

# Desempenho de Serviços de Comércio Eletrônico

# Sumário

1. Introdução e Conceitos
2. Metodologia para Avaliação de Desempenho
3. Métricas para Avaliação de Desempenho
4. Técnicas de Avaliação de Desempenho
5. Avaliação Experimental de Desempenho
6. Caracterização da Carga de Trabalho
7. Avaliação de Desempenho no Modelo Café
8. Exemplo: Avaliação de Desempenho da Livraria Jambreiro

# Introdução e Conceitos

- O que é avaliar desempenho
  - avaliar a execução de um trabalho de forma quantitativa, atribuindo valores numéricos para métricas de desempenho escolhidas
- Entendimento de desempenho
  - estudo do desempenho de um sistema com o objetivo de identificar fatores e causas que determinam seu desempenho
- Graus de entendimento de desempenho
  - quais recursos são utilizados
  - como os recursos são utilizados
  - quais são os efeitos de alterações do sistema no desempenho
  - quanto o desempenho varia quando são alterados os parâmetros do sistema

# Metodologia para Avaliação de Desempenho

- Definição de objetivos
- Seleção de métricas de interesse
- Seleção da técnica de avaliação
- Caracterização da carga de trabalho
- Aplicação da técnica de avaliação
- Análise e interpretação dos resultados

# Métricas para Avaliação de Desempenho

- Desempenho comercial: exemplos de métricas
  - volume de vendas
  - taxas de conversão
- Desempenho computacional: exemplos de métricas
  - número de requisições atendidas simultaneamente
  - tempo de resposta para uma operação específica

# Métricas de Desempenho Computacional

- Quantificam a utilização de recursos e a qualidade do serviço oferecido pelo servidor
  - Gerais: quantificam o desempenho do servidor como um todo
  - Especializadas: quantificam o desempenho de um componente específico

# Métricas de Desempenho Computacional

- Taxa de Serviço:
  - frequência de execução de operações
- Tempo de resposta:
  - tempo necessário para executar uma tarefa
- Questões:
  - variância entre serviços
  - custos dependentes da entrada

# Desempenho Computacional

## Fatores

- Heterogeneidade
  - componentes diversificados
- Multiprogramação
  - múltiplas requisições simultâneas
- Diversidade da carga de trabalho
  - comportamento do cliente e custo do serviço varia
- Custo de monitoração
  - minimizar a intrusão



# Métricas de Desempenho Comercial

- Específicas do negócio
- Devem estar em conformidade com o projeto do servidor
- Exemplos
  - Vendas por segundo
  - Leilões criados por hora

# Técnicas de Avaliação de Desempenho

- Modelagem analítica
  - o sistema é modelado como uma rede de filas
  - os resultados de desempenho são analíticos
- Simulação
  - o sistema é modelado por um simulador
  - há compromisso entre nível de detalhe e custo de execução
- Experimentação
  - é a avaliação de desempenho do sistema real
  - os eventos do sistema são monitorados

# Avaliação Experimental de Desempenho

- Perfis de Desempenho

- são importantes na avaliação de sistemas dinâmicos complexos
- definem o desempenho do servidor sob várias perspectivas
- descrevem numericamente características importantes da execução de um ou mais processos
- são obtidos através de **monitoração**

- Dimensões

- processos
- segmentos de código
- categorias de processamento

# Monitoração de Desempenho

- Dificuldades:

- Estado do servidor como um todo é distribuído e não pode ser registrado facilmente
- Valores das medidas de desempenho são probabiísticos
- Experimentos podem não ser reproduzíveis

# Monitoração de Desempenho

## Escopo

- Tipos:
  - Locais: sumarizam desempenho de um processo ou processador
  - Globais: sumarizam o comportamento geral do servidor, incluindo as interações
- Estratégias:
  - Contagem
  - Amostragem
  - Registro

# Monitoração de Desempenho Instrumentação

- Compromisso: nível de detalhe do perfil
  - quantidade de informação X intrusão
- Estratégias
  - hardware
  - software
  - híbrido

# Monitoração de Desempenho

## Dimensões

- Localização
- Categoria
  - ortogonalidade
  - completeza
  - significância
- Tempo

