x)$, then $\alpha = \alpha_k$. We could use this algorithm to realize the classification for the security characteristics of grid computation packets:

Under the condition of have known $P(\omega_j)$, $p(x \mid \omega_j)$, $j = 1, \cdots, c$ and x as the computation packets in the node that want to be discerned, we could Bayes equation to figure out the posterior probability:

$$P(\omega_j \mid x) = \frac{p(x \mid \omega_j) P(\omega_j)}{\sum_{i=1}^{c} p(x \mid \omega_i) P(\omega_i)} \tag{4}$$

$$j = 1, \cdots, c$$

With the posterior probability and decision table, use this equation to figure out the condition risks $R(\alpha_i \mid x)$ under the condition α_i, $i = 1, 2, \cdots, \alpha$

$$R(\alpha_i \mid x) = \sum_{i=1}^{c} \lambda(\alpha_i \mid \omega_j) P(\omega_j \mid x)$$

$$i = 1, 2, \cdots, \alpha \tag{5}$$

$$R(\alpha_k \mid x) = \min_{i=1,\cdots,\alpha} R(\alpha_i \mid x) \tag{6}$$

Compare those a condition risks $R(\alpha_i \mid x)$, choose the decision α_k with the minimum condition risk, that is:

$$R(\alpha_k \mid x) = \min_{i=1,\cdots,\alpha} R(\alpha_i \mid x) \tag{7}$$

As to the invalid computation packets, because their security categorized characteristics are relatively clear, we can adopt convenient method to judge and eliminate them. Generally we can adopt Chernoff limit or Bhattacharyya coefficient to confirm upper bound of the error rate.

4 Conclusion

We have analyzed the grid security characteristic classification and classification algorithm for the computation packets in grid node. This information is very useful for us to design or optimize the grid security system. While deciding the knowledge parameters of computation packets in grid node, we can get the security characteristics dynamically, classify them and decide the most suitable security scheme to the node resource of the grid system.

The judgment and mark of invalid computation packets offers the important basis for whole grid system to collect rubbish and detect task stop. If a grid node that is attacked, or a threaten computation packet coming, these would influence the computations security characteristic

weight, so the characteristic of the computation packet changed. We can judge this by index it in the security knowledge base, if its character matches to the invalid computing package, we decide the computation packet is invalid, and then eliminate it at once.

References:

[1] BAJKUMAR B. High performance cluster computations architectures and systems [M]. Beijing: Electronic Industry Press, 2001: 527—529.

[2] MIKE S, COLIN U. Third International Conference on Peer—to—Peer Computing (P2P'03) September 01 — 03, 2003 Link? ping, Sweden Grid Security: Lessons for Peer—to—Peer Systems IT Innovation Centre.

[3] MOORE P C, JOHNSON W R, RICHARD J. Detry adapting globus and kerberos for a secure ASCI Grid [EB/OL]. http://csdl. computer. org/comp/proceedings/sc/2001/1990/00/19900054abs. htm

[4] MO Z Y, LI X M. Parallel multigrid computations for anisorropic diffusion problems [J]. Numerical Computation And Computer Application, 1997, 9 (3): 218—229.

[5] BIAN Z Q, ZHANG X G. Pattern recognition [M]. Beijing: Tsing-hua Univisity Press, 2000: 177—179.

[6] http://www. ivdgl. org/testsite/griphyn/documents.

[7] MIRON LIVNY. The Grid Console [EB/OL]. http://www. cs. wisc. edu/condor/bypass/examples/grid—console/.

基于知识库的网格安全分析与研究

萧海东, 李建华, 蒋兴浩, 陈丽亚

(上海交通大学信息安全工程学院, 上海 200030)

摘 要: 知识网格如今研究的越来越普遍, 基于知识挖掘提出了一种独特的解决网格节点安全性的方法。这对于建立网格安全体系结构非常有用。这种结构提供了一种网格安全的基础知识挖掘服务, 并利用分支定界法则做了仔细的讨论。

关键词: 网格知识; 安全策略; 网格安全

中图分类号: TP393.08