



基于GPU的LCS算法加速机制研究与实现

作者：张常志 牟澄 黄小红 马严

报告人：张常志

北京邮电大学

目录

1

背景介绍

2

CUDA简介

3

算法并行化研究和实现

4

实验及结果分析

5

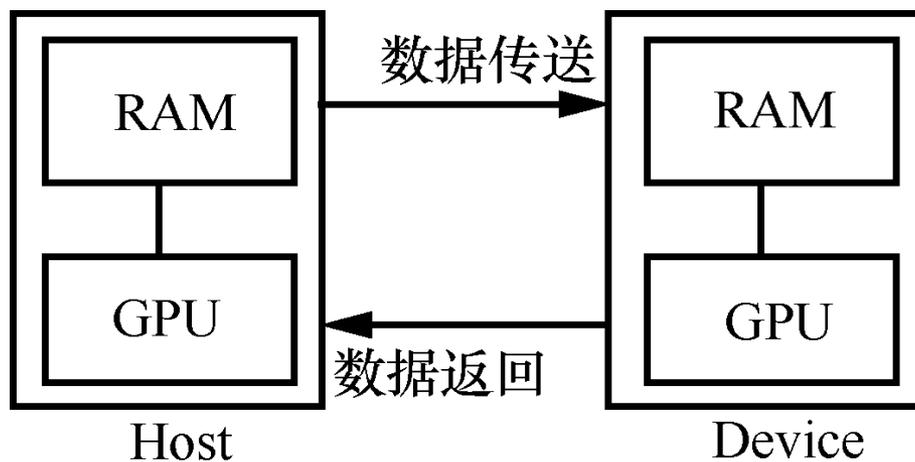
结束语

1.背景介绍

- 传统的基于端口识别应用层协议的算法
 - 准确率低
 - 基本被淘汰
- 基于负载的协议识别技术
 - 即DPI(Deep Packet Inspection)技术
 - 被广泛应用
- LCS(Longest Common Subsequence)算法
 - 即最长公共子序
 - 缺点：时间复杂度高，算法效率低

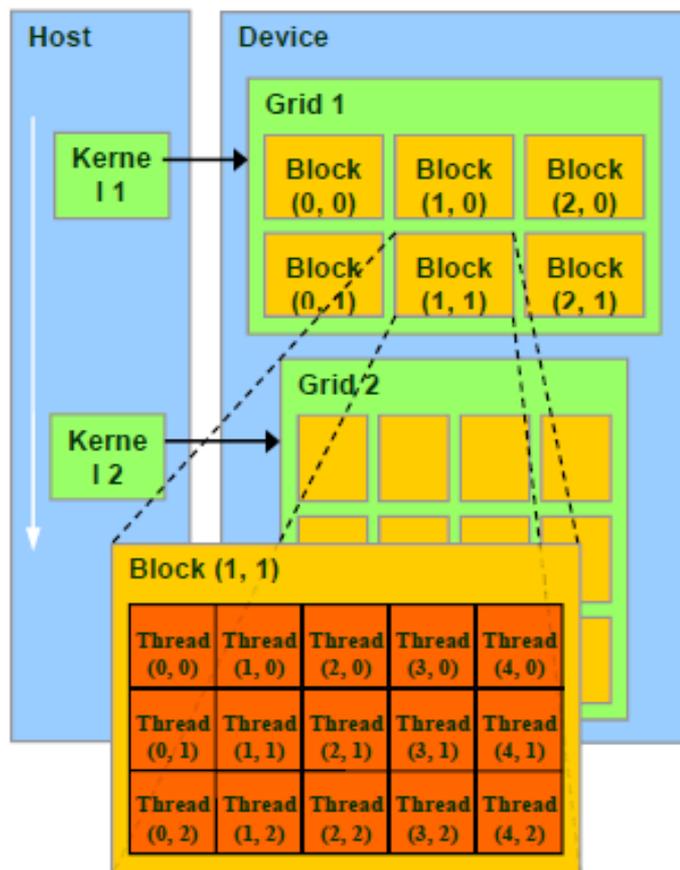
2. CUDA简介

- 提出了基于CUDA(Compute Unified Device Architecture)平台的解决方案
- CPU和GPU数据传送结构图



2. CUDA简介

- GPU的并行线程组织结构



3.算法并行化研究和实现

- 矩阵算法

- S1=“hang0an” S2=“09ngzhan”

	h	a	n	g	0	a	n
0	0	0	0	0	1	0	0
9	0	0	0	0	0	0	0
n	0	0	1	0	0	0	1
g	0	0	0	1	0	0	0
z	0	0	0	0	0	0	0
h	1	0	0	0	0	0	0
a	0	1	0	0	0	1	0
n	0	0	1	0	0	0	1

