

3 Performance Analysis Using the MIPS R10000 Performance Counters

文章主要测试 MIPS R10000 中对非侵入式监视各种处理器事件的支持：多级存储器层次结构，基于硬件的高速缓存一致性和推测执行技术

- 首先解释如何使用一组集成的硬件机制，操作系统抽象和性能工具收集性能数据
- 阐述使用计数器和分析工具帮助开发人员分析和调整应用程序

Introduction

- MIPS R10000 为计数各种类型的事件提供硬件支持：缓存未命中，内存一致性操作，分支错误预测以及几类已发布和分级指令
- 计数器可以用于：预测未来处理器和内存系统的应用程序性能和可伸缩性；分析架构权衡；生成地址跟踪或地址统计；分析总线或通信流量；评估编译器转换；分析内核；测试硬件和表征工作负载

Hardware and Operating System support

Hardware Support

- 应用程序通过内核级进程读取和写入寄存器来查询性能计数器的状态；R10000 硬件实现了两个 32 位计数器，每个计数器都可以配置为监视 16 种不同事件类型中的任何一种；可以监视总共 30 种不同的事件类型，因为两个事件类型（“循环”和“分级指令”）可以由两个计数器中的任何一个监视
- 计数器还提供内核级中断，当其中一个计数器溢出时触发该中断；该功能允许计数器产生在任何指定事件中是周期性的中断，并且允许通过在每次溢出之后设置初始计数器值来在软件中控制周期
- 性能计数器并不影响处理器性能，其实现代价很小；唯一的问题在于设计环节性能事件的选取，以及相应的信号如何导出

Operating system and library support

- 操作系统虚拟化共享计数器硬件并呈现更灵活和可共享的抽象；OS 层位于硬件上方，并且可以独立直接访问计数器硬件，而用户程序则无权访问；操作系统具有唯一访问权限，因此必须提供简单有效的接口，引入最小的必要开销
- OS 抽象包括两种计数模式：用户模式和系统模式
- 内核为用户程序维护 32 个不同的 64 位虚拟计数器，计数器通过编程接口呈现给开发人员
- 内核根据需要在调度程序中断（时钟节拍）边界上切换或多路复用事件，计算共存于每个物理计数器上的指定事件

- OS 接口允许开发人员为每个事件指定频率，内核在该事件上同步，将用户信号传递给进程；通过为缓存未命中溢出中断设计适当的处理程序，可以进行各种实验，例如内存地址跟踪
- perfex 库提供了基于 OS 接口构建的易于使用的界面

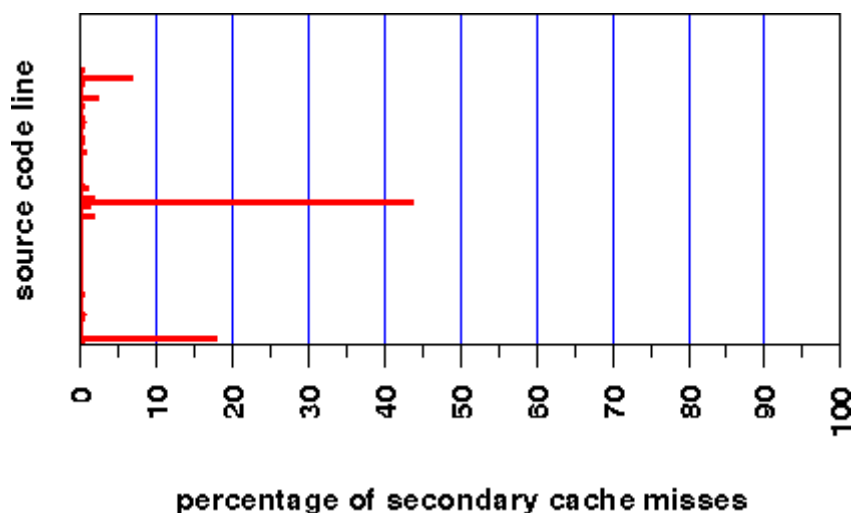
Developer Tools

- 两种性能分析工具：命令行界面 perfex；SGI 性能工具 SpeedShop
- perfex：监视两个兼容的事件计数器，使用 OS 多路复用功能对所有计数器进行采样；对于多线程应用程序，报告每个线程的数据；perfex 识别导致性能问题的处理器事件，SpeedShop 性能工具识别遇到该问题的程序源部分
- SpeedShop：根据硬件计数器的溢出情况提供统计分析；在函数和近似源行级别显示程序源中触发溢出的位置

Examples

Cache-miss Profiling

- SpeedShop 形成的直方图如上所示，其中横坐标代表二级 Cache 的缺失率，纵坐标代表引起缺失的指令位置



- 通过分析源代码，Cache 缺失的原因是二级 Cache 的一个该行无法存放下该 *for* 循环需要的数组大小

Multi-level Memory Hierarchy Profiling

- 分析实例：Alternating Direction Implicit (ADI) method for solving a set of three-dimensional partial differential equations

- 图 4 显示了 X, Y 和 Z 扫描的性能与网格大小的关系：时间由内存操作占主导地位：转换后备缓冲区（TLB）未命中的时间约为 45%，第二级数据缓存未命中时为 45%，第一级数据缓存为 5% 未命中