# Exemplo: O Servidor Pampulha

- Servidor WWW
  - objetivo: ilustrar conceitos de avaliação de desempenho experimental
  - limitação: atende somente uma requisição por vez



# Servidor Pampulha

```
while (ativo){
    if ((nfd = accept(fd,&cliente,&tam)) < 0){
        perror("accept:");
        continue;
    }
    if (le_req(nfd,url) <= 0){
        close(nfd);
        printf("nao consegui ler req\n");
        continue;
    }
    obtem_doc((char*)url,(char*)buf,&buftam);
    escrito=escreve_resposta(nfd,buf,buftam);
    if (escrito != buftam)
        printf("escrevi apenas %d de %d bytes\n", escrito,buftam);
    close(nfd);
}
```

# Aquisição de Perfis: prof

%	cum.	self		self	total	
time	seconds	seconds	calls	us/call	us/call	name
78.86	9.36	9.36	10000	935.41	935.41	obtem_doc
11.54	10.73	1.37	10000	137.03	137.03	le_req
9.27	11.83	1.10	10000	110.35	110.35	escreve_resposta
0.33	11.87	0.04				main

# Aquisição de Perfis: Instrumentação

```
gettimeofday(&t1,(void*)0);  
obtem_doc((char*)url,(char*)buf,&buftam);  
gettimeofday(&t2,(void*)0);
```

# Aquisição de Perfis: Instrumentação

```
d.tv_sec = t2.tv_sec - t1.tv_sec;  
d.tv_usec = t2.tv_usec - t1.tv_usec;  
if (d.tv_usec < 0){  
    d.tv_sec --;  
    d.tv_usec += 1000000;  
} else if (d.tv_usec > 1000000){  
    d.tv_sec += d.tv_usec/1000000;  
    d.tv_usec %= 1000000;  
}
```

