

CSCW 对分布式系统的继承和发展

陈立挺 鲁东明 潘云鹤

(浙江大学人工智能研究所 杭州 310027)

摘要 The relationship between CSCW and general-purpose distributed systems is described in three respects in this paper. There exists difference between CSCW and general-purpose DS. To a large extent, CSCW inherits concepts and techniques from DS, and however it possesses particular requirements and techniques which general-purpose DS lacks.

关键词 Computer supported cooperative work, Distributed systems, Coordination, Cooperative multiuser interfaces

“计算机支持的协同工作(Computer Supported Cooperative Work, 简称 CSCW)”这个词是 Grief 和 Cashman 在1986年提出来的,代表那些支持用户群(User Groups,在不引起歧义的情况下,本文对群和组不加区别)共同工作的计算机系统^[1]。在传统的群体工作方式中,人与人之间以及组织与组织之间的合作和协同是一个自然特性。在基于计算机的工作环境中,它仍然是一个固有特性。早期支持群的应用只是被动地感受到这一点,在计算机性能迅速提高与高速通讯网络发展的条件下,协同工作逐渐变成多用户系统的强烈要求。桌面会议系统集中体现了这种要求,它使得多个组通过共享窗口进行交互,实时地讨论会议主题。

事实上, CSCW 不是一个短时期内形成的概念,它与分布式系统有着直接的联系。在 CSCW 与分布式系统的关系上,存在的两种极端是:一是认为 CSCW 是分布式系统中普通的一类应用;另一极端则视 CSCW 为一类全新的系统。前者不利于清晰地反映 CSCW 不同于一般分布式系统的特性和要求,后者不利于概念和技术上对分布式系统的继承。

分布式系统是一个由许多处理单元和多个存储设备通过网络联接起来的统一的计算机系统。计算机支持的协同工作指人们在计算机的帮助下就同一个产品、同一个研究课题、同一个会议主题或同一个学术问题进行共同工作。这两个描述并不能很精确地反映出 CSCW 与分布式系统的关系。但是可以看到分布式系统强调机器之间的相互联系,而 CSCW 着重人通过计算机怎样协同或合作。

CSCW 与分布式系统的关系可以从三个方面来看。首先,从网络计算的层次上看,一般分布式系统

和 CSCW 是网络计算的两个分支,它们在系统目标和主要特征等方面都有很大的差别;其次, CSCW 系统和分布式系统具有相似的结点网络分布结构,这个共同的基础使 CSCW 在很大程度上能继承分布式系统的概念和技术,如基本的体系结构、通信、命名和安全、事务以及多副本等,但在要求上又和分布式系统有很大的不同,特别在某些具体的技术上,如协调控制、一致性和并发控制等,这种概念上的继承和要求上的差异尤其明显;再次, CSCW 中的有些要求及其技术在一般分布式系统中是没有的,也就是说,这些要求是 CSCW 特有的,如协同多用户界面以及多媒体通讯等,这些方面是对一般分布式系统的发展。

本文将着手从上述三个方面描述 CSCW 与分布式系统的关系。对于协调控制以及并发控制和一致性问题,将着重描述它们在 CSCW 中的特性和特有的要求,它们在 CSCW 和一般分布式系统中的共同问题也将简略提及。协同多媒体界面和多媒体通讯单独分两节描述。

一、CSCW 和一般分布式系统的差异

从系统整体,我们总结出 CSCW 系统和一般分布式系统在下面四方面的差异:

(1) 系统目标。分布处理通常有技术和组织两方面的目标,前者是分布式系统发展的主要推动力,具体而言,是要改善可靠性和坚定性,缩短响应时间,降低成本和提高性能^[2]。CSCW 系统的目标在于为一组用户的协同与合作提供支持,通过使用户能感知协同的更加直观的交互方式来模拟人在传统群体工作环境中的共同工作。分布式系统的技术目标仍

是 CSCW 考虑的因素,但不是最直接的目标。

(2)主要特征。分布式系统的主要特征是容错和并行性。这两个特征保证分布式系统获得比集中式系统高的可靠性和计算性能。CSCW 系统则以支持协同和用户的地域分散性为主要特征。

