

多线程编程模型的研究与应用

吴 丹, 傅秀芬, 苏 磊, 林乔捷

(广东工业大学 计算机学院, 广东 广州 510090)

摘要: 在同步机制的基础上, 结合局域网内网审计系统项目的实际开发经验, 总结了流水线模型、对等工作组模型、客户端/服务器模型、主从架构模型和其它模型等共 5 种常见多线程编程模型, 并介绍如何将这些模型运用到局域网内网安全审计系统中。

关键词: 多线程; 互斥量; 条件变量; 编程模型

中图分类号: TP31

文献标识码: A

文章编号: 1007-7162(2008)01-0047-03

在多线程编程模型中, 线程间是相互独立而又相互依赖的. 使用多线程编程模型编程就是将进程的任务划分为执行的线程, 每一个线程为一个顺序的单控制流, 而所有线程都是并发、并行并且是异步执行的, 这样, 多线程编程模型具有实现进程并行计算、节省内存空间、减少系统管理开销、快速切换、(线程)通信易于实现等优点^[1]. 更重要的是多线程编程模型提供了一种新型的模块化编程思想和方法, 这种方法能清晰地表达各种独立事件的相互关系, 但是这种多线程编程模型的并发和异步优点也带来了一定的复杂度: 并发和异步机制带来了线程间资源竞争无序性. 因此需要引入同步机制来消除这种复杂度和实现线程间共享数据, 以一致的顺序执行一组操作.

由上可见, 如何使用同步机制来消除线程并发、并行和异步执行而带来的复杂度是多线程编程中最重要的问题, 也是多线程编程模型研究的核心问题.

1 多线程编程模型

线程是机器中连续的、顺序的执行属性集合^[2], 多线程编程提供一种新型的模块化思维方式和编程方式, 它将独立的或者低耦合的执行代码显式地分离, 并将这些低耦合的执行代码设计为相对独立运行的线程, 这种相对独立运行的线程必须相互依赖和相互合作才能提供一套完善而又完整的接口、服务或系统, 在此总结和介绍流水线模型、对等工作组模型、客户端/服务器模型、主从架构模型和其它模型等 5 种常见多线程编程模型.

1.1 流水线模型

流水线模型^[34]是每个线程接受上一个线程传

递过来的数据, 然后执行本线程的逻辑代码, 接着把数据处理的结果传递到下一个线程, 其中第一个线程从用户端接收数据或者接收初始化数据, 最后一个线程负责把数据输出, 整个过程就像一个串行的流水线. 其模型如图 1 所示.

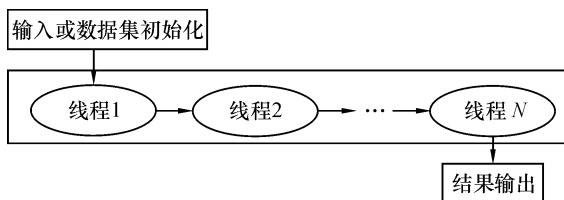


图 1 流水线模型

在流水线模型中, 整个环节的数据处理过程看似串行运行, 但其本质还是并发或并行运行的, 它与真正的串行运行程序是有本质区别的, 其主要区别在于前一线程在处理数据的同时, 后一线程可以处理前一线程传过来的结果. 流水线模型具体算法简单描述为

- 1) 上锁互斥量.
- 2) 等待本线程处理数据完毕.
- 3) 传递线程结果给到下一线程.
- 4) 发送“数据准备好”信号, 唤醒相应线程来处理已准备好的数据.

5) 解锁互斥量.

1.2 对等工作组模型

在流水线模型^[4]中, 后一个线程要等待前一个线程的处理结果, 然后依次把结果传递下去, 直至最后输出结果, 这是一种并行处理的方法. 但是在现实系统中, 还有一些数据间的耦合度不高, 它们不需要等待前一数据的结果, 而且它们的结果也不影响后一线程的运行, 像这种低耦合度的数据集可

以分别采用不同的逻辑处理,而这正是对等工作组模型所要处理的问题。

在对等工作组模型中,数据不是由一个线程独立处理,而是把低耦合度的数据分别由一组线程独立处理,而这一组独立线程之间是平等的,它们都独立地从数据集中读取数据,然后在线程内进行处理。工作组模型本质上是并发、并行和异步,其模型如图 2 所示。