3. 算法并行化研究和实现

- 经过改进后所得匹配矩阵

		h	a	n	g	0	a	n	h
	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	1	0	0	0
9	0	0	0	0	0	0	0	0	0
n	0	0	0	1	0	0	0	0	0
g	0	0	0	0	2	0	0	0	0
z	0	0	0	0	0	0	0	0	0
h	0	1	0	0	0	0	0	0	0
a	0	0	2	0	0	0	1	0	0
n	0	0	0	3	0	0	0	2	0

3.算法并行化研究和实现

- 矩阵算法
 - 优点：时间复杂度小
 - 缺点：空间复杂度大，如两个100K的文件需要消耗10G的内存，不现实
 - 该算法的并行化受文件大小的限制

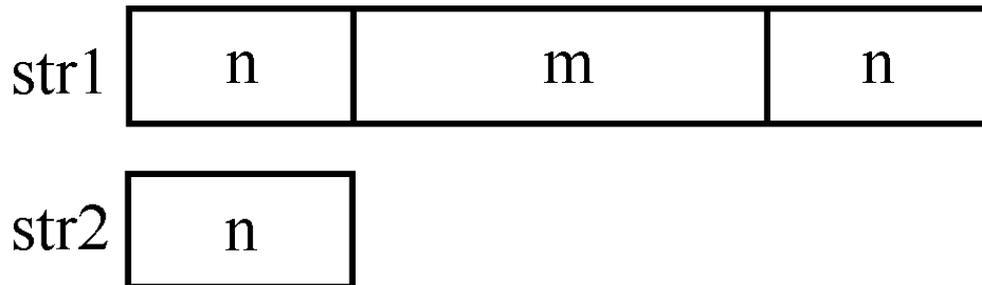
3.算法并行化研究和实现

- BF算法

- BF(Brute Force)算法是普通的模式匹配算法，BF算法的思想就是将目标串S的第一个字符与模式串T的第一个字符进行匹配，若相等，则继续比较S的第二个字符和T的第二个字符；若不相等，则比较S的第二个字符和T的第一个字符，依次比较下去，直到得出最后的匹配结果
- 时间复杂度大，空间复杂度小
- 适合并行化