(3)透明性。分布式系统采用面向系统的途径来控制计算机的协同(这里协同指分布式系统中计算机如何紧密地关联),也就是把控制看作处理分布问题并把这些应用与应用隔离。比如资源的共享存取,在绝大多数的分布式系统中,是在掩盖其他用户存在的情况下进行的。共享具有透明性,因为用户意识不到别人的活动。这显然与 CSCW 的要求矛盾。CSCW 系统在很多情况下要求协同感知,这种感知不仅意味组的成员能感觉到其他成员“在哪里”与他合作,并且还包括了解其他成员的具体工作过程。例如在 Team WorkStation^[3], Kasmer (SCL media space)^[4]等原型或系统中,通过摄像机获取其他成员的脸部表情和身体动作的图象。而在视频会议中,与组成员的“见面”更是必不可少的基本要求。

(4)自治性和协同。网络中各个结点的自治性是分布式系统的一个特性。但是事实上,分布式系统中结点之间协同(含义如上)的增强会抑制自治性。从我们的观点来看,CSCW 中代表每个用户的机器是“高度自治”的,而互相之间又可紧密协同。分布式系统中自治性与协同的相互制约是由于面向系统的观点和实现分布的透明性引起的,对资源的全局管理引起自治性的丧失。

简而言之,CSCW 强调多个用户自主合作的观点,一般分布系统着重系统整体计算性能和可靠性的提高。一般分布式系统和 CSCW 特别在分布透明性这一点出现了差别。

二、CSCW 中的协调控制

CSCW 最本质的特征是支持一组用户协同工作,也就是要为一组用户提供一个协调控制机制。在一般分布式系统中,“协调”是指对许多进程或线程的调度和控制,这同样也是 CSCW 系统的要求。但在 CSCW 系统中,“协调”有其进一步的含义。在基于计算机的协同工作系统中,“协调”的意思是对各个用户的工作进行调整和集成以使其完成一个更大的目标,也就是要对一组通过计算机执行一项集体活动的进程(在很多场合下被称为 agents)之间的相互依赖和交互进行管理^[5]。

2.1 工作类别对协调控制的影响

人们共同工作,对不同类别的问题采取不同的协同形式加以解决。有两类极端的问题,即非预组织的问题和规定的任务。非预组织的问题是那些需要从许多用户那里接受创造性输入的那一类问题,这类问题通常无法预先详察和描述,如软件的设计;相反,规定的任务则反映那些例行过程性的协同机制,用来解决那些已有现成的群体解决方法的问题,比如在一些大组织里进行的开发票过程。协同的问题介于这两者之间。规定的任务很好地反映了对协同的细化的控制。而对于非预组织问题,则需要协同系统提供很大程度的自由来解决它。后一种情况反映了 CSCW 最典型的协调控制。

2.2 协调控制的要求

CSCW 本质上支持人与人之间协同丰富的形式,CSCW 系统中的协调控制机制应该反映这种丰富性。很精确地列举 CSCW 协调控制的要求还需要更多的对实际系统的观察,但下面这些要求是必需的:(1)用户必须获知工作进行的流程即来龙去脉;(2)群的结构和组织必须能显式地辨识;(3)支持群以动态和预先无法了解的方式工作;(4)群本身可以动态地改变;(5)协调控制的作用是便于协同而不是制约它。

2.3 协调控制内容

协调控制的复杂性受应用性质的影响,具体的协调控制要求包括路径选择、行动的时序安排、存取控制、一致性检查和并发控制、冲突解决、版本控制和修改传播等^[5]。分布式系统的协调控制为了透明性而预先嵌入体系结构,而不能在应用级或用户级定制。这可能不适合 CSCW 的要求,但是可以通过把协调控制机制与策略分开来达到定制的要求。对于 CSCW,并发控制和一致性问题尤其显得重要。

三、并发控制和一致性问题

数据的一致性和并发控制是分布式系统中研究最多的问题之一,也是 CSCW 系统中最重要的问题之一。CSCW 系统中(分布式系统也一样)可能引起数据不一致的原因有对数据的并发存取以及结点故障。前者引起的一致性问题通常也称作准确性问题,后者由于结点与外界失去联系而导致自身拥有的信息过时。在 CSCW 系统中,为了减少对 CPU 和 I/O 设备的竞争冲突和远程存取时延(也为了提高可靠性),共享数据常采用多副本方式存放在各个结点,这也是一般分布式系统采用的典型方法。