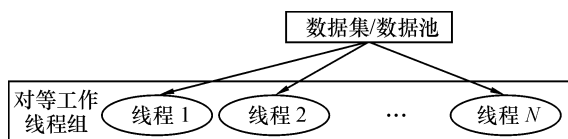


图 2 对等工作组模型

对等工作组模型具体算法简单描述为

- 1) 上锁互斥量。
- 2) 阻塞条件变量,直到有任务到来。
- 3) 解锁互斥量。
- 4) 有任务到来时重新上锁互斥量。
- 5) 复制一个数据包,并把该数据包从工作链表中删除。
- 6) 解锁互斥量,以便让其它线程从数据集/数据池中获取数据包。
- 7) 处理复制过来的数据包。
- 8) 上锁互斥量。
- 9) 数据包数量减少 1 个,如果数据包数量小于等于 0 说明已经处理完现有的全部任务,则发广播包通知其它所有线程本线程已经处理完毕。
- 10) 广播完信号后解锁互斥量。

1.3 客户端/服务器模型

客户端/服务器模型^[5]和工作组模型刚好相对,工作组中是各个线程到数据集或数据池中获取相应数据,然后并发或并行地处理数据,而客户端/服务器模型则是并发或并行地输入数据,然后统一由服务器进行处理。在客户端/服务器模型中,客户端线程请求服务器线程对一组数据进行统一处理,服务器线程独立运行,客户端线程可以等待服务器线程返回处理结果,也可以并发或并行地执行其它操作,一般地,在实际开发中客户端应该并发或并行执行其它操作,这样可以提高系统的执行性能。其模型如图 3 所示。

客户端/服务器模型具体算法简单描述为

- 1) 上锁服务器端互斥量。
- 2) 等待客户端请求的到来。
- 3) 从请求列表中复制一个请求。
- 4) 解锁服务器互斥量。

5) 服务器端处理复制过来的客户端请求。

6) 如果客户端请求中要求“同步”机制,则重新上锁服务器端互斥量,发送“请求已处理完成”通知客户端,解锁服务器互斥量。

7) 如果客户端请求是“退出”操作,则退出。

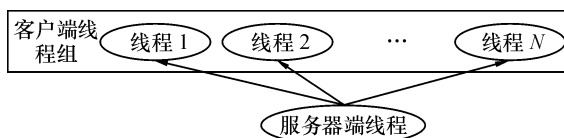


图 3 客户端/服务器模型

1.4 主从架构模型

上面描述的 3 种模型都具有一些共性: 1) 都有限制并发线程的数量。众所周知,系统资源是有限的,无限地使用线程不仅不会提升性能,反而会使性能下降。所以在实用软件系统中,限制并发线程的数量是很有必要的。2) 都没有一个线程能统一调度和平衡所有线程,这会造成线程管理和调度的混乱。而主从架构模型恰恰能弥补这两点不足。

主从架构模型类似对等工作组模型,但其主要特点是模型中有一个线程能统一调度和平衡所有的线程,而对等工作组模型则没有这一特点。主从架构模型如图 4 所示。

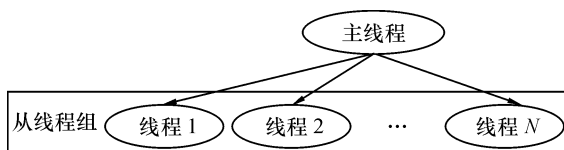


图 4 主从架构模型

根据以上流水线模型、对等工作组模型、客户端/服务器模型 3 个基本模型的具体算法不难写出该模型的逻辑处理,限于篇幅关系,这里略去。另外值得一提的是从线程组都可以使用流水线模型、对等工作组模型、客户端/服务器 3 种基本模型,从而组成一套完整而实用的多线程系统。

1.5 其它模型

多线程并发编程模型除了上面的 4 种外,其实还有很多,这就需要根据项目自身的要求来调整相关模型,但总的来说它们基本上能从上述 4 种模型中找到一个较相近的模式,以下再列出几种其它实用模型。

1) 结合主从结构模型和对等工作组模型,主线程应该能根据需要自动调整对等工作组的任务分配,这样不至于有某些线程处于忙状态,而另一些线程长时间处于空闲状态。

2) 在所有合作线程中,有些线程由于业务逻辑不太复杂而耗时较少,相反另一些线程则耗时较长,

当这两种线程以流水线模型加对等工作组模型方式合作时, 需要确保它们在往下执行前都要求到达某个特定点^[6].

