



Universidade Federal de Santa Catarina (UFSC)
Programa de Pós-Graduação em Engenharia de Automação e Sistemas (PPGEAS)

Como garantir a confiabilidade dos requisitos temporais?

Karila Palma Silva
karila.palma@posgrad.ufsc.br

Rômulo Silva de Oliveira
romulo.deoliveira@ufsc.br

Novembro de 2016

1 / 53

Considerações Iniciais

- Sistemas de Tempo Real (STRs)
 - Sistemas computacionais que estão submetidos, além de a requisitos de natureza lógica, a requisitos de natureza temporal
 - Os resultados devem estar corretos não somente do ponto de vista lógico, mas também devem ser gerados no prazo correto
 - Podem ser classificados de acordo com a criticalidade dos seus requisitos temporais:
 - STRs críticos ou *hard real-time*
 - STRs não críticos ou *soft real-time*
 - Para garantir o cumprimento dos requisitos temporais, a análise de tempo como parte de um processo de certificação se baseia no conhecimento adequado do tempo de resposta no pior caso *WCRT*

3 / 53

Conteúdo

- 1 Considerações Iniciais
- 2 Tempo de Resposta no Pior Caso
- 3 Panorama Atual
- 4 Teste
- 5 Métodos Estatísticos
- 6 Motivação
- 7 Trabalhos Relacionados
- 8 Experimento
- 9 Considerações Finais

2 / 53

Considerações Iniciais

- Tempo de Resposta no Pior Caso - *WCRT*
 - Delimitam estímulos/respostas
 - Processamento devem terminar dentro de prazos
 - Se terminar fora de prazo sistema falha (falha temporal)
 - Obtenção do *WCRT*
 - A partir do *WCET* usando métodos analíticos
 - Diretamente, como em um teste temporal

4 / 53

Previsibilidade

- Previsibilidade determinista - PD
 - Todos os prazos serão respeitados
 - Antecipação determinista do comportamento temporal do sistema
 - Está cada vez mais difícil devido ao uso de elementos de hardware complexos
- Previsibilidade probabilística - PP
 - Fornece uma probabilidade para o cumprimento dos requisitos temporais
 - Antecipação probabilística do comportamento do sistema
 - Baseada em estimativas ou simulação que estipulam probabilidades dos prazos serem atendidos
 - Podem acarretar em incertezas relacionadas ao ajuste dos dados ao modelo
- Para suportar a demanda computacional das aplicações executadas em STRs do futuro, é necessário o desenvolvimento de abordagens que garantam com altos níveis de confiança a PP

5/53

Panorama Atual

- Na indústria
 - Geralmente utilizam medição fim a fim (testes) para estimar o limite superior do tempo de resposta, para um subconjunto de possíveis cenários de execução
 - Os testes são baseados na especificação de requisitos do programa
 - Nenhum conhecimento de como o programa está implementado (caixa preta)
 - Não tem disponível uma modelagem formal do *hardware*
 - É difícil determinar se os valores obtidos para o *WCRT* são seguros
 - O *WCRT* pode ocorrer com pouca frequência e as condições para que ele aconteça são normalmente desconhecidas
 - Trabalham com margens de segurança para lidar com o problema da incerteza de que o pior caso é coberto por testes
 - As margens de segurança não são o resultado de modelos analíticos, mas de experiências em aplicações semelhantes

6/53

Panorama Atual

- *Linux - Realtime Testing Best Practices*
(http://elinux.org/Realtime_Testing_Best_Practices)
 - Programas de teste
 - Medição
 - Estresse
 - Teste de *Hardware*
 - *Target*: o sistema em teste
 - *Host*: sistema de controle, e também coleta os dados
 - *Logger*: máquina especial usada para causar interrupções no *target*, e registrar o tempo que ele leva para responder
 - Problemas e técnicas - Desempenho de teste em TR no Linux

7/53

Panorama Atual

- Problemas e técnicas - Desempenho de teste em TR no Linux
 - Teste de estresse
 - *Nicholas McGuire* afirmou: teste de estresse pode ser pobres
 - Como o sistema é forçado? (memória, caminho de código ruim)
 - Teste de latência
 - *Ingo Molnar* acredita que o sistema tem alguma latência de hardware que mascara as capacidades do RTOS subjacentes
 - Número de amostra
 - *Ingo Molnar*: a diferença de latência é bastante significativa entre 1.000.000 amostras e 100.000.000 amostras
 - Aumentar a taxa de interrupções geradas pelo *logger*
 - Cenário de teste
 - *Nicholas McGuire*: o grande problema com os testes RT é que estão olhando para o bom caso, assumem as operações bem sucedidas
 - Os piores casos, quando executar nos caminhos de erro do *kernel*, em seguida, uma aplicação trivial, pode induzir instabilidade no sistema