与分布式数据库系统类似,在 CSCW 系统中,

更新引起的保证一致性的处理必须解决三个方面的问题^[2]:(1)由于对同一数据项的并发更新引起的数据对象准确性问题;(2)若数据对象内部数据存在关联性,典型的例子如:以关键字(keys)相关联的数据库表(tables)和代表同一实体的不同媒体形式的信息,则存在数据对象内部一致性问题;(3)数据对象多个副本之间的一致性问题。

并发控制方法效率的提高主要得益于分布式数据库系统并发控制和多用户操作系统资源共享问题的研究。在分布式数据库系统中,为了提高系统的响应性和事务吞吐量,已经提出了一些为人熟知的并发控制方法,如主控制副本、时间戳以及多重协议同步等。CSCW 与一般分布式系统比较,有它自己的要求:

(1)由于 CSCW 的协同感知性,用户修改过程显式化,对于修改传播的时延要求提高,也就是说对其响应要更加迅速;

(2)要解决用户显式地撤消自己失误动作引起的一致性问题,这在非基于事务执行模式的系统中是一个需要重新考虑的难题;

(3)CSCW 系统是一个包含多维数据的媒体空间^[4],这使得基于事务执行模式的分布式数据库系统的并发控制方法不适合它的特征。

除此之外,我们认为 CSCW 和一般分布式系统的并发控制面临一个新的问题。在高速网络下,传输开销往往不再是瓶颈,新的瓶颈是与网络数据传输率无关的传播时延。以前的并发控制协议没有考虑这种变化,致使分布式数据库系统在网络速度上升时性能虽有所提高但迅速饱和,所以需要写出新的并发控制协议隐藏传播时延,有效地把可利用的网络带宽转换成事务吞吐量。在 CSCW 中,则希望提高用户行动的响应速度。Send-on-demand 协议是对这个问题的一个尝试,但它着重于提高系统的吞吐量,而没有把提高响应速度作为直接的目标,而 CSCW 的并发控制协议必须考虑这一点。

四、协同多用户界面

由于协同工作的动态性和灵活性,多用户界面成为它的难点之一。为了支持协同,协同用户必须允许用户感知其他人的活动。协同多用户界面的目的就是要建立和维持一个公共上下文(common context),允许一个用户的活动能反映在其他用户的屏幕上。这个公共上下文通过共享应用的信息来实现,对共享信息的实时的表达和操作是协同多用户界面

的主要功能。

多用户界面通过活动的传播来支持共享的程度被称为界面耦合,有以下三级界面耦合下的共享^[5]:

(1)表示级共享(紧密耦合)。每个用户前展现的是公共信息空间相同信息的相同显示,这被称为 WYSIWIS(你见即我所见);

(2)视图级耦合(中等耦合)。每个用户拥有对相同信息的表示,但表示可以不同,比如对同一数据,同一时刻不同用户可以分别通过列表形式和图形方式进行交互;

(3)对象级共享(松耦合)。每个用户拥有不同信息的表示。比如几个用户可以各自编辑同一文件的不同部分。

由于用户的知识、角色和秉性各不相同,协同多用户界面而只限定一种 WYSIWIS 共享是不适合的,因此一种不太严格的 WYSIWIS 被提了出来,在这种方式中共享窗口和私人窗口并存于同一屏幕中。更加灵活地,协同多用户界面应该支持多种显示和不同的视图。

协同多用户界面是与传播控制、并发控制以及多媒体同步紧密相关的。

五、多媒体的实时同步通讯

信息的媒体种类多样化是由最终用户的要求推动的,在 CSCW 系统中,由于直观的用户交互要求,采用多媒体不只是一种增强交互的冗余手段,而已成为一个基本的特性。多媒体的本质包含信息媒体多样性和集成性。这方面的研究已经非常丰富,在 CSCW 中,多媒体的研究主要解决网络通讯限制带来的问题,即多媒体的实时和同步问题。

(1)反馈(feedback)和馈通(feedthrough)的快速响应。反馈是用户自己行动的屏幕显示;馈通指用户屏幕对在其他机器上的用户的行动响应而引起变化;

(2)媒体间的连续同步。这特别指连续媒体如音频、视频之间的同步,以保持语义上的一致;