2 多线程编程模型的应用

在局域网内网审计系统项目中, 根据业务需求以及对系统性能的要求, 主要模块划分为网络包获取模块、管理线程模块、网络包获取线程模块、网络包重组和业务逻辑处理线程模块以及通信模块, 每个模块对应一个线程. 其框架见图 5

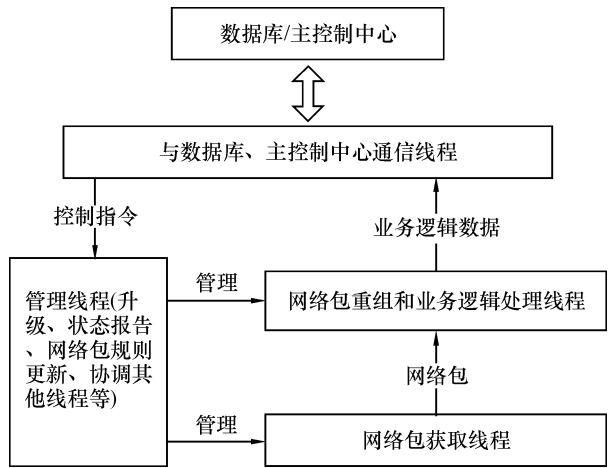


图 5 局域网内网审计项目框架图

该项目涉及到 4 个线程, 上述模型在此发挥了很大的作用. 管理模块负责系统的自动升级、状态报告、规则更新操作以及协调多线程共同工作, 从它在系统中起的作用看, 它与其他 3 个线程之间的协调主要应用了主从架构模型.

网络包获取线程模块主要负责抓取网络包, 但是对抓取的网络包不做任何处理, 既不做 IP 包重组, 也不做 TCP 流重组, 更不做网络包过滤的工作, 它只负责把网络包抓取过来, 然后把抓取到的网络包原封不动地转发到网络包重组和业务逻辑处理线程模块, 需要完成的业务流程都集中在该线程处理. 网络包重组和业务逻辑处理线程模块根据业务规则

处理有关网络包, 然后把生成的业务逻辑数据传递给通信模块, 由通信模块把有关数据转发给主控制中心. 以上这一过程中的线程合作可以使用流水线模型实现.

3 结束语

多线程编程模型提供了一种新型的模块化编程模型, 清晰地表达各种独立事件的互相关系. 本文结合局域网内网审计系统项目, 总结了流水线模型、对等工作组模型和客户端/服务器模型、主从架构模型和其它模型等共 5 种常见多线程编程模型, 在局域网内网审计系统项目中, 利用这些基本模型及其组合模式开发了性能较好的模块; 另外, 这些模型与异构平台、具体实现以及采用各种编程语言是无关的, 它们也可以根据项目的实际需求来组合使用. 下一步的研究工作包括:

- 1) 进一步深入研究各线程的控制和调度, 达到各线程间的负载平衡.
- 2) 研究多线程的数据结构和并行算法, 以达到提高线程最大粒度和程序性能的目标.
- 3) 各种模型的性能及其一致性评估.

参考文献:

[1] 骆斌, 费翔林. 多线程技术的研究与应用 [J]. 计算机研究与发展, 2000, 37(4): 407-412

[2] Bu tenhof D R. POSIX 多线程程序设计 [M]. 于磊, 曾刚, 译. 北京: 中国电力出版社, 2003

[3] 姚念民, 郑名扬, 鞠九滨. 基于流水线的高性能 Web 服务器 [J]. 软件学报, 2003, 14(6): 1127-1133

[4] 刘云生, 陈世俊. 基于多线程的流水线查询处理方式 [J]. 计算机应用, 2004, 24(6): 54-59

[5] 王成耀. 基于多层客户/服务器模型的多线程应用实现框架 [J]. 计算机工程与应用, 2001, 37(21): 147-149

[6] 戴华东, 杨学军. 以操作系统为中心的存储一致性模型——线程一致性模型 [J]. 计算机研究与发展, 2003, 40(2): 351-359

Study and Application of Multithread Programming Model

Wu Dan, Fu Xiu-fen, Su Lei, Lin Qiao-jie

(Faculty of Computer Science, Guangdong University of Technology, Guangzhou 510090, China)

Abstract Multithread technology has been adopted in developed mainstream operating systems recently. In this paper, based on synchronization mechanisms, experience of project and other references, some models of multithread programming are summed up, including three common models, one extended model and other models. Finally, those models of multithread programming are used to develop the Intranet network security auditing system.

Key words multithread, mutex, conditional variable, program model