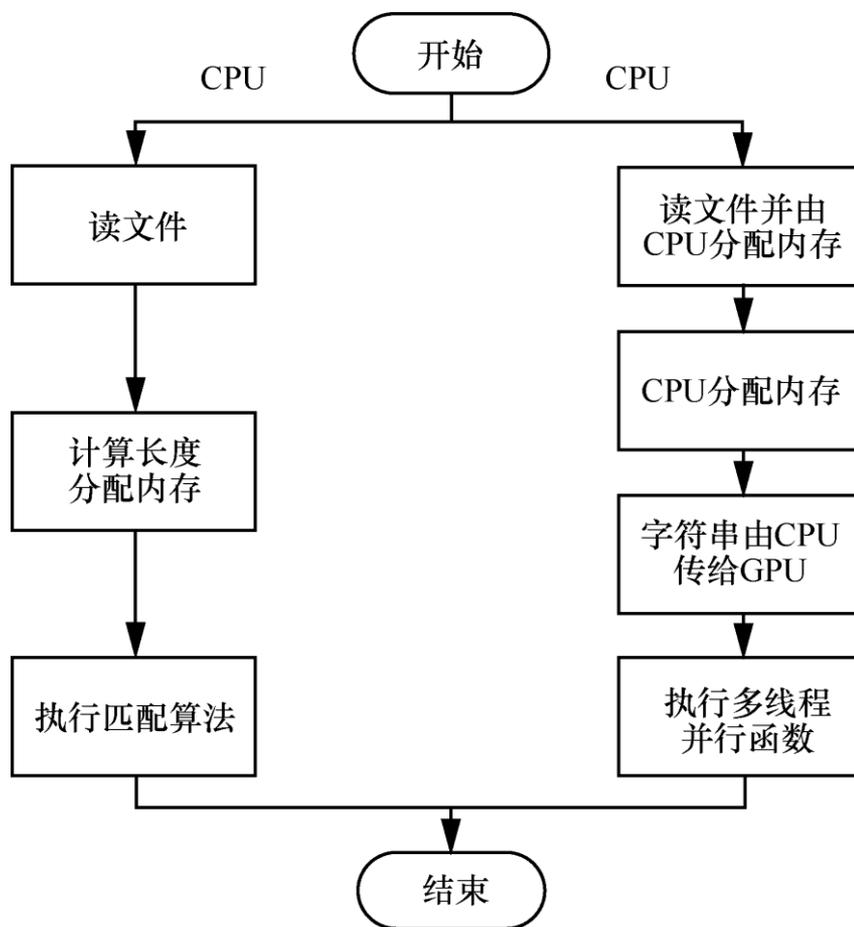
3.算法并行化研究和实现

- 并行化后的BF算法
 - 将其中一个文件进行扩展
 - 一趟比较作为一个thread来执行



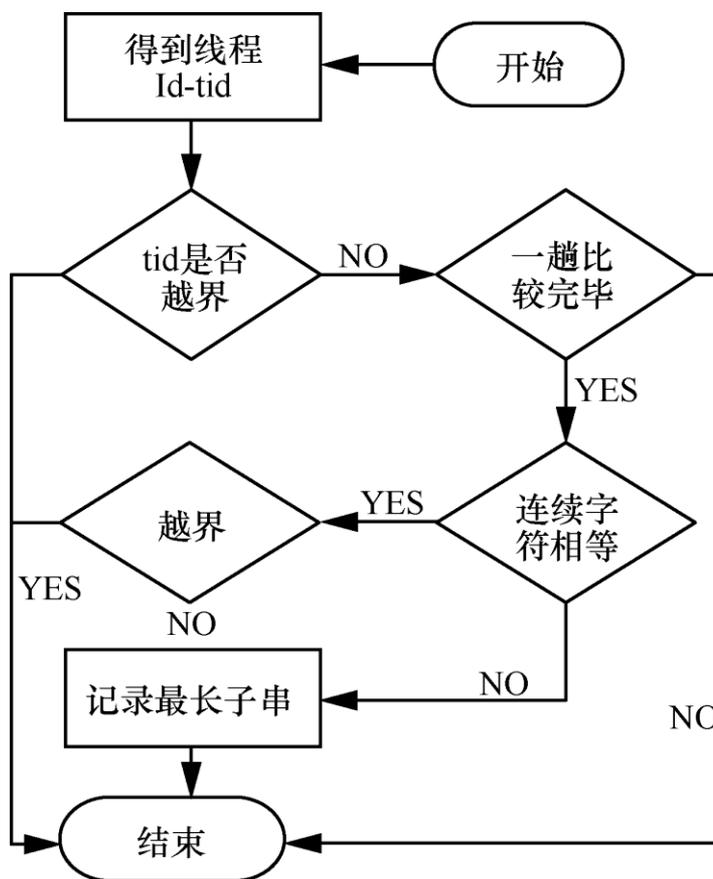
3.算法并行化研究和实现

- CPU与GPU算法流程图



3.算法并行化研究和实现

- 关键算法（核函数）流程图



4.实验及结果分析

- 实验环境
 - 系统搭载服务器环境为Ubuntu10.04
 - CUDA各软件版本 3.20
 - GPU型号GeForce 9800 GT
- 实验数据
 - 两个大小基本一致的文件

