

# Índice de Carga

Paulo Sérgio Lopes de Souza

Grupo de Sist. Distribuídos e Prog. Concorrente

Orientador: Prof. Dr. Marcos José Santana

março / 1999

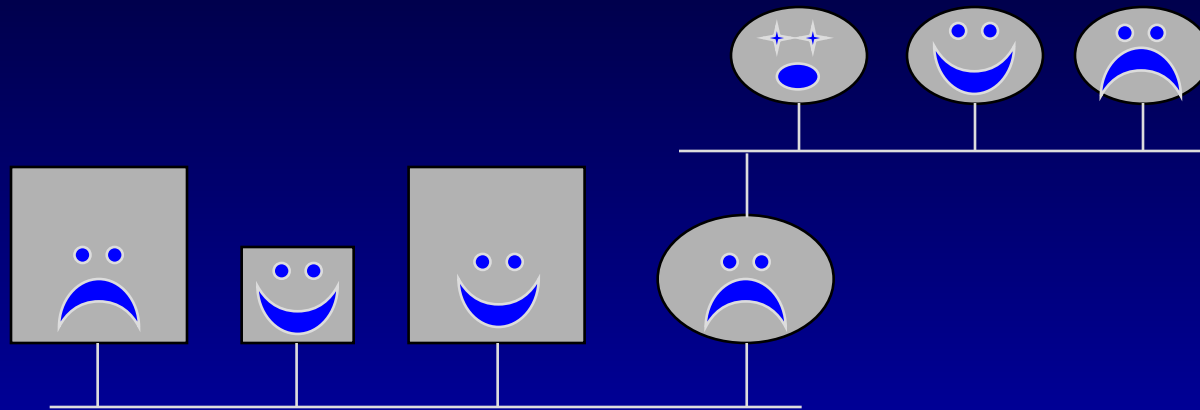


# Conteúdo

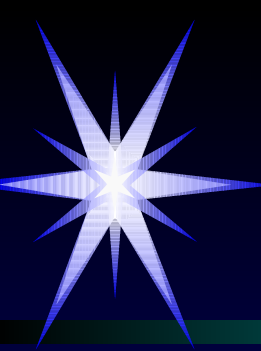
---

- Revisão Bibliográfica
- Problemas Encontrados na DPWP
- Propostas para Possíveis Soluções

# Índice de Carga



- Como determinar a carga atual do sistema?
  - Quais fatores determinam a carga?
  - Objetivos: aumentar eficácia e a satisfação
    - O significado de eficácia e satisfação pode variar.



# Índice de Carga

- O índice de carga quantifica a carga do sistema
  - é um valor não negativo que varia proporcionalmente à carga atual do sistema;
  - sinônimos (nomenclatura variada):
    - descritor da carga de trabalho
    - medida da carga de trabalho
  - finalidade: indicar comportamento futuro com base no comportamento atual/passado.

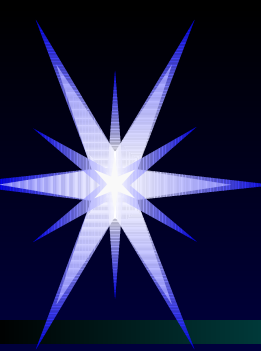


# Índice de Carga

- Qualidades/propriedades de um índice de carga:  
critérios dependem do objetivo do escalonador

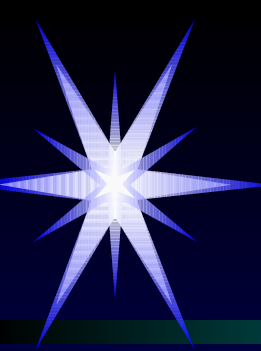
de um modo geral:

- 1) capaz de refletir a carga atual de um host;
- 2) capaz de prever a carga em um futuro próximo
- 3) ser estável, ignorar flutuações de carga (noise)
- 4) estar relacionado com o índice de desempenho do escalonador, para facilitar análises de desempenho.