8/53

Teste de Software

- A utilização de medições para estimar o *WCRT* de uma tarefa corresponde a uma atividade de teste não-funcional (temporal)
- Para auxiliar na formalização da atividade de teste existem alguns elementos essenciais, dos quais pode-se destacar:
 - **Caso de teste:** descreve uma condição particular a ser testada e é composto por valores de entrada, restrições para a sua execução e o resultado ou o comportamento esperado
 - **Procedimento de teste:** é uma descrição dos passos necessários para executar um caso de teste (ou um grupo de casos de teste)
 - **Critério de teste:** utilizado para selecionar e avaliar casos de teste de forma a aumentar as possibilidades de provocar falhas. Quando não há ocorrência de falha, é utilizado para estabelecer um nível elevado de confiança na correção do produto

9/53

Teste de Software

- Os critérios de teste podem ser utilizados como:
 - **Critério de cobertura dos testes:** permitem a identificação de partes do programa que devem ser executadas. Para garantir a qualidade do *software* e indicar quando este foi suficientemente testado
 - **Critério de adequação de casos de teste:** avalia se os casos de teste definidos são suficientes ou não, para a avaliação de um produto ou uma função
 - **Critério de geração de casos de teste:** define as regras e diretrizes para geração dos casos de teste de um produto que esteja de acordo com o critério de adequação definido anteriormente

10/53

Teste de software

- Teste de *software* é uma área de fundamental importância para a garantia da qualidade do *software*
- São atividades que envolvem tempo e custos elevados, mas que precisam ser realizadas durante todo o processo de construção de um *software*
- Existem problemas cuja solução não é trivial. Para esses problemas, têm sido exploradas técnicas de busca e otimização tentando encontrar uma solução ótima ou perto da ótima
- Dando origem as linhas de pesquisa:
 - **Engenharia de *software* baseada em busca** (SBSE - *Search-Based Software Engineering*)
 - **Teste de *software* baseada em busca** (SBST - *Search-Based Software Testing*)

11/53

Teste de software

- O termo *SBSE* consiste na aplicação de técnicas heurísticas para buscar soluções aproximadas
- Os seguintes critérios devem ser considerados:
 - O espaço de busca do problema deve ser grande, que inviabilize a busca da solução através de um método exato em tempo polinomial
 - Inexistência de soluções completas, isto é, solução para o problema para qualquer instância em tempo polinomial
 - Existência de uma função objetivo adequada
 - As soluções candidatas devem ser encontradas em tempos aceitáveis
- Soluções ótimas ou sub-ótimas são procuradas em um espaço de soluções candidatas, tendo como guia, em geral, uma função objetivo, que distingue as soluções melhores das piores

12/53

Teste de software

- A abordagem *SBST* está sendo muito aplicada na área de testes, verificando anomalias no comportamento do *software*
- A definição do *WCET* de um determinado *software* pode ser entendida como um problema de busca complexo
- O resultado calculado depende da arquitetura computacional onde está instalado o *software*
- A natureza dos problemas da área são problemas do mundo real, geralmente relacionados a objetivos que não podem ser caracterizados por um conjunto de equações lineares, e assim, difíceis de serem tratados por métodos determinísticos
- Este problema tem sido modelado como um problema de busca e aplicado algumas soluções heurísticas para calcular *WCET*

13/53

Teste de software

- Uma abordagem de teste de *software* baseado em busca (SBST)

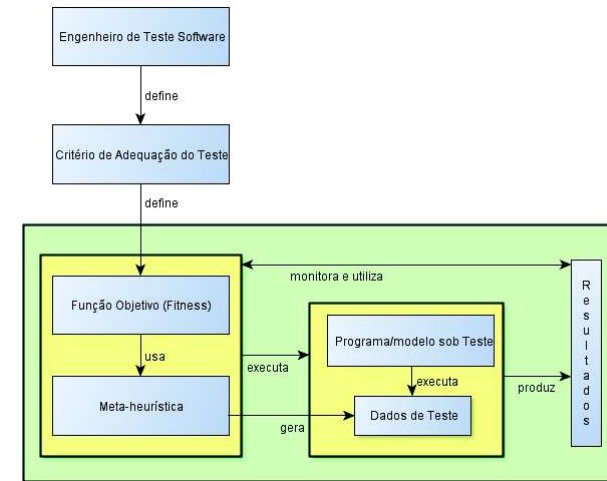


Figura: Solução de otimização na geração de dados de teste

14/53

Teste para STR

- A natureza assíncrona, depende do tempo, de aplicações em tempo real acrescenta um elemento potencialmente difícil ao conjunto de testes - **o tempo**
- Os recursos disponibilizados para teste devem possibilitar a verificação lógica e temporal
- O projetista do caso de teste tem que considerar não apenas os casos de teste convencionais, mas também
 - A manipulação de eventos (o processamento de interrupção)
 - A temporização dos dados e
 - O paralelismo das tarefas (processos) que manipulam os dados
- O estado do sistema pode influenciar no processamento correto ou erro sobre os mesmos dados de teste
- Testes de *software* devem levar em consideração o impacto das falhas do *hardware* sobre o processamento do *software*
- As falhas podem ser difíceis de simular realisticamente