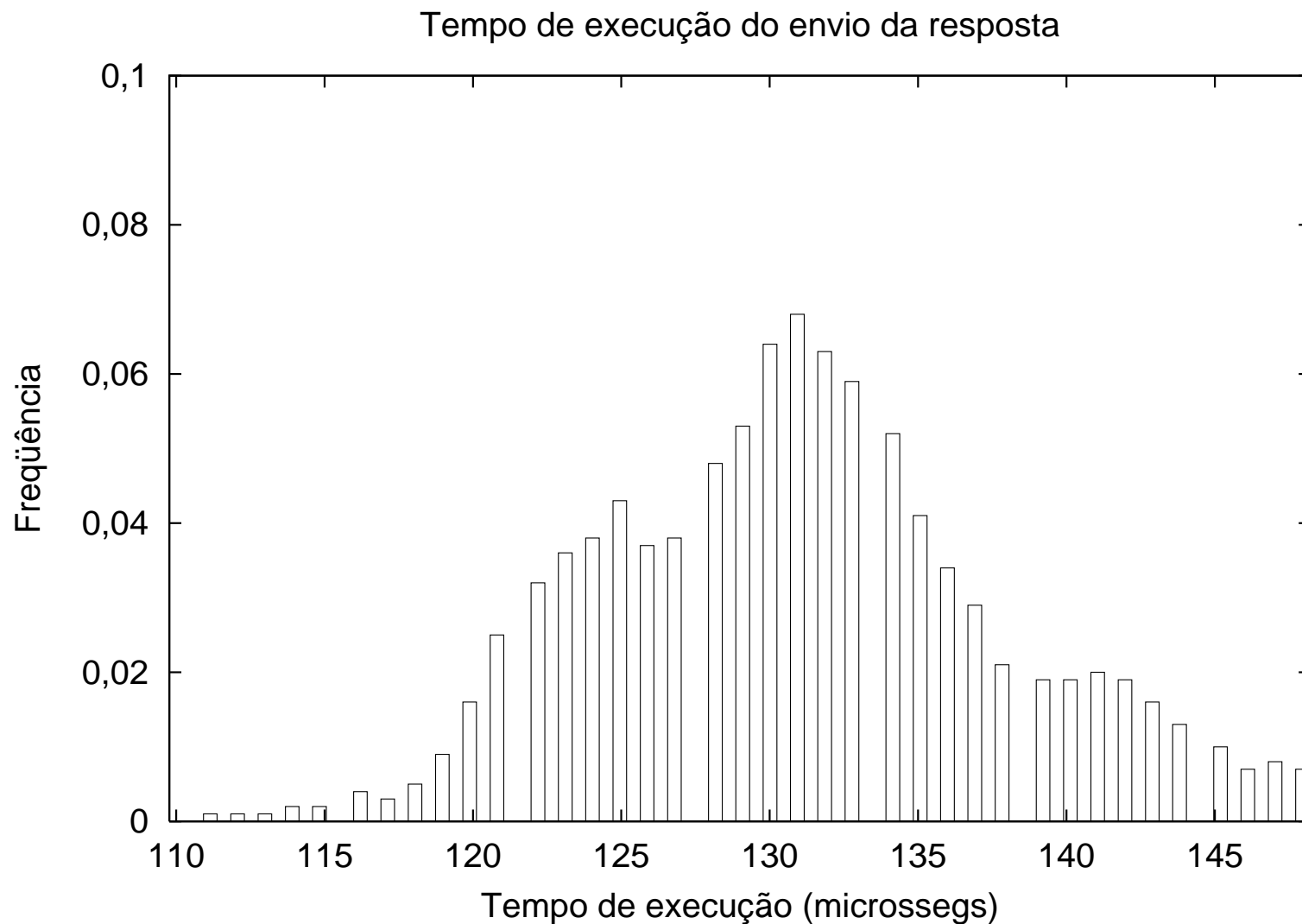
# Aquisição de Perfis: Instrumentação

<i>Passo</i>	<i>Tempo (microssecs)</i>
Recebimento de requisição	169,219
Obtenção de documento	994,091
Envio de resposta	132,604

# Aquisição de Perfis: Instrumentação

```
gettimeofday(&t1,(void*)0);  
fprintf(log,"Inicio: %d.%,6d\n",t1.tv_sec,t1.tv_usec);  
obtem_doc((char*)url,(char*)buf,&buftam);  
gettimeofday(&t2,(void*)0);  
fprintf(log,"Fim: %d.%,6d\n",t2.tv_sec,t2.tv_usec);
```

# Instrumentação: Avaliação de Resultados



# Caracterização da Carga de Trabalho

- Objetivos da caracterização de carga
  - quantificar aspectos da interação entre cliente e servidor que influenciam o desempenho
  - definir uma representação ou modelo para a carga
- Características fundamentais da carga: servidor WWW
  - processo de chegada das requisições
  - características específicas das requisições
    - tamanho da requisição
    - localidade de referência



# Caracterização da Carga de Trabalho

- Servidor de comércio eletrônico: baseada em sessões
  - seqüência de requisições de um cliente
- Parâmetros
  - Frequência das sessões
  - Tamanho das sessões
  - Composição das sessões
  - Frequência das requisições
  - Parâmetros das requisições

# Café: Avaliação de Desempenho

- Baseada em experimentação sobre plataformas reais
- Utilização
  - Implementação: Atende aos requisitos de projeto?
  - Produção: Desempenho é satisfatório?

Quais os fatores determinantes do desempenho?

# Café: Avaliação de Desempenho

- Nível Funcional
  - Métricas de Desempenho
  - Estratégia de caracterização de carga
- Nível de Execução
  - Estratégia de monitoração

# Especificação de Métricas de Desempenho

- Descrição
  - define informações quantificadas e sua utilização
- Tipo: geral ou especializada
- Unidade de medição
- Eventos de interesse:
  - marcos para medições de desempenho

# Métrica: Taxa de Serviço

- Descrição

- mede o número de operações completadas por unidade de tempo

- Exemplos

- servidores Web: número de requisições HTTP por segundo
  - servidores C.E.: número de serviços concluídos por segundo

# Especificação da Métrica Taxa de Serviço

- Servidor Web Pampulha
- Pampulha: taxa de serviço
  - Descrição
    - quantifica o número de requisições atendidas por unidade de tempo
    - permite avaliar a capacidade de atendimento do servidor
  - Tipo: global
  - Unidade de medição: requisições atendidas por segundo
  - Eventos de interesse: execução da função `escreve_resposta`