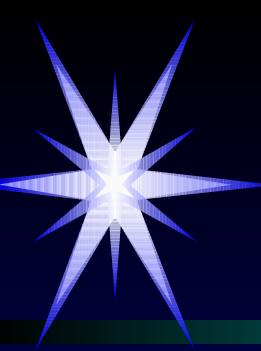
# Índice de Carga

- Classificação dos índices:
  - baseados no comprimento da fila do recurso
  - baseados no % de utilização do recurso
  - específicos ou genéricos
  - exemplos:
    - comprimento da fila de CPU
    - % de utilização da CPU
    - % utilização CPU + % ocupação da memória



# Índice de Carga

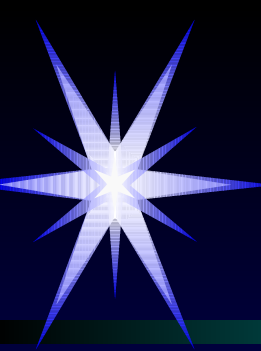
- Como um índice de carga é obtido e calculado?
  - obtido diretamente no kernel ou por uma aplicação
  - cálculo simples, pouco overhead (custo/benefício)
  - valores instantâneos x valores médios com “refinamentos”
- Fatores que afetam o índice de carga:
  - são vários (hardware, multiusuário, time-sharing, ...)
  - destaque:
    - classes de software e/ou carga de trabalho.



# Índice de Carga

- Classes de Software:
  - há diversas classificações:(batch, interativas,...)
  - aqui, para simplificar:
    - CPU-bound, I/O-bound, balanceadas
  - afeta diretamente a escolha do índice
    - vários estudos feitos (predizer o uso de recursos)
- mais informação=>maior qualidade escalonamento
- essa informação nem sempre está disponível:
  - dificuldade em obtê-la antes ou durante a execução





# Índice de Carga

- Carga de Trabalho
  - o que é a carga trabalho ?
    - seleção das aplicações utilizadas no sistema
    - caracterização da demanda pelos recursos
  - essencial para estudar índices de carga
  - pode ser:
    - real / natural ou sintética / artificial
  - background workload:
    - multiusuário



# Índice de Carga

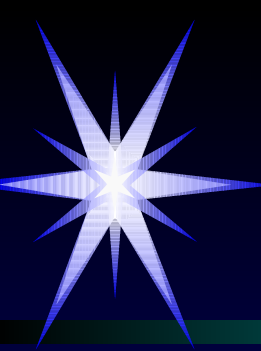
- Estudo de Ferrari e Zhou (1988)
  - considerou carga de trabalho de vários meses:
    - VAX-11/780 com Unix 4.3BSD
    - selecionou 30 aplicações mais usadas (tipo e frequência)
    - construiu “job scripts” com seqüências desses comandos
      - carga de trabalho sintética ou artificial
    - três níveis de scripts: (L)light, (M)moderate e (H)heavy
    - porém estudo foi feito em 6 Sun-2 workstations
    - Escalonador no SO, transparente, sem migração



# Índice de Carga

Caracterização dos níveis da carga de trabalho

<b>Tipo</b>	<b>Utilização da CPU</b>	<b>Comprimento médio Fila CPU</b>
Light (L)	30 – 45 %	0.3 – 0.7
Moderate(M)	60 – 70 %	1.0 – 1.8
Heavy (H)	70 – 85 %	1.8 – 3.0



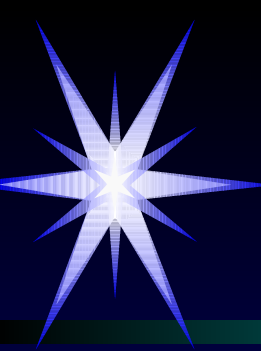
# Índice de Carga

- Fatores considerados no Estudo de Ferrari:
  - 1) índices de carga estudados:
    - comprimento da fila de CPU atual
    - comprimento médio da fila de CPU (exponencial)
    - soma das filas (CPU, paginação/swapping I/O e memória)
    - média de utilização de CPU em um período (T).

