异构分布式系统可信调度模型关键技术研究

（文献综述报告）

2009 级博士生 唐小勇

调度模型及调度算法的研究一直是分布式系统研究的热点问题之一. 一般并行应用都采用任务优先图模型(即有向无环图DAG)，其中节点代表任务，有向弧或边代表具有相互依赖关系的通信，如任务优先约束关系. 经典的基于DAG调度算法是启发式表调度算法，其基本思想是首先计算DAG任务优先权，然后依优先权降序排列，最后依次调度各任务. 然而经典表调度算法[1-5]主要研究同构计算系统，不适合异构计算系统. 近年来提出一些改进的针对异构计算系统表调度算法，如启发式映射(MH)算法[6]，动态级调度(DLS)算法[7]，最小分级时间片(LMT)算法[8]，关键路径(CPOP)算法和异构环境下最早完成时间(HEFT)算法[9,10]. 文献[9,10]证明HEFT算法无论是平均调度长度，还是加速比都明显优于DLS, MH, LMT和CPOP算法. 但上述算法优化目标仅局限于调度长度、加速比等经典调度性能，而没有考虑到异构分布式系统的可靠性、安全性等系统可信性性能需求.

近十年来，系统可信性研究受到了国际上广泛的重视. 在美国，DARPA、NSF、NASA、NSA、NIST、FAA、FDA 和其他DoD机构都积极参与关于高可信软件和系统的研究开发. 为了实现系统可信性的目标，人们从20世纪70年代之后就在做着不懈的努力. 包括从应用程序层面，从操作系统层面，从硬件层面来提出的TCB相当多. 最为实用的是以硬件平台为基础的可信计算平台（Trustec Computing Platform），它包括安全协处理器、密码加速器、个人令牌、软件狗、可信平台模块（Trusted Platform Modules，TPM）以及增强型CPU、安全设备和多功能设备. 这些实例的目标是实现：数据的真实性、数据的机密性、数据保护以及代码的真实性、代码的机密性和代码的保护.

国内的可信计算领域，主要有：武汉瑞达公司研制的安全计算机于2004年10月通过国家密码管理委员会主持的技术鉴定，成为国内第一款自主研制的可信计算平台. 2005年联想集团的“恒智”芯片和可信计算机相继研制成功. 同年，兆日公司的TPM芯片研制成功. 这些产品都相继通过国家密码管理委员会的鉴定和认可. 此外，同方、方正、浪潮、天融信等公司也都加入了可信计算的行列. 武汉大学、中科院软件所等高校和研究机构也都开展了可信计算的研究.

国内外在可信性研究方面已经开展了多年的工作，但是由于传统的可信机制主要关心软件行为的安全性和可依赖性问题[11]，且主要面向封闭或同一管理域内展开，面向开放环境、跨组织和管理域的动态性异构分布式可信研究尚处于起步阶段. 封闭环境中，用户的授权由权威源集中管理，信任的作用较为隐蔽；在开放环境中，主体间的信任关系对授权的作用逐步凸现出来，合理利用信任能够有效简化复杂的授权管理任务[12]. 在这种背景下, Blaze 等人为了强调信任在授权中的特殊作用，于1996 年提出了信任管理(trust management)的概念，并将信任管理定义为：“一种描述和解释安全策略、安全凭证以及用于直接授权关键安全

操作的信任关系的统一方法”[13]. 信任管理系统把授权决策转换为一种满足性验证(proof of compliance)问题：“凭证集C 是否能证明请求R 满足本地策略P”. 信任管理的初期研究主要关注授权语言与CCA 算法，如PolicyMaker[14]，KeyNote[15]， SPKI/SDSI[16]和DL[17]等，其中PolicyMaker 和KeyNote 是Blaze 等人先后提出的两个信任管理语言， 与SPKI 类似，以能力传播的方式表达权威委派策略. DL 是基于逻辑程序的信任管理语言，主要关注语言表达能力和计算复杂性. 近期出现的几个信任管理系统将角色引入到策略语言中，如RT[18] 和Cassandra[19]等. RT 在SDSI 的基础上提出了ABAC(attribute based access control)的思想，并考虑了多级访问请求过程中的委派问题. Cassandra 提出了更为通用的信任管理语言, 试图支持所有典型安全策略. 目前国内也相继开展了信任管理方面的研究，Tang将信任度评估模型集成到信任管理系统中[20]，增加了系统的动态适应性，Hong 等人基于RT 研究了ABAC 策略的委派深度约束问题[21].

