



## SALÃO DE INICIAÇÃO CIENTÍFICA XXVIII SIC

paz no plural



<b>Evento</b>	Salão UFRGS 2016: SIC - XXVIII SALÃO DE INICIAÇÃO CIENTÍFICA DA UFRGS
<b>Ano</b>	2016
<b>Local</b>	Campus do Vale - UFRGS
<b>Título</b>	Suscetibilidade a falhas críticas em algoritmos de classificação
<b>Autor</b>	CAIO BRIGAGÃO LUNARDI
<b>Orientador</b>	PAOLO RECH

# Susceptibilidade a falhas críticas em algoritmos de classificação

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Caio Brigagão Lunardi – Paolo Rech (Orientador)

Alta eficiência energética é um fator imprescindível em inúmeros cenários atuais, tanto em sistemas de alto desempenho (HPC), quanto em sistemas embarcados. No cenário de HPC, atualmente, são empregadas placas de vídeo (GPUs), dispositivos de computação altamente paralela. Tais dispositivos são utilizados em sistemas gigantes de HPC, como o Titan, o segundo supercomputador mais potente do mundo em 2015, que opera com mais de 18 mil GPUs. Porém, com um número de dispositivos tão elevado, devem ser consideradas as falhas nestes dispositivos induzidas por radiação atmosférica. No Titan, dados reais evidenciam uma taxa de um erro não detectável a cada 8h de serviço. Para reduzir tais efeitos, se provou necessária uma análise detalhada sobre o comportamento dos algoritmos e dispositivos quando expostos a estas falhas, para que possam ser desenvolvidas técnicas de prevenção adequadas.

Alguns dos algoritmos mais relevantes na computação de alta performance e de alta criticalidade são os de classificação e ordenamento. Estes são empregados nas mais diversas categorias de aplicações e simulações científicas, tais quais rastreamento de estrelas de acordo com sua intensidade luminosa, codificação de vídeo, transações bancárias, sistemas de interface humana, sistemas de segurança para detecção de pedestres, entre outras. É notável que muitas destas aplicações requerem a computação de uma imensa quantidade de dados ou necessitam de respostas em tempo real, evidenciando a necessidade de implementações de alta performance em dispositivos eficientes.

No trabalho desenvolvido, foram identificadas três implementações de algoritmos de classificação e ordenamento que eram extremamente eficientes tanto em relação ao tempo de resposta quanto à capacidade de processar quantias enormes de dados. Os algoritmos escolhidos foram o QuickSort, o MergeSort e o RadixSort, todos desenvolvidos pela própria empresa que desenvolve as GPUs, permitindo otimizações específicas para estes dispositivos. O QuickSort foi implementado com a tecnologia Dynamic Parallelism (DP), que permite à GPU realizar mais cálculos úteis sem necessitar de excessivas sincronizações com o sistema nativo. O MergeSort é um algoritmo que se baseia na filosofia “dividir para conquistar”, dividindo o problema de ordenamento inicial em diversos outros problemas pequenos e independentes, permitindo um amplo uso da arquitetura paralela das GPUs. O RadixSort é um algoritmo altamente otimizado para GPUs, baseado na biblioteca Thrust, feita em linguagem CUDA. Esta implementação emprega a tecnologia DP para redistribuir o trabalho entre os vários núcleos da GPU.

O objetivo do trabalho foi o de identificar as falhas ocorrentes na computação dos algoritmos descritos anteriormente quando as GPUs eram expostas a falhas induzidas por radiação de nêutrons. Foram realizados testes experimentais em colaboração com o Los Alamos National Laboratory (LANL), nos Estados Unidos, onde os dispositivos testados foram submetidos a um fluxo de nêutrons na faixa entre 10 e 750 MeV, simulando a radiação presente ao nível do mar, com fluxo de 10 ordens de magnitude superior. Tais testes experimentais foram imprescindíveis para coleta de dados estatisticamente relevantes. Foi concluído que, dependendo da aplicação, boa parte das falhas podem ser facilmente detectadas e então corrigidas com métodos propostos no trabalho, métodos estes que não impactam significativamente o desempenho da aplicação. Existem, porém, outros tipos de falhas críticas, como, por exemplo, interrupções no funcionamento do dispositivo ou corrupção de grandes partes do resultado com erros que não são facilmente detectados sem que seja necessário um maior custo de desempenho.