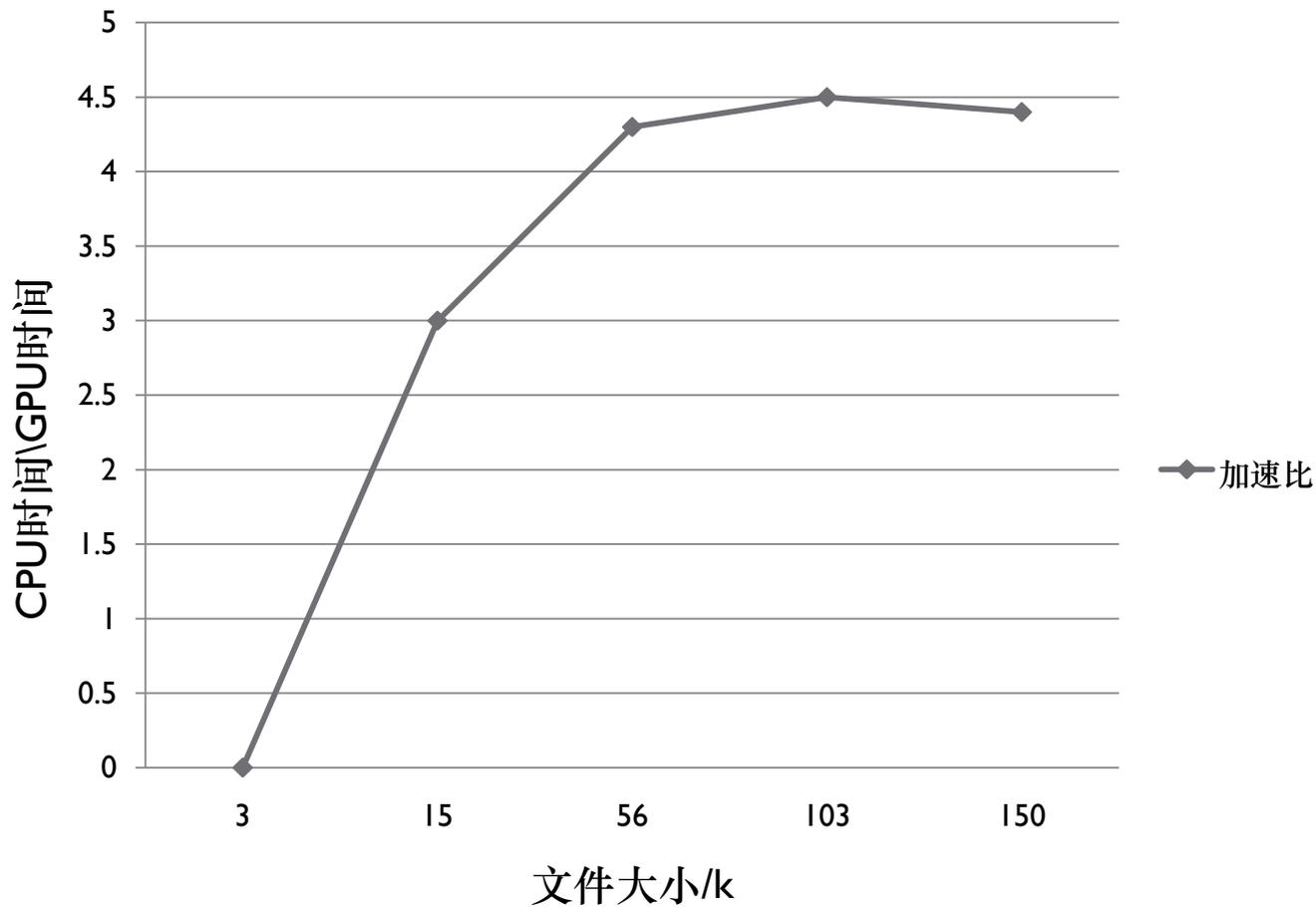
4.实验及结果分析

- GPU和CPU运行时间比对

File1大小/k	File2大小/k	CPU运行时间/s	GPU运行时间/s	加速比
3.31	3.5	0	0	0
15.1	15.6	3	0	3
56.6	56.6	47	11	4.3
103	104	165	37	4.5
150	150	342	78	4.4

4.实验及结果分析

- 加速比曲线



4.实验及结果分析

- 加速比（CPU运行时间/GPU运行时间）大概固定在4.5左右
- 验证了GPU对LCS算法确实有加速作用，但是加速的倍数还是比较小，这主要是由于核函数中控制语句过多，影响了并行化效果

5.结束语

- 本文通过对LCS算法的并行化分析及CUDA平台的搭建，验证了GPU的加速作用，对算法的并行化作了初步的探索
- 同时发现，该平台的搭建还比较繁琐，编程有一定的难度，CUDA的学习理解还是一个复杂的过程
- 从如何减少逻辑控制语句方面进行更深入的分析



谢谢!