valores obtidos diretamente no kernel

tamanho instantâneo de cada fila:

média de 1s, coletando valores a cada 10ms



# Índice de Carga

- 2) Intervalo T para determinação da média:
  - usado para o refinamento exponencial
    - comprimento das filas dos recursos
    - utilização média dos recursos
- 3) Carga de Trabalho:
  - carga de trabalho afeta o índice escolhido:
    - 6 Sun-2's => 2H, 2M, 2L e 6M
- 4) Intervalo P para atualização da informação:
  - importante em políticas periódicas (local x global)
  - curto : maior overhead
  - longo: informação antiga, < desempenho, instabilidade



# Índice de Carga

Carga de trabalho(2H,2M,2L), P=10s

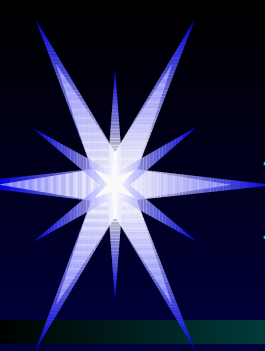
Índice de Carga	T. de Resposta	Ganho %
sem escalonamento	53.3 +/- 0.83	---
fl de CPU	35.0 +/- 0.68	34.4
média fl CPU (T = 1s)	33.8 +/- 0.65	36.6
média fl CPU (T= 4 s)	33.1 +/- 0.39	37.9
média fl CPU+I/O+MEM (T = 4s)	32.2 +/- 0.45	39.6
média fl CPU (T = 20s)	37.0 +/- 1.20	30.6
média fl CPU+I/O+MEM (T=20s)	35.6 +/- 0.12	33.3
média fl CPU (T=60s)	39.7 +/- 1.69	25.5
média fl CPU+I/O+MEM (T = 60s)	40.0 +/- 0.56	25.0
carga média UNIX (T = 60s)	37.2 +/- 0.85	30.2
utilização CPU (T = 10 s)	38.5 +/- 2.10	27.8
utilização CPU (T = 60s)	42.9 +/- 1.36	19.5