15/53

Casos de Teste em STR

- **Teste de tarefa**
 - Testar cada tarefa independente, descobrir erros em lógica e função
- **Teste comportamental**
 - Simular o comportamento de um STR e examinar este em consequência de eventos externos
- **Teste intertarefas**
 - Tarefas assíncronas são testadas com diferentes taxas de dados e carga de processamento - erros de sincronização intertarefas
 - Teste de tarefas que se comunicam via fila de mensagens ou armazenamento de dados - erros no dimensionamento
- **Teste de sistema**
 - Testes para descobrir erros na interface *software-hardware*
 - A maioria dos STR processa interrupções. É essencial testar a manipulação desses eventos *booleanos*

16/53

Validação de STR

- Para estimar a confiabilidade do *software* é utilizado teste estatístico
- É importante ter um processo bem definido e certificado para o desenvolvimento de sistemas críticos de segurança
- O processo deve incluir a identificação e a monitoração de perigos potenciais
- A validação de segurança pode ser realizada por meio da análise baseada em
 - Experiências
 - Ferramentas ou
 - Equipes que simulam ataques ao sistema
- Os casos de segurança coletam todas as evidências que demonstram que um sistema é seguro

17/53

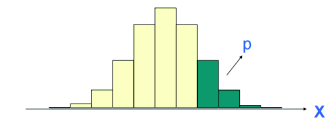
Teoria de Valores Extremos - TVE

- Os fundamentos da teoria de valores extremos foram inicialmente expostos por *Fisher & Tippett* (1928)
- Definiram os três tipos possíveis de distribuições assintóticas dos valores extremos: *Weibull*, *Gumbel* e *Fréchet*
- Um problema foi identificar o tipo de distribuição mais adequada para uma determinada amostra de dados
- *Jenkinson* (1955) desenvolveu a estimação dos parâmetros da distribuição generalizada de valores extremos (GEV)

19/53

Métodos Estatísticos

- A partir de resultados reais obtidos (amostra), conjectura (hipótese) sobre o comportamento de variáveis
- Hipótese nula (H_0): aceita como verdadeira até haver prova estatística em contrário
- Hipótese alternativa (H_1): hipótese que será aceita, se os dados mostrarem evidências suficientes para a rejeição da hipótese nula
- Calcula o p-valor e compara com o nível de significância
- Testes usuais:
 - Teste de hipótese
 - Teste de proporção



18/53

Teoria de Valores Extremos - TVE

- GVE unifica as três distribuições, e tem função de distribuição acumulada de probabilidade dada por

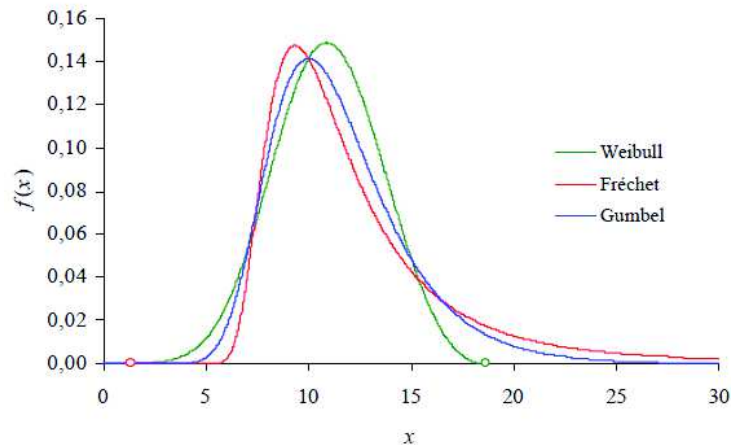
$$F(x) = \exp \left\{ - \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{\frac{1}{\xi}} \right\} \quad (1)$$

- Parâmetros: μ localização, σ escala, e ξ forma
 - *Weibull*: $\xi < 0$, cauda curta delimitada
 - *Gumbel*: $\xi = 0$, cauda ilimitada ($\xi \rightarrow 0$)
 - *Fréchet*: $\xi > 0$, cauda pesada ilimitada

20/53

Teoria de Valores Extremos - TVE

- Função densidade de probabilidade da distribuição GVE para *Weibull* $\xi = -0,3$, *Gumbel* $\xi \rightarrow 0$ e *Fréchet* $\xi = 0,3$, com $\mu = 10$ e $\sigma = 2,6$