在基于信誉的机制方面，1997年，Abdul-Rahman 等学者从信任的概念出发，提出了一个信任评估的数学模型，并给出了一种虚拟社区中基于信誉的信任解决方案[22]. 此后, Napster, eBay 等网上交易系统都采用了信誉机制来提供交易参考. 随着新的网络应用模式的兴起，围绕P2P 环境下不同的信誉模型和激励模式人们进行了众多的探索，其中比较有代表性的如基于抱怨的信任管理方法、PeerTrust、基于Bayesian Network 的信任模型、P2PRep 和基于PageRank 思想的EigenRep[23]等. 在基于信誉的机制研究中，围绕建立分布式的信誉管理机制[24]，以及基于模型实现对于恶意和不良行为的识别与防范正在成为新的研究热点[25].

国内在这方面也做了深入的研究，清华大学的林闯教授等[26] 利用随机模型分析了网络安全中的可信性，在可信网络方面取得了很大的进展；北京大学的唐文等[27] 运用模糊集合理论对信任管理问题进行了建模，提出了信任关系的推导规则，为开放网络环境中的信任管理研究提供了一个有价值的新思路. 同济大学的袁禄来[28,29]针对网格环境下一个网格节点推荐另一个网格节点所提供的推荐证据是不完备、不精确、不完全可靠的，导致很多信任评估方法的结果误差很大，利用D-S理论进行信任研究并提出基于信任的调度算法(TDLS). 文献[30-33]从不同的角度对可信系统进行了有效的研究，满足了一定的实际需求.

从社会学角度看，信任关系是最复杂的社会关系之一，是一个很难度量的抽象的心理认知，当实体之间的信任关系不能明确定义的时候，它也是不稳定的，给它的管理和评估带来了困难. 信任也是与上下文相关的一个动态过程，随着时间的变化，实体之间的行为上下文可能会动态地变化，并且具有时间滞后性的特点，也就是说，新的信任关系的评估依赖于时间和行为上下文. 因而，信任研究一个非常复杂的课题，虽然许多研究者从不同分布式应用角度应用各种方法进行信任定量计算，取得了一些有效的应用，但这些模型大多没有综合考虑各种可能的输入因子，例如：大多数模型没有风险机制，没有考虑服务者的声誉，不能很好地消除恶意推荐对信任评估的影响，没有解决初始信任值如何获得，准确定量评价信任精度不高等问题. 本课题将从心理学和经济学中的信任研究入手，进一

步研究信任关系，尤其是动态信任关系的相关性质、信任的表述和信任值的计算方法. 本课题拟用概率论中的Bayesian方法表示不确定性的信任关系，受经济学品牌形象理论[34]的启发，采用多人合作与非合作微分对策技术建立信任值计算二价偏微分方程，实现信任的动态表示，以获得信任值的精确计算公式.

上述基于信任的可信研究都是从系统安全性的角度出发，而没有充分考虑分布式系统底层需求. 最近几年，从系统底层任务在处理机上的的执行行为进行动态安全保证越来越受到广大研究人员的重视. S.Song最先提出一种可信的网格节点遗传调度算法(STGA)[35]，她把节点分为安全、危险和f可能安全三类，然后提出相应调度方法. Lin和Yang采用整数线性编程技术和启发式搜索方案相结合的实时安全保证调度算法[36]. Dogan提出一种高效的多QoS静态异构分布式系统调度算法(QSMTS_IP) [37]，其中包括安全性QoS. T. Xie 和X. Qin研究并提出了多个基于分布式系统的安全调度算法[38,39]，但其工作主要集中在调度时安全开销分析机制. 以上研究工作都是从任务调度角度进行系统安全性的有益探讨，但由于研究工作都是基于全互连网络，且假定节点和任务安全级，这与实际的任意网络结构异构分布式系统不符，且不能适应计算资源和任务安全需求动态变化. 另一方面，他们的研究都假定任务具有独立性，而没有研究任务的相互约束性. 针对上述问题，本课题将以动态的信任管理和信任值计算为基础，建立任务执行行为的动态安全性评估机制，从安全性角度研究可信调度模型和调度算法，提高异构分布式系统的可信性.