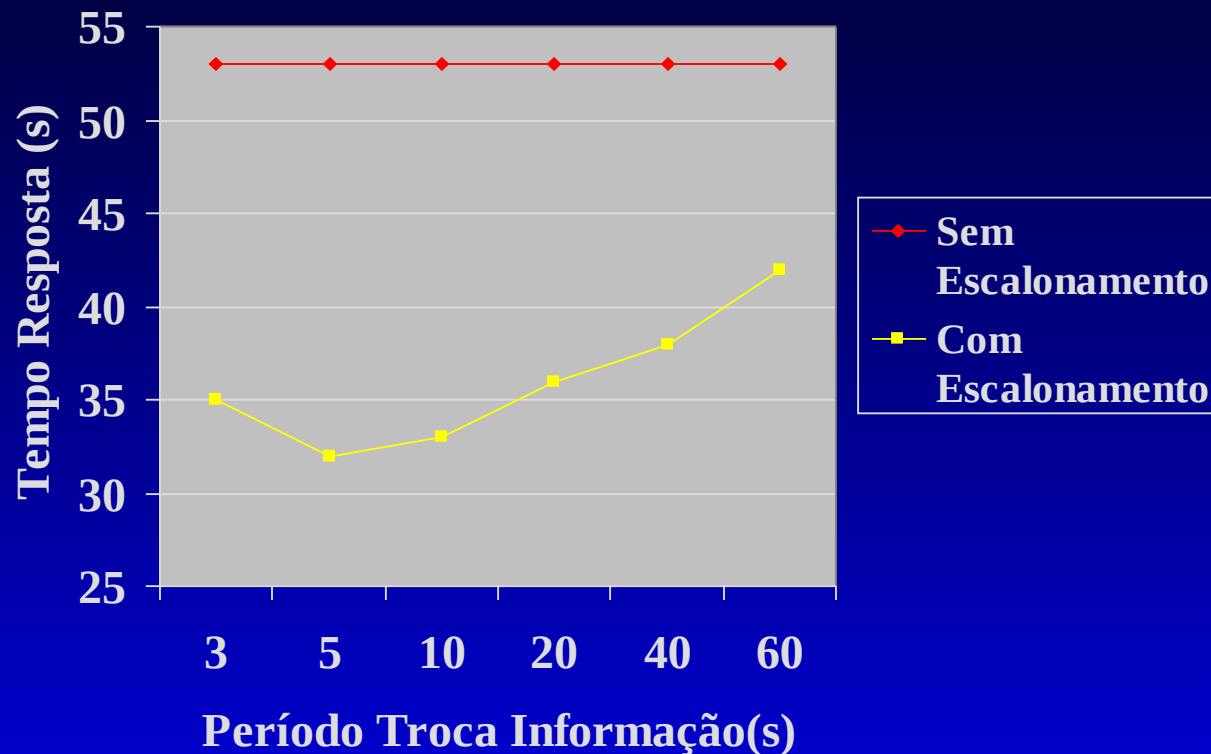
# Índice de Carga

Carga de trabalho(6M), P=10s

Índice de Carga	T. de Resposta	Ganho %
sem escalonamento	49.5 +/- 0.27	---
fl de CPU	42.3 +/- 0.79	14.5
média fl CPU (T= 4 s)	39.9 +/- 0.63	19.4
média fl CPU+I/O+MEM (T = 4s)	36.5 +/- 0.91	26.3
média fl CPU+I/O+MEM (T=20s)	45.2 +/- 0.89	8.7
média fl CPU+I/O+MEM (T = 60s)	47.1 +/- 1.34	4.9
carga média UNIX (T = 60s)	47.9 +/- 1.12	3.2
utilização CPU (T = 10 s)	44.0 +/- 1.97	11.1
utilização CPU (T = 60s)	48.6 +/- 1.34	1.8



# Índice de Carga



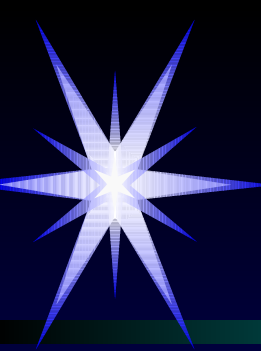
Tempo de resposta sob vários períodos P  
Carga de trabalho: 2H,2M,2L  
Índice: média fila CPU+I/O+MEM T=4s





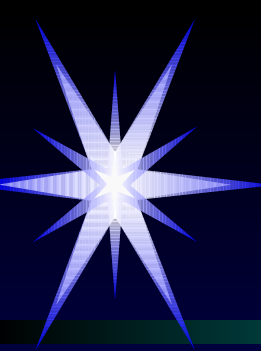
# Índice de Carga

- Fechando...
  - comprovada a importância do índice de carga
    - 2H,2M,2L: 19,5 - 39,6%      6M: 1,8 - 26,3%
  - índices baseados no comprimento da fila são melhores
    - baseados na média são melhores (evitam oscilações)
    - intervalo T depende de quanto a carga varia
    - índice específico (ex: fila CPU) tende a ser melhor
      - união de índices não mudou resultado significativamente
      - depende da carga de trabalho e de quanto se sabe sobre ela
  - valores vindos do kernel são mais precisos e melhoram o desempenho



# Índice de Carga

- Fechando...
  - cargas de trabalho diferentes não alteraram resultado
    - fila CPU+I/O+MEM-4s e Utilização de CPU 60s  
melhor e pior nos dois casos comparados
  - intervalo P depende:
    - custos comunicação e da carga de trabalho
    - informação mais atualizada X overhead
  - outros estudos:
    - possuem diferenças (índices e procedimentos adotados)
    - baseados nas filas dos recursos são melhores



# Índice de Carga

- questões em aberto:
  - heterogeneidade configuracional
  - heterogeneidade arquitetural
  - limites de saturação
    - como generalizar a determinação da saturação de um host?
- falhas na comparação:
  - limitou as classes de software
    - estudar o relacionamento efetivo com as classes de software
  - usou apenas o tempo de resposta médio como objetivo
    - estudar o relacionamento com o índice de desempenho (métrica)