21 / 53

Teoria de Valores Extremos - TVE

- TVE é uma estrutura de inferência estatística com o objetivo de descrever comportamento extremo raro
- Sua aplicação varia de hidrologia, financiamento, gestão de riscos, telecomunicações, e foi aplicada para derivar limites superiores probabilísticos do tempo de execução de tarefas
- Uma estimativa do pior caso de uma tarefa pode ser construído a um determinado nível de precisão estatística por meio do uso de
 - Medição para recolher dados
 - TVE de generalizar os processos que produzem estas medições
- Pode oferecer uma abordagem alternativa para testes exaustivos, bem como para análise de tempo complexa

22 / 53

Teoria de Valores Extremos - TVE

- TVE é normalmente referida como uma técnica baseada na medição e consistem em
 - (i) Medir quanto tempo uma tarefa leva para executar um conjunto representativo de suas entradas
 - (ii) Seleção a partir dessas medições de uma amostra de máximos
 - (iii) Modelo de valor extremo que melhor se ajuste aos valores máximos selecionados
 - (iv) Uma vez que o modelo derivado é validado, obtenção da estimativa de pior caso

23 / 53

Métodos Estatísticos

- Métodos estatísticos quando aplicado corretamente podem fornecer confiabilidade suficiente, para a estimativa do *WCET* obtido através de medição
- É necessário um bom ajuste dos dados ao modelo
- Os tempos medidos devem seguir distribuições de probabilidade conhecidas
- As variáveis aleatórias associadas as medições devem ser independentes e identicamente distribuídas

24 / 53

Motivação

- Motivação
 - A composição de tempo atualmente para *software* sofisticados trabalham com margem de segurança para lidar com o problema da incerteza de que o pior caso é coberto por seus experimentos
→ **Motivando a busca de métodos estatístico para serem utilizados neste contexto, aumentando a confiabilidade da garantia temporal**
 - Conforme *Ingo Molnar* a diferença de latência é bastante significativa entre amostras de tamanho diferentes
→ **Motivando a busca de métodos para garantir que o tamanho da amostra é bom para gerar boas estimativas de pior caso**
 - Conforme *Nicholas McGuire* o grande problema com os testes RT é que estão olhando para o bom caso, assumem as operações bem sucedidas
→ **Motivando a busca de métodos para a geração dos piores casos**

25 / 53

Trabalhos Relacionados

- [CHONG, Lee Kee et al. (2013)] - *A Statistical Response-Time Analysis of Real-Time Embedded Systems*
 - Cálculo da estimativa do *WCRT* probabilística (*pWCRT*)
 - Considera sistema caixa preta
 - Utiliza de técnicas estatísticas para derivar limite superior do tempo de resposta de tarefas em sistemas embarcado de tempo real
- Assumem que a distribuição do tempo de resposta das tarefas converge para uma distribuição de *Gumbel* - **É necessário uma análise mais cuidadosa**
- Considera tarefas não-bloqueio - **Poderia realizar a análise por partes da tarefa, considerando o bloqueio separado**

26 / 53

Trabalhos Relacionados

- [Lima, George et al. (2016)] - *Extreme Value Theory for Estimating Task Execution Time Bounds: A Careful Look*
 - Apresentam um estudo cuidadoso sobre a aplicabilidade e a eficácia da TVE para derivar limites de tempos de execução de tarefas probabilísticos (*pWCET*)
 - Projetaram uma plataforma embarcada equipada com cache aleatória de tamanhos configuráveis
- Estimativas *pWCET* depende da distribuição dos dados de entrada - **Classificar a sensibilidade da tarefa com relação aos dados de entrada**
- Mostram que é necessário utilizar a GVE, pois ser restrito a uma única família de distribuição pode gerar valores inseguros ou sobrestimado - **Ser cuidadoso no teste de goodness-of-fit**
- Confirmam que randomização de *hardware* favorece a aplicabilidade da TVE, apesar de não garantir que a distribuição dos dados podem ser analisável via TVE - **Investigar métodos para garantir analisabilidade por TVE (modificar a PDF)**

27 / 53

Trabalhos Relacionados

- [Griffin, D; Burns, A. (2010)] - *Realism in Statistical Analysis of Worst Case Execution Times*
 - Considera o uso da TVE para modelar o tempo de execução no pior caso
 - Mostram que estimativa estatística sacrifica o realismo, é necessário tomar medidas para que a segurança do modelo não seja um acidente não intencional deste sacrifício
 - Para garantir a segurança do modelo é necessário provar os pressupostos de estatísticas para TVE ou adaptar a aplicação da análise estatística
 - Adaptação da aplicação requer um conjunto adicional de restrições sobre as medidas utilizadas para gerar o modelo estatístico
- **Realizar uma análise cuidadosa da aplicabilidade da TVE**