可信研究的另一个重要内容是可靠性，软件容错技术是提高分布式系统可靠性的一种有效途径，也是目前的研究热点，其中活动/备用技术[40]是一种重要的软件容错模型. 在活动/备用计算模型中，每个任务都有主、从两个版本，从版本是主版本任务的一个副本. 同一时间只有主版本的任务运行，当主版本运行出现故障时，从版本任务接替主版本任务的工作继续执行. 为了保证系统的容错特性，主任务与从任务不能分配到同一个节点机上. 许多学者对分布式系统中具有主/从多个版本的任务调度问题做了大量的研究，K.Ahn、J.Kim等提出一种延迟被动复制调度方法来实现可靠性[41]，文献[31]提出了在分布式实时系统中同时调度具有容错需求与无容错需求进程的混合调度算法，文献[42]引入可靠性代价概念对异构系统中的可靠性进行了评估并提出了最大化系统可靠性的调度算法. 然而，以上研究都假设分布式系统是由单个处理机构成的同构全互连系统，实际上，由于集群、网络技术的发展，异构系统容错调度已经成为研究热点. 同构系统和异构系统调度的最大区别是：在同构系统中，同一任务在所有处理机上的运行时间及处理机发生故障的频率相同；而异构系统则不一定相同，这主要是由于异构系统中不同的软、硬件配置所致. 文献[43]基于可靠性代价提出了一种针对异构分布式系统的容错调度算法，但其假设任务之间是相互独立. 而实际应用程序中，大部分任务之间是相互约束的. 文献[44]采用泊松概率分布理论分析处理机的可靠性，并提出针对具有相互约束限制任务的容错调度算法，但该算法假设异构系统连接网络是全互连且具有高可靠性，这与实际的异构分布式计算系统不

符，因为异构系统本质上具有网络任意连接性和异构性，所以系统网络的不可靠性在评价系统可靠性方面不能忽略. 其次，文献[45]假设处理机故障时只局限于同时一台处理机故障，这与实际系统存在多理机故障不符. 本课题试图从系统观点出发研究处理机和通信网络的可靠性，采用Poisson随机过程理论建立任意处理机网络异构分布式系统可靠性模型，最后提出依据任务执行行为失败率的任务容错技术和最优可靠通信路经查找算法为一体的任务调度算法，从而提高系统可信性.

综上所述：本课题的研究目标是提出一种基于行为的异构分布式计算系统可信任务调度模型，该模型包含系统可靠性分析、信任管理机制、安全开销计算、并行应用任务可信评估机制、考虑可信的任务调度器等模块. 可靠性分析主要研究多理机失败、网络通信故障与任务执行行为的函数关系，实现任意处理机网络最优可靠通信路经的动态查找. 信任管理机制拟采用概率论中的 Bayesian 方法表示不确定性的信任关系，并用多人合作与非合作微分对策技术建立信任值计算二价偏微分方程，实现信任的动态表示，从而获得信任值的精确计算公式. 安全开销计算将以信任值的精确计算为基础，依据用户的安全需求和计算节点能向用户提供的信任级，实现安全开销的精确表示. 以任务在系统上执行行为的可靠性和任务的安全开销为基础，实现任务可信性的动态评估. 最后，针对具有优先约束关系的并行任务，提出基于任务执行行为的考虑可信性的可信任务调度模型和可信任务调度算法，从而有效提高异构分布式系统性能.

