基于Hpctoolkit的显存泄露检测

# 摘要

基于GPU加速的异构体系发展越来越快，越来越多的高性能程序以及深度学习框架使用GPU进行加速。具有显式显存管理的语言要求程序员手动取消分配该程序不再需要的显存空间。程序员经常在编写程序的过程中忘记释放申请的空间从而导致程序崩溃等不可预测的现象发生。在长时间运行的应用程序中，显存泄漏尤其严重。另外，由于检测到的显存泄漏很少，每次泄漏的大小不固定，这是最难检测到的错误之一。在我们开发的程序中，显存泄漏仍然是一个很重要的问题。

在具有显式显存管理的程序中提出了一种基于上下文的算法来保护显存避免泄漏。

基于分配-释放匹配的泄漏检测算法：检测异构程序中未被释放的显存大小及代码位置。

# 关键词

显存管理；显存泄漏；代码分析；

# 介绍

新兴的超级计算机越来越多地使用GPU加速器。 这些GPU加速的系统不仅提供比仅使用常规多核处理器构建的系统更高的性能，而且这些加速的系统还提供更高的电源效率。 此类GPU加速系统的使用日趋广泛，促使研究人员开发了新技术来分析这些系统的性能。要开发适用于GPU加速系统的性能工具，我们需要回答两个问题：1）我们要收集哪些数据？ 2）我们如何收集它？迄今为止，关于异构体系结构性能分析的最新工作（例如[18，21]）集中在识别GPU内核中的性能问题上。 GPU内核级问题很重要，这只是更大问题的一方面。 整个应用程序性能分析对于调整大型GPU加速的应用程序同样重要。 此类分析需要性能数据的系统级视图。 因此，数据收集的问题简化为决定哪种系统级分析可以最好地增强标准组件级配置文件和跟踪。Luk等人的研究。 [10]和宋等。 [22]演示了在CPU和GPU之间动态分区应用程序的工作，对于为各种应用程序提供高效率非常重要。

# 背景

## Hpctoolkit

Hpctoolkit是一套集成的工具套件，可支持对顺序和并行程序进行度量、分析，可视化展示应用程序性能。基于Hpctoolkit工具扩展针对NVIDIA GPU以及AMD DCU的显存分析功能。

HPCTOOLKITis是一套集成的工具，支持对顺序和并行程序的应用程序性能的度量、分析、属性和表示。在完全优化的并行程序中，hpctoolkit可以精确地确定和量化可伸缩性瓶颈，其测量结果仅为几个百分点。最近，hpctoolkit中增加了新的功能，可以在不需要任何编译器支持的情况下收集完全优化的代码的调用路径配置文件，可以在多线程程序中精确定位和量化瓶颈，可以使用newuser inte研究性能信息和源代码。[1]

RiceUniversity正在开发HPCToolkit [2]，一种用于准确测量和查明性能瓶颈的性能工具包。HPCToolkit使用新颖的技术来测量和分析并行程序。 特别是，它使用硬件性能计数器的统计采样，并将度量标准归因于它们发生的调用上下文和程序结构，包括循环和内联过程。

HPCToolkit [3]包括用于测量并行程序的完全优化的可执行程序的性能，分析应用程序二进制文件以将测量值与程序结构相关联的组件，以及用于查明性能瓶颈的新颖分析技术。 我们设计了HPCToolkit来执行以下操作[2]:

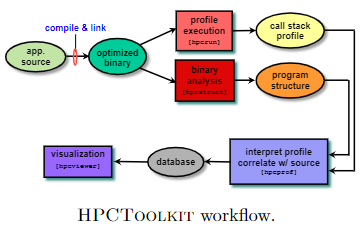
•Work at binary level for language indepen-dence.This enables HPCToolkit to supportmeasurement and analysis of multi-lingualcodes with external binary-only libraries.

•Profile rather than adding code instrumenta-tion.Sample-based profiling is less intrusivethan code instrumentation and requires onlyvery modest data volume.app. sourceoptimizedbinarycompile & linkcall stack profileprofile execution[hpcrun]binary analysis[hpcstruct]interpret profilecorrelate w/ source[hpcprof]databasevisualization[hpcviewer]program structureHPCToolkitworkflow.

•Collect and correlate multiple performance metrics.Performance problems typically cannotbe diagnosed with only one species of event.

•Compute derived metrics to aid analysis.Synthetic metrics, such as memory bandwidthconsumed, often provide insight for optimization.

•Attribute costs very precisely.HPCToolkitis unique in its ability to associatemeasurements with dynamic calling context, loops, and inlined code.



# 分配-释放匹配算法

# NVIDIA GPU显存分析

# AMD DCU显存分析

# 可扩展性

# 总结

# 参考文献

[1]HPCTOOLKIT: Tools forperformance analysis ofoptimized parallel programs

[2] HPCToolkit: performance tools for scientific computing

[3] Rice University 2008 HPCToolkit performance tools.http://www.hipersoft.rice.edu/hpctoolkit

[4] Effective Sampling-Driven Performance Tools for GPU-Accelerated Supercomputers