28 / 53

Trabalhos Relacionados

- [Zempléni András (2004)] - *Goodness-of-fit in extreme value applications*
 - O teste de *Kolmogorov-Smirnov* (KS) não tem limites livre de distribuição em caso de parâmetros estimados e não são trabalhados para este caso. O método em si não é forte
 - O teste de *Anderson-Darling* é baseado nas estatísticas de teste, não denotando dependência dos parâmetros
- Realizar uma análise cuidadosa do teste de *goodness-of-fit*

29 / 53

Trabalhos Relacionados

- [Hansen, Jeffery et al. (2009)] - *Statistical-based WCET estimation and validation*
 - Apresentam uma abordagem baseada na medição, que produz simultaneamente uma estimativa *WCET*, e uma previsão da probabilidade de que um futuro tempo de execução irá exceder a estimativa
 - Abordagem estatística baseada na TVE para construir um modelo do comportamento da cauda do valor medido do tempo de execução
 - Validaram a abordagem usando um conjunto de dados industriais, composta por mais de 150 componentes incluídos na amostra e quase 200 milhões de tempos de execução da amostra
 - Mostraram que o método baseado em TVE pode produzir estimativas *WCET* para os quais a probabilidade de ultrapassagem é mais previsível e controlável do que utilizando o momento máximo de execução observado
- Como saber se o tamanho da amostra é suficiente para dar boas estimativas? Quando parar de testar?

30 / 53

Trabalhos Relacionados

- [Lu, Yue et al. (2011)] - *A New Way about using Statistical Analysis of Worst-Case Execution Times*
 - Revisitam o problema da utilização de TVE, na análise (*WCET*) de programas em execução em um único processador
 - Propõem um método de análise *WCET* estatístico que consiste de um novo mecanismo de amostragem
 - Propõem um método para a estimativa do *WCET* com uma certa probabilidade previsível de ser ultrapassado
- Realizar uma análise cuidadosa da aplicabilidade da TVE e do teste de *goodness-of-fit* (*Gumbell* e *KS* podem não ser confiáveis)

31 / 53

Trabalhos Relacionados

- [BÜNTE, Sven et al. (2011)]: investigam métrica relativa a segurança que compara vetores de testes no que diz respeito à forma como o comportamento de pior caso observado de partes do programa é exercido
 - É mostrado empiricamente que os critérios de cobertura de código estruturais a partir do domínio de teste funcional pode produzir estimativas *WCET* inseguras
 - Apresentam uma técnica de geração de dados de entrada (*Balanced Path Generation*) que combina as vantagens dos critérios de cobertura avaliados e geração de dados de entrada aleatória
- [BÜNTE, Sven; ZOLDA, Michael; KIRNER, Raimund. (2011)]: apresenta técnica de geração de dados de entrada, combinam verificação do modelo e algoritmos genéticos, a fim de otimizar heurísticamente o conjunto de medições em termos de segurança

32 / 53

Experimento

- Investigação de métodos para teste de tempo de execução e tempo de resposta, que possibilitem obter estimativas de confiabilidade para os requisitos temporais
- Utilizar medição e técnicas estatísticas adequada
- Na medição definir:
 - Caso de teste
 - Procedimento de teste
 - Critério de teste (cobertura, adequação e geração de casos de teste)
- Técnicas estatísticas
 - Classificar a tarefa com relação a sensibilidade aos dados de entrada
 - Identificar quando parar de testar (tamanho da amostra)
 - Obter um limite superior para o tempo de resposta no pior caso dado um grau de confiança (TVE)

33 / 53

Plataforma

- A plataforma escolhida para realização do experimento foi a *BeagleBone*
 - Equipada com um processador Texas Instruments (TI) AM335X
 - Com um núcleo Advanced RISC Machines (ARM) Cortex™-A8
- Medição do *WCRT* - *BeagleBone Black* com Sistema Operacional (SO) Linux
- Medição do *WCET* - *BeagleBone White* possui uma interface de depuração através da qual é possível utilizá-la sem o emprego de SO, obtendo-se assim maior controle sobre seu *hardware*

34 / 53

Análise

- Análise do *WCRT*
 - Medição fim a fim
 - Tarefa **Bsort**: realiza a ordenação de um vetor de 100 elementos utilizando o método *bubble sort*
 - Dados de entrada que acredita-se gerar maior tempo de resposta (vetor ordenado em ordem decrescente)
 - Tamanho das amostras: 10^5 , $5 \cdot 10^5$ e 10^6

35 / 53

Análise Exploratória - *WCRT*

- Resumo da análise exploratória de três amostras - **Bsort**

Tabela: Análise exploratória das amostras

Amostra	Min	Max	Média	Des. Pad	Assim.	Curt.	Tamanho
I	320958	2464959	356043,1835	31636,8700	10,07685	366,7016	10^5
II	320666	4407417	360700,5412	31156,7421	11,5611	734,2644	$5 \cdot 10^5$
III	320625	2465166	361479,0708	30747,2884	7,7994	197,8586	10^6