参考文献:

1 M.K. Dhodhi, I. Ahmad, A. Yatama,et al. An integrated technique for task matching and scheduling onto distributed heterogeneous computing system. J. Parallel Distrib. Comput., 2002, 62 (9):1338–1361

2 A. Radulescu, A.J.C. van Gemund. Low-cost task scheduling for distributed-memory machines. IEEE Trans. Parallel Distrib. Systems, 2002, 13 (6):648–658

3 H.J. Park, B.K. Kim. An optimal scheduling algorithm for minimizing the computing period of cyclic synchronous tasks on multiprocessors. J. Systems Software, 2001, 56 (3):213–229

4 TANG XiaoYong, LI KenLi & D Padua. Communication Contention in APN List Scheduling Algorithm. Science in China Series F: Information Sciences, 2009, 52(1):59-69

5 Sinnen Oliver, Sousa, Leonel. List scheduling: extension for contention awareness and evaluation of node priorities for heterogeneous cluster architectures. Parallel Computing, 2004, 30(1):81-101

6 H. El-Rewini, T.G. Lewis. Scheduling parallel program tasks onto arbitrary target machines. J. Parallel Distrib. Comput., 1990, 9 (2):138–153

7 G.C. Sih, E.A. Lee. A compile-time scheduling heuristic for interconnection-constrained heterogeneous machine architectures. IEEE Trans. Parallel Distrib. Systems, 1993, 4 (2):175–187

8 M. Iverson, F. Ozuner, G. Follen. Parallelizing existing applications in a distributed heterogeneous environment. in: Proceedings of Heterogeneous Computing Workshop, 1995, 93–100

9 Liu,G.Q, Poh,K.L, Xie,M. Iterative list scheduling for heterogeneous computing. Journal of Parallel and Distributed Computing, 2005, 65(5):654-665
10 H. Topcuoglu, S. Hariri, M.-Y. Wu. Performance-effective and low complexity task scheduling for heterogeneous computing. IEEE Trans. Parallel Distrib. Systems, 2002,13 (3):260–274
11 Yu Bin, Munindar SirIsh P. An evidential model of distributed reputation management. In: Proceedings of First International Joint Conference on Autonomous Entities and Multi-Entity Systems, 2002, pp.294 -301
12 Aberer K, Despotovic Z. Managing trust in a peer-2-peer information system. In: Proceedings of the l0th International Conference on Information and Knowledge Management , New York, 2001
13 Blaze M, Feigenbaum J, Strauss M. Compliance checking in the policymaker trust management system. In: Proceedings of the Financial Cryptography'98. Berun: Springer-Verlag, 1998, pp. 254-274
14 Blaze M, Feigenbaum J, Lacy J. Decentralized trust management. In: IEEE Symp Secur Privacy, 1996, pp.164-173
15 Blaze M, Feigenbaum J, Ioannidis J, et al. RFC 2704: The KeyNote Trust Management System Version2, 1999
16 Ellison C M, Frantz B, Lampson B, et al. SPKI Certificate Theory. RFC, 1999, 2693
17 Li N H. Delegation Logic: A Logic-based Approach to Distributed Authorization. PhD thesis. New York: New York University, 2000
18 Li N H, John C M, Winsborough W H. Design of a role-based trust management framework. In: Proceedings of IEEE Symposium on Security and Privacy. Wiley: IEEE Comput Soc Press, 2002, pp.114-130
19 Moritz Y B, Peter S. Cassandra: Flexible trust management, applied to electronic health records. In: Proceedings of the 17th IEEE Computer Security Foundations Workshop (CSFW'04)
20 Tang Zhuo, Lu Zhengding, Li Kai.Time-Based Dynamic Trust Model Using Ant Conony Algorithm, Wuhan University Journal of Natural Sciences, 2006, 11(6):1462-1466
21 Hong F, Zhu X, Wang S B. Delegation depth control in trust-management system. In: Proceedings of the 19th International Conference on Advanced Information Networking and Applications (AINA'05). Taipei: IEEE Computer Society, 2005, pp.411-414
22 Abdul-Rahman A, Hailes S. Supporting trust in virtual communities. In: Proceedings of 33rd Hawaii International Conference on System Sciences, Maui, 2000
23 Kamvar S D, Schlosser M T, Garcia-Molina H. The eigentrust algorithm for reputation management in P2P Networks. In: Proceedings of WWW, May 2003
24 Christin N, Andreas S, Weigend, et al. Content availability, pollution and poisoning in file sharing peer-to-peer networks. In: Proceedings of EC05, HongKong, 2005, pp.68-77
25 Khopkar T, Li X, Resnick P. Self-selection, slipping, salvaging, slacking, and stoning: The impacts of negative feedback at eBay. In: Proceedings of EC05, HongKong, 2005, pp.223-231
26 林闯，汪洋，李泉林．网络安全的随机模型方法与评价技术．计算机学报，2005, 28(12) :1943-1956

