

Identifying and Predicting Timing-Critical Instructions to Boost Timing Speculation

Xin, Jing and Joseph, Russ. *Identifying and predicting timing-critical instructions to boost timing speculation*. Proceedings of the 44th Annual IEEE/ACM International Symposium on Microarchitecture, MICRO-44 '11. pages 128–139. ACM, Porto Alegre, Brazil, 2011.

Walisson Ferreira Pereira
RA 115168

I. OBJETIVOS

Os autores propuseram realizar simulações ao nível de porta lógica para explorar a erro temporal de localidade em instruções estáticas nos quais os padrões de uso de dados são sensíveis as falhas de atraso no trajeto. Outro objetivo foi a avaliação de mecanismos simples que predizem instruções prováveis de produzirem erros temporais no pipeline e fazer preparativos para a redução do custo de recuperação.

II. PROCEDIMENTOS

Os experimentos realizados focaram no impacto que a predição de erro temporal e o preenchimento de erros têm num processador ambarcado de alto desempenho e baixo consumo energético. Um processador de 8 estágios foi simulado junto a um modelo detalhado a nível de porta lógica de unidades inteirais funcionais com erros temporais na execução do pipeline.

Os modelos implementados são suficientemente completos para virtualmente suportar todas as intruções inteiras em modo usuário sem as instruções MVI SIMD. Baseado-se em ciclos, os autores conduziram o simulador a nível de porta lógica com valores de programa que permitissem modelar o acoplamento entre o padrão controle com o fluxo de dados e os erros temporais na execução das unidades.

III. RESULTADOS

Com o objetivo de medir a acurácia da medição, os autores exploraram as estratégias de predição de erros temporais *Last Value*, MDC (*Miss Distance Counter*) com 2, 3 e 4 bits, e Sat (*Saturating Counter*) com tamanho de 32 a 2048 entradas. A estratégia MRE (*Most Recent Entry*), foi avaliada de 2 até 128 entradas.

Os resultados da análise de sensibilidade das estratégias em relação ao seu custo, o MDC com 4 bits obteve as melhores acurácias. A estratégia MRE obteve melhores números de acertos de predição (PPV, de *Positive Predictive Value*). Em contra partida, o MRE teve a pior sensibilidade e o MDC com 4 bits o pior PPV. Outra análise feita foi a penalidade sofrida pelas estratégias, e a partir desses resultados, os autores escolheram realizar uma análise mais detalhadas a estratégia MDC com 2 bits e 2048 entradas e o MRE de 32 entradas.

Os autores avaliaram a eficiência da predição de erros em aplicações independentes do ajuste de tensão. Nesse teste, a tensão (VDD) de operação era fixa em 70%, 80% e 90% da tensão nominal, obtendo redução média de 80% das penalidades causadas por erros temporais.

Os autores também testaram com a técnica de ajuste de tensão *Hill Climbing*, *Sampling* e *Fuzzy Controller* e comparado seu atraso energético com um base livre de erro, onde o sistema opera em máxima tensão.

Geralmente a previsão de erro e o preenchimento podem economizar mais energia trabalhando acima da tensão crítica e mascarando os erros de predição de tempo. Dentre outros *benchmarks* realizados pelos autores, ambos os preditores (MRE e MDC de 2 bits) reduziram o atraso energético em aproximadamente 21% no compilador gcc.

IV. DISCURSÕES E CONCLUSÕES

A fronteira da tensão crítica é uma limitação que a especulação temporal enfrenta, pois impede sobrecarregar o processador além de ponto específico. O trabalho dos autores podem ser aplicados no projeto com técnicas de otimização para reduzir o impacto da fronteira crítica de tensão e impulsionar a especulação temporal.

Os modelos analíticos limitam a análise a pequenas porções do processador. Por isso, os autores trabalharam com simulação, e mesmo assim, isso ainda era o gargalo em sua pesquisa. Eles esperam que outros pesquisadores desenvolvam novas metodologias e modelos para acelerar a simulação.

A variação de parâmetros pode alcançar alguns ganhos ao se explorar diferentes sensibilidades de entrada, porém mesmo com a variação de parâmetros, os erros temporais de localidade ainda são proeminentes, pois em geral um pequeno número de instruções é responsáveis pela maioria dos erros temporais e há uma tendência de instruções estáticas executarem na mesma unidade funcional.

Nas conclusões finais, os autores concluem que as instruções estáticas são mais previsíveis e possuem taxas de erros estáveis (erro de temporização por localidade); e que a maioria dos erros temporais ocorrem devido a um pequeno número de instruções estáticas que tendem a terem taxas de erros estáticas.