# Métrica: Tempo de Resposta

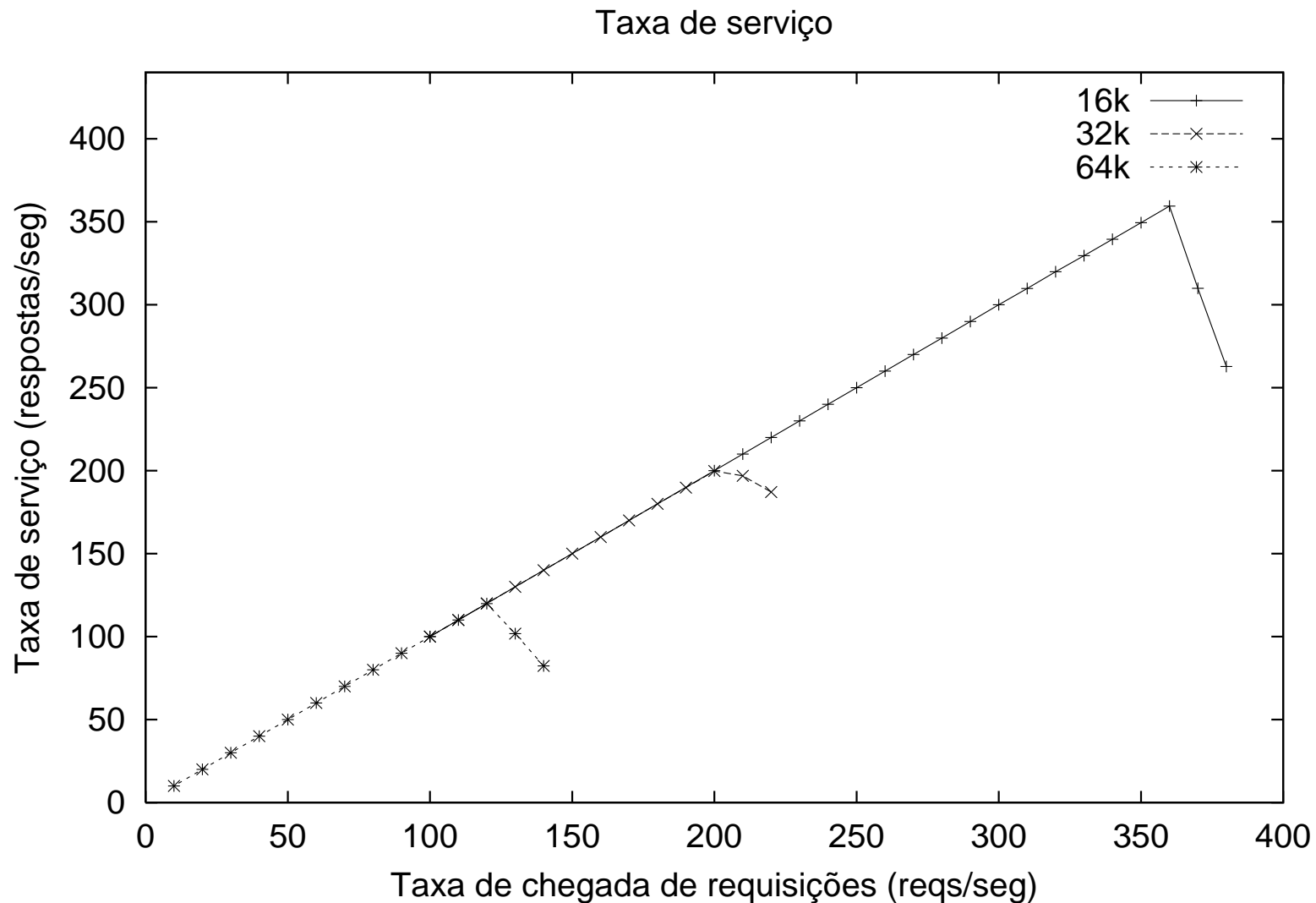
- Descrição
  - tempo decorrido entre o início e o fim da execução de um tarefa
- Característica: composto por diversos tempos
  - tempo de computação
  - tempo de espera ou tempo nas filas
  - tempo de comunicação