(3)媒体信息到达各个用户的同步。也就是保证多媒体信息同时到达所有用户,如果达不到这一点,很可能丧失协同的条理和逻辑。

存在网络带宽瓶颈和 I/O 设备速度差异的情况下,要达到上述要求是很困难的。比如第(3)点在连续媒体下尤其困难。事实上,我们认为,实时的要求和同步的要求是相互制约的,这就迫使协议必须具

(下转第67页)

描述。

定义4(循环算子 While X_1 Do X_2 End) 当条件 X_1 满足时,执行 X_2 描述,当条件 X_1 不满足时,跳过 X_2 描述,执行 End 后的描述。

定义5(分情形算子)Case X_1

While Y_1

Z_1

While Y_2

Z_2

.....

While Y_n

Z_n

End

当 X_1 的值为 Y_1 时,执行 Z_1 描述;当 X_1 的值为 Y_2 时,执行 Z_2 描述;当 X_1 的值为 Y_n 时,执行 Z_n 描述。

定义6(无条件转移算子 Goto X_1) 当遇见该算子时,无条件执行 X_1 描述。

虽然 Goto 算子会带来控制结构上的不清晰,但有时少量使用,会使控制比较方便,所以,我们还是提供了该算子。

(上接第40页)

有灵活性,应能支持应用根据自己的特点在实时和同步之间选择偏重的一方或者采取折衷的方案。这方面的研究非常活跃。

结束语 通过 CSCW 系统与一般分布式系统的比较,说明了 CSCW 对分布式系统在技术和某些概念上的继承性,提出了 CSCW 系统的特征和要求。这些特征和要求反映在上述几个方面。从中我们得出结论:CSCW 是对一般分布式系统的继承和发展。

对各个部分的详细描述超出了本文的范围。另外文中没有提到对 CSCW 的网络支持结构以及 CSCW 体系结构,由于 CSCW 还是一个迅速发展的领域,通用的体系结构远未定型,但这方面的尝试非常多。

CSCW 被认为是一个多学科的研究领域,除了计算机科学,它还要吸取组织学和心理学的知识,这反映了 CSCW 是一个高度突出人的因素的系统。

通过以上的五个算子,可以使基本描述进行适当的组合,这就为组装软件系统带来了极大的方便性。也可以使用这些算子,把基本描述组装成构件,放入构件库中,为以后重用提供支持,例如,把菜单生成组装成一个菜单管理构件,把系统初始化有关的描述组装成系统维护构件等。

结束语 本文提出的过程控制语言,已在我们开发的一个实际 MIS 项目中进行了初步试验,该项目是财政部的“八五”科研项目,其中的一个子系统已在试点单位投入运行,并又推广应用到另外两个单位。我们相信,该语言经过完善后,一定会有很好的实用价值。

参考文献

- [1] K. Brockschmidt, Inside OLE2, Microsoft Press, 1994
- [2] Richard M. Adler, Emerging Standards for Component Software, IEEE on Computer, March 1995
- [3] Meyer B., Object-Oriented Software Construction, Prentice Hall, Englewood Cliffs, N. J., 1988
- [4] L. Osterweil, Software Processes are Software Too, Proc. of the Ninth Intl. Conf. on Software Engineering, Monterey, California, April 1987

主要参考文献

- [1] Tom Rodden and Gordon S Blair, Distributed systems support for computer supported cooperative work, Comput. Communi., 15(8)1992
- [2] 杨德元(编译),分布式数据库管理系统概论,清华大学出版社,1987
- [3] Hiroshi Ishii et al., Iterative Design of Seamless Collaboration Media, CACM, 37(8)1994
- [4] Sara A. Bly et al., Media Spaces: Bringing People Together in a Video, Audio, and Computing Environment, CACM, 36(1)1993
- [5] Ting-Peng Liang et al., When client/Server Isn't Enough: Coordination Multiple Distributed Tasks, COMPUTER, 27(5)1994
- [6] Sujata Banerjee et al., Performance analysis of the send-on-demand: A distributed database concurrency control protocol for high-speed networks, Comput. Commun., 17(3)1994
- [7] Richard Bentley et al., Architectural Support for Cooperative Multiuser Interfaces, same to [6]