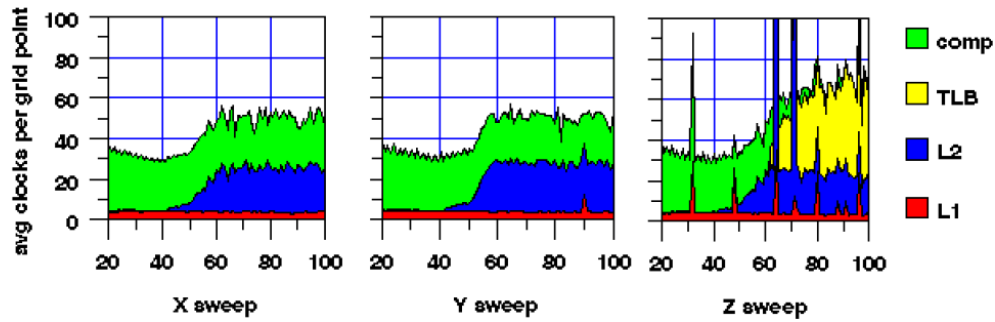


Figure 4: Time for X, Y, and Z sweeps of ADI as a function of grid size.

- 分析该行为后的优化措施有：array padding, loop interchange, loop fusion
- 优化之后的性能分析图 5 所示, Y-sweep 的二级 Cache 和 comp 缺失率下降；Z-sweep 的 TLB, 二级 Cache 以及 comp 缺失率下降

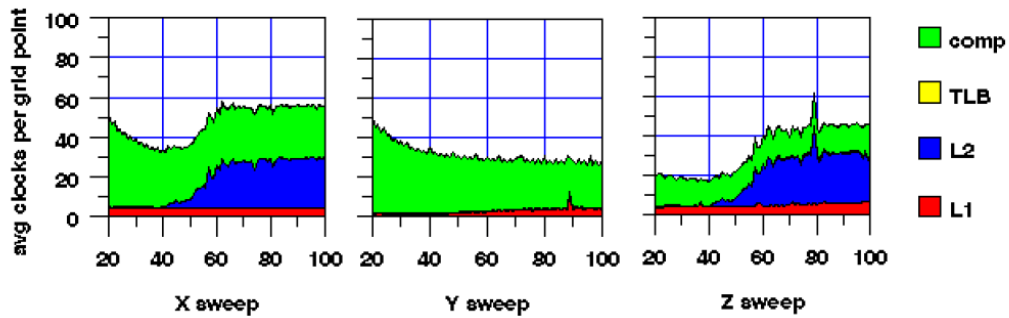
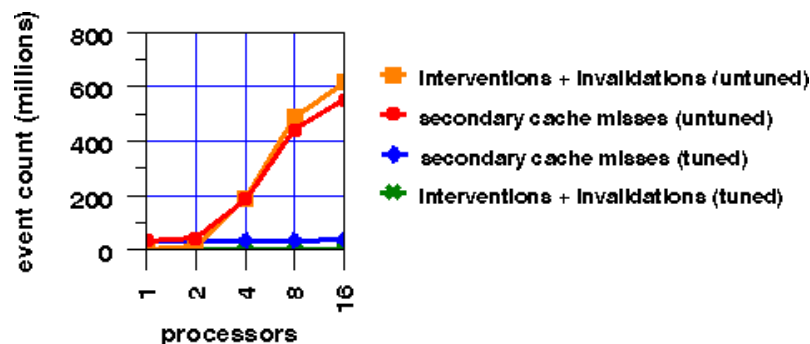


Figure 5: Time for optimized X, Y, and Z sweeps as a function of grid size.

Multiprocessor Sharing

- 天气建模程序的多处理器性能分析
- 如下图所示，二级 Cache 缺失导致性能问题



- 随着处理器数量的增加，辅助数据缓存未命中数会增加；干预和无效的总数密切跟踪二级缓存未命中的数量，表明高度共享
- 多处理器的错误共享导致了性能的下降
- 优化之后：一致性事件数量和并行加速的显著性能提升

Branch Prediction

- 二维卷积子程序的性能分析
- 如图 8 所示，分支预测的错误率很高

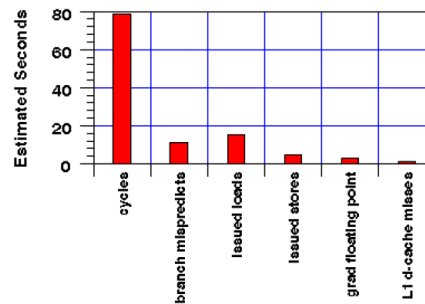


Figure 8: Estimates of event costs in the original convolution subroutine.

- 错误原因：由卷积掩模上的最短内环（3x3）引起的，R10000 使用基于历史信息的分支预测器；对于嵌套的 *for* 循环，硬件通常会预测最内层循环上的所有分支都被采用，因此将在内循环的最后一次迭代中重复进行不正确的预测。由于该程序中使用了短内循环，误预测率预计接近三分之一
- 优化之后：分支预测的正确性提高，如图 9 所示

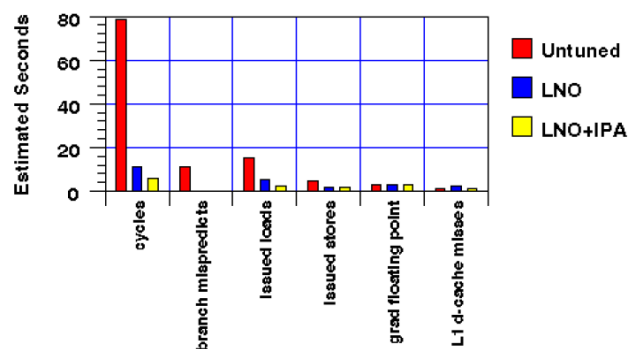


Figure 9: Performance improvements from loop nest optimization and inter-procedural analysis.