Tabela: Análise exploratória dos blocos máximos

Amostra	Des. Pad	Assim.	Curt.	Tamanho
I	28432,2492	2,8835	38,3003	2000
II	26836,5635	6,4110	165,5986	10000
III	30996,6080	8,6845	208,9602	20000

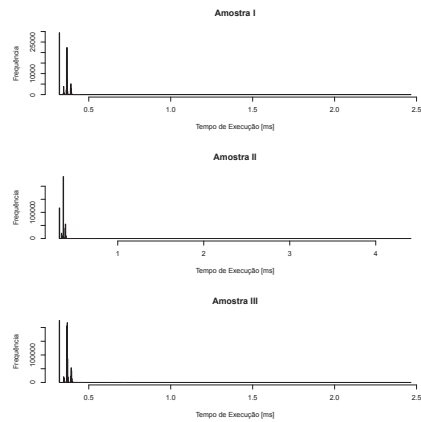
Tabela: Teste de proporção das amostras

Amostra	Média	p-valor	Min	Max	Tamanho
I	356043,2	0,6475	0,6445	0,6504	10^5
II	360700,5	0,5678	0,5664	0,5692	$5 \cdot 10^5$
III	361479,1	0,7328	0,7319	0,7337	10^6

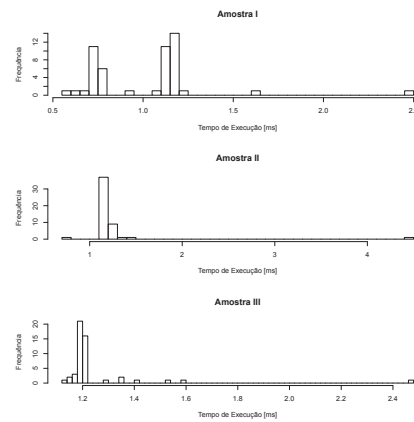
36 / 53

Análise da Tarefa - WCRT

• Amostra



• Bloco máximo



37 / 53

Análise - WCRT

• Considerações da análise do WCRT da tarefa Bsort

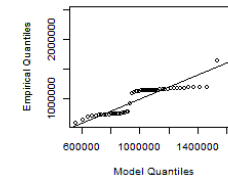
- Difícil garantir a confiabilidade da medição do WCRT executando com Linux
- Atividades do SO tornam o tempo de resposta não-suave
- Existência de picos e vales
- Técnicas estatísticas tornam-se inapropriadas

39 / 53

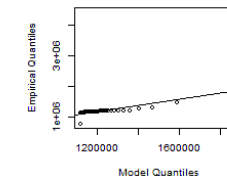
Análise da Tarefa - WCRT

• GEV (L-moments)

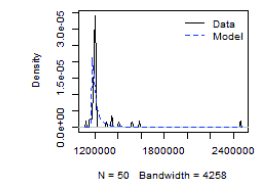
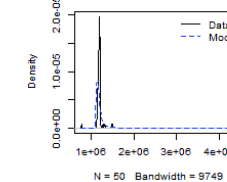
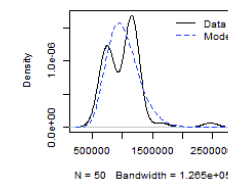
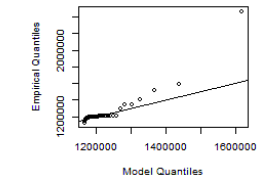
• Amostra I



• Amostra II



• Amostra III



38 / 53

Análise

• Análise do WCET

- Medição fim a fim
- Tarefa Bsort
- Dados de entrada que acredita-se gerar maior tempo de resposta (vetor ordenado em ordem decrescente)
- Tamanho das amostras: 10^3 , $5 \cdot 10^3$ e 10^3

40 / 53

Análise Exploratória - WCET

- Resumo da análise exploratória de três amostra - **Bsort**

Tabela: Análise exploratória das amostras

Amostra	Min	Max	Média	Des. Pad	Assim.	Curt.	Tamanho
I	813207	813747	813481	95,4925	-0,0202	-0,37602	10 ³
II	813131	813749	813390,2964	82,8431	0,6995	0,6340	5.10 ³
III	813166	813802	813491,6381	85,0683	-0,0041	-0,0897	10 ⁴