27 唐文, 陈钟. 基于模糊集合理论的主观信任管理模型研究. 软件学报, 2003, 14 (8) :1401-1408
28 袁禄来, 曾国荪, 王伟. 基于Dempster-Shafer证据理论的信任评估模型. 武汉大学学报 (理学版), 2006, 52(5):627～630
29 袁禄来, 曾国荪, 姜黎立等. 网格环境下基于信任模型的动态级调度. 计算机学报, 2006, 29(7):1217-1224
30 朱峻茂, 杨寿保, 樊建平, 陈明宇. Grid 与P2P 混合计算环境下基于推荐证据推理的信任模型. 计算机研究与发展, 2005, 42 (5):797～803
31 秦啸, 庞丽萍, 韩宗芬, 李胜利. 分布式实时系统的容错调度算法. 计算机学报, 2000, 23(10):1056-1063
32 王功明, 关永, 赵春江, 吴华瑞. 可信网络框架及研究. 计算机工程与设计, 2007, 28(5):1016-1019
33 李小勇, 桂小林. 大规模分布式环境下动态信任模型研究. 软件学报, 2007,18(6):1510−1521
34 Anca E. Cretu, Roderick J. Brodie. The influence of brand image and company reputation where manufacturers market to small firms: A customer value perspective. Industrial Marketing Management, 2007, 36(2):230-240
35 S. Song, Y.-K. Kwok, K. Hwang. Trusted Job Scheduling in Open Computational Grids: Security-Driven Heuristics and A Fast Genetic Algorithms. Proceedings of the International Symposium on Parallel and Distributed Processing, Colorado, USA ,2005
36 M. Lin, L.T. Yang. Schedulability driven security optimization in real time systems, in: The 1st International Conference on Availability, Reliability and Security, 2006,314-320
37 Dogan, F. Ozguner. LDBS: a duplication based scheduling algorithm for heterogeneous computing systems. Proceedings of the International Conference on Parallel Processing, B.C., Canada, 2002, 352-359
38 T. Xie, X. Qin. Scheduling security-critical real-time applications on clusters. IEEE Trans. Comput, 2006, 55 (7):864 – 879
39 T. Xie, X. Qin. Performance evaluation of a new scheduling algorithm for distributed systems with security heterogeneity. Journal of Parallel and Distributed Computing, 2007, 67:1067-1081
40 刘东, 张春元, 李瑞, 黄影, 李毅. 软件容错模型中的容错实时调度算法. 计算机研究与发展, 2007, 44(9):1495-1500
41 K.Ahn, J.Kim, S.Hong. Fault-tolerant real-time scheduling using passive replicas. In:Proc. Pacific Rim Int. Symposium on Fault-Tolerant Systems, December, 1997, pp.15-16
42 Shatz S.M., Wang J.P., Goto M.. Task allocation for maximizing reliability of distributed computer systems. IEEE Transactions on Computer, 1992, 41(9): 1156-1168
43 Dogan, F. Ozguner. Matching and scheduling algorithms for minimizing execution time and failure probability of applications in heterogeneous computing. IEEE Transactions on Parallel and Distributed Systems, 2002, 13(3):308-323
44 Xiao Qin, Hong Jiang. A novel fault-tolerant scheduling algorithm for precedence constrained tasks in real-time heterogeneous systems. Parallel Computing, 2006, 32:331-356
45 Atakan Dogan. Matching and Scheduling Algorithms for Minimizing Execution Time and Failure Probability of Application in Heterogeneous Computing. IEEE Transactions on Parallel and Distributed Systems, 2002, 13(3):308-323