Tabela: Análise exploratória dos blocos máximos

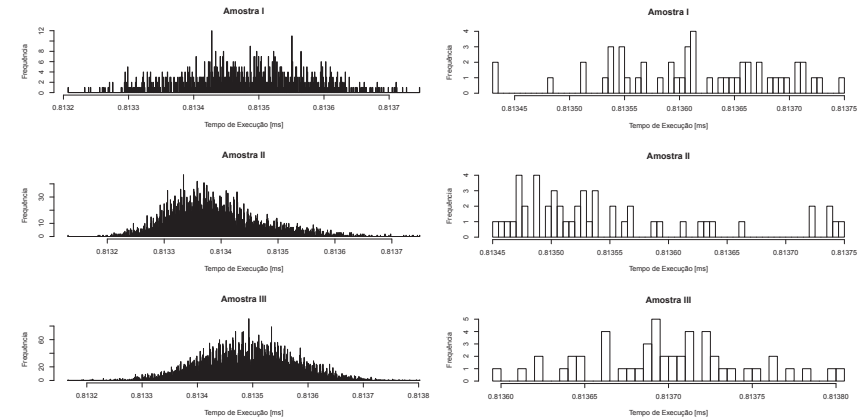
Amostra	Des. Pad	Assim.	Curt.	Tamanho
I	13,9994	-0,7785	2,1605	20
II	12,0906	0,3184	0,9485	100
III	8,1548	-0,1988	-0,2623	200

Tabela: Teste de proporção das amostras

Amostra	Média	p-valor	Min	Max	Tamanho
I	813481	0,501	0,4696	0,5324	10 ³
II	813390,3	0,4396	0,4258	0,4535	5.10 ³
III	813491,6	0,5029	0,4930	0,5127	10 ⁴

Análise da Tarefa - WCET

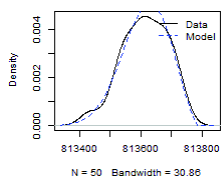
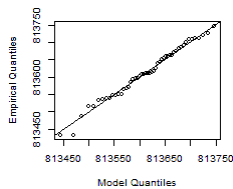
- Amostra
- Bloco máximo



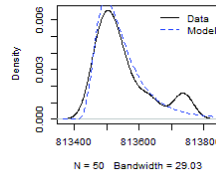
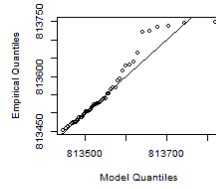
Análise da Tarefa - WCET

- GEV (L-moments)

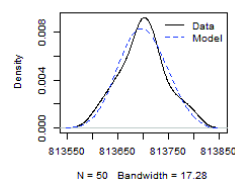
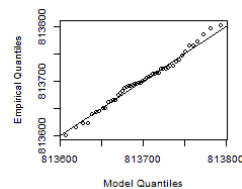
- Amostra I



- Amostra II



- Amostra III



Análise - WCET

- Considerações da análise do WCET da tarefa **Bsort**

- Os tempos de execuções são mais suaves
- Existência de picos e vales menos frequentes
- Técnicas estatísticas podem prover informações necessárias de garantia temporal, quando aplicadas corretamente

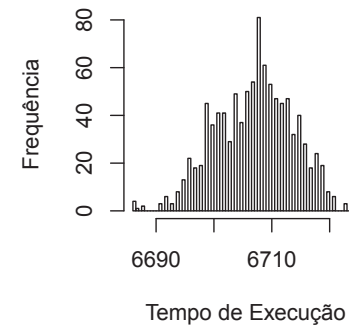
Análise - WCET

- Análise do WCET
 - Medição fim a fim
 - Tarefas:
 - **CNT**: contagens de números não negativos em uma matriz
 - **ST**: realiza cálculos estatísticos para duas matrizes
 - Tamanho das amostras: 10^3 , 10^4 e 10^6
 - Dados de entrada gerados aleatórios em faixas diferentes
 - Teste *Kolmogorov-Smirnov* (KS): sensibilidade aos dados de entrada
 - No teste KS - Tem como hipótese nula que as distribuições são iguais
 - Nível de significância utilizado 1% ($\alpha = 0,01$)
 - A hipótese nula não podem ser rejeitadas se o p-value obtido no teste for maior que o nível de significância
 - GEV - Estimativa do WCET

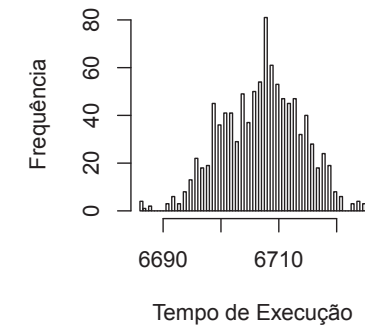
45/53

Sensibilidade aos dados de entrada

Histograma CNT - 1K



Histograma CNT - 1K

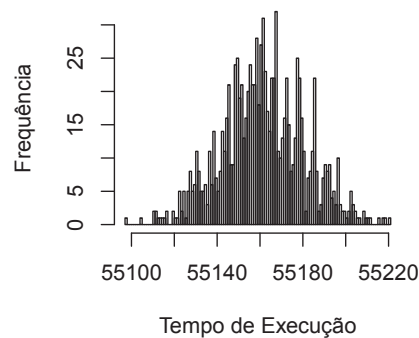


- KS: Obteve-se supremo $D = 0,0299$ e $p\text{-value} = 0,7591$
- Não há evidências de diferença entre as duas distribuições ao nível de significância de 1%, portanto CNT não é sensível aos dados de entrada

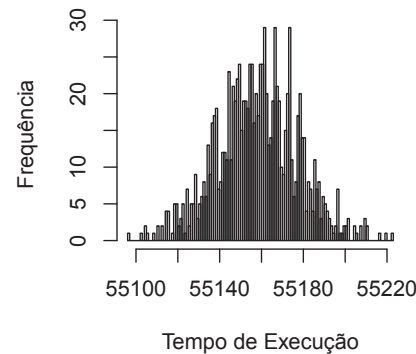
46/53

Sensibilidade aos dados de entrada

Histograma ST - 1K



Histograma ST - 1K



- KS: Obteve-se supremo $D = 0,0739$ e $p\text{-value} = 0,00837$
- Há evidências de diferença entre as duas distribuições ao nível de significância de 1%, portanto ST é sensível aos dados de entrada

47/53

Quando parar de testar

- Dobrar o tamanho da amostra até não obter mais nenhum ganho na análise
- Utilizar o tamanho da amostra somado com um valor delta, iterando até quando for necessário
- Analisar a dinâmica do sistema

... é suficiente?

48/53

Quando parar de testar

Tabela: Análise exploratória das amostras - CNT

Amostra	Min	Max	Média	Des. Pad	Assim.	Curt.	Tamanho
CNT_1K	6686	6725	6707	6,8320	-0,1598	-0,1917	10 ³
CNT_10K	6680	6726	6706,5332	7,1549	-0,2586	-0,0571	10 ⁴

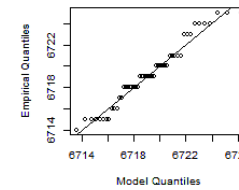
Tabela: Análise exploratória das amostras - ST

Amostra	Min	Max	Média	Des. Pad	Assim.	Curt.	Tamanho
ST_1K	55097	55221	55162	19,2297	0,0289	0,02392	10 ³
ST_10K	55093	55254	55159,4737	19,0890	0,0157	-0,0139	10 ⁴
ST_100K	55080	55258	55161,9969	19,6518	0,0635	0,0266	10 ⁵

- Diferença pequena na análise exploratória das amostras de tamanhos 10³ e 10⁴

Estimativa do WCET dado um grau de confiança

- GEV (Lmoments) - CNT



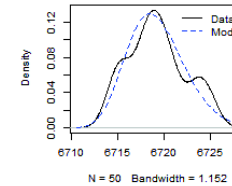
- Amostra: 10³
- Número de bloco máximo: 50

Tabela: Amostra

Tamanho	Max
10 ⁴	6726

Tabela: Estimativa do WCET

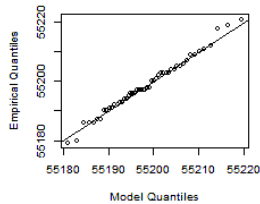
C(10 ^x)	WCET
10 ⁻²	6727.2629
10 ⁻³	6729.8559
10 ⁻⁴	6731.6293
10 ⁻⁵	6732.8445
10 ⁻⁸	6732.9397
10 ⁻¹⁶	6735.4495



- Há evidências que o tamanho da amostra utilizada seja suficiente, porém, observa-se falta de aderência dos dados à distribuição

Estimativa do WCET dado um grau de confiança

- GEV (Lmoments) - ST



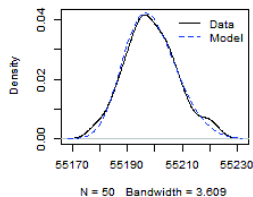
- Amostra: 10³
- Número de bloco máximo: 50

Tabela: Amostra

Tamanho	Max
10 ⁴	55254
10 ⁵	55258

Tabela: Estimativa do WCET

C(10 ^x)	WCET
10 ⁻²	55222.2369
10 ⁻³	55229.2280
10 ⁻⁴	55233.7482
10 ⁻⁵	55236.6768
10 ⁻⁸	55236.8988
10 ⁻¹⁶	55242.0221



- Tamanho da amostra utilizada não é suficiente

Considerações Finais

- Obter estimativas de confiabilidade para os requisitos temporais de *software* sofisticado, onde geralmente não é possível a modelagem formal do *hardware*, é difícil
- Garantia do WCET por teste de *software* e técnicas estatísticas
- Usar métodos analíticos para obter o WCRT
- Métodos para geração de caso de teste e critério de cobertura são importantes
- Testes estatísticos propostos até o momento são inexatos
- Análise estatística rigorosa
 - Amostra grande o suficiente
 - Aplicabilidade correta dos métodos estatísticos
 - Aderência à uma distribuição (bom ajuste dos dados ao modelo)
 - Garantia de confiança no valor obtido e a probabilidade do tempo de resposta a ser excedido