# Especificação da Métrica Tempo de Resposta

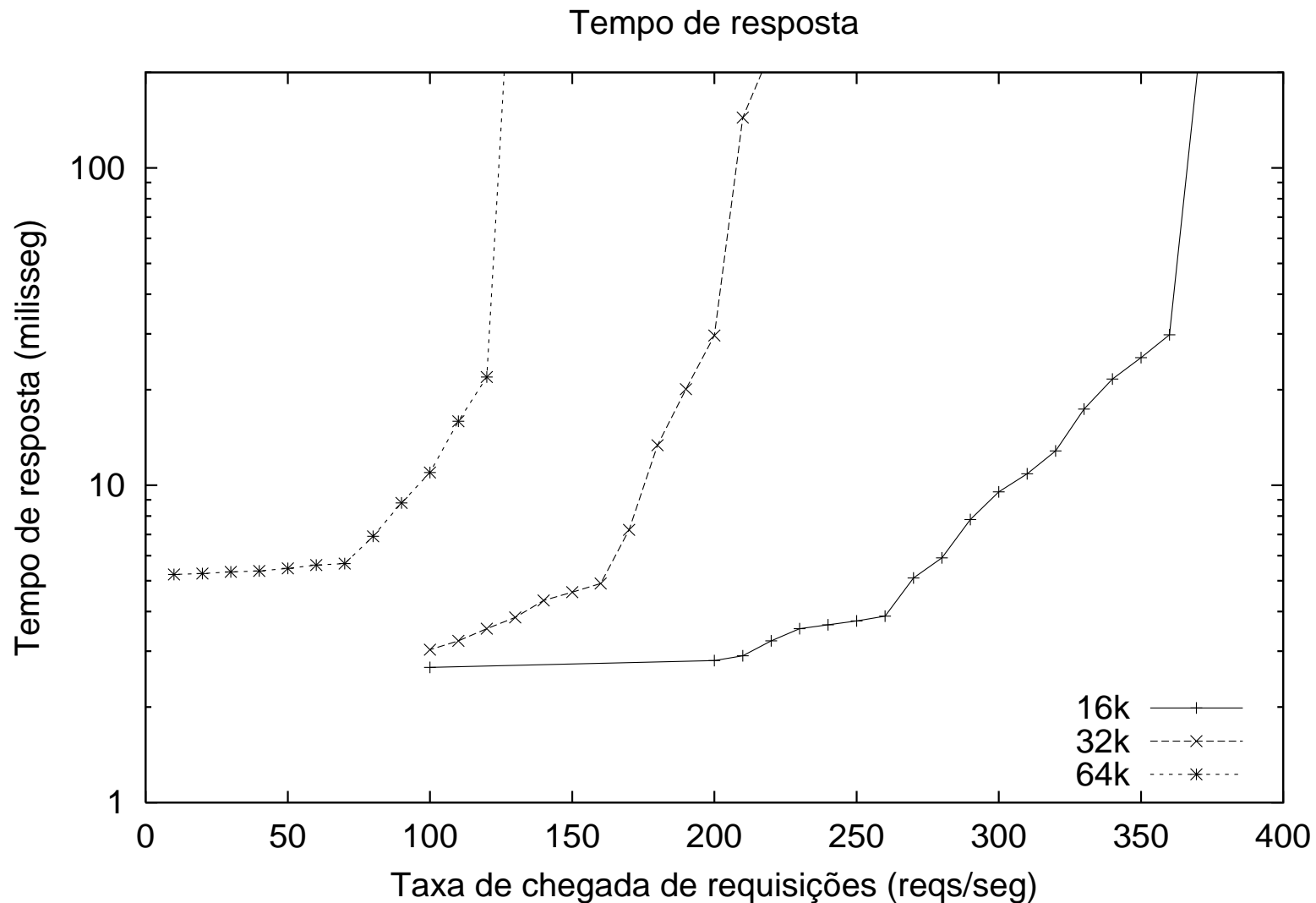
- Servidor Web Pampulha
- Pampulha: tempo de resposta
  - Descrição
    - quantifica a latência de resposta do servidor para uma requisição
    - auxilia o controle de qualidade de serviço e a identificação de degradação de desempenho
  - Tipo: global
  - Unidade de medição: segundos
  - Eventos de interesse: início da recepção de uma requisição e conclusão do envio da resposta



# Taxa de serviço do servidor Pampulha



# Tempo de resposta do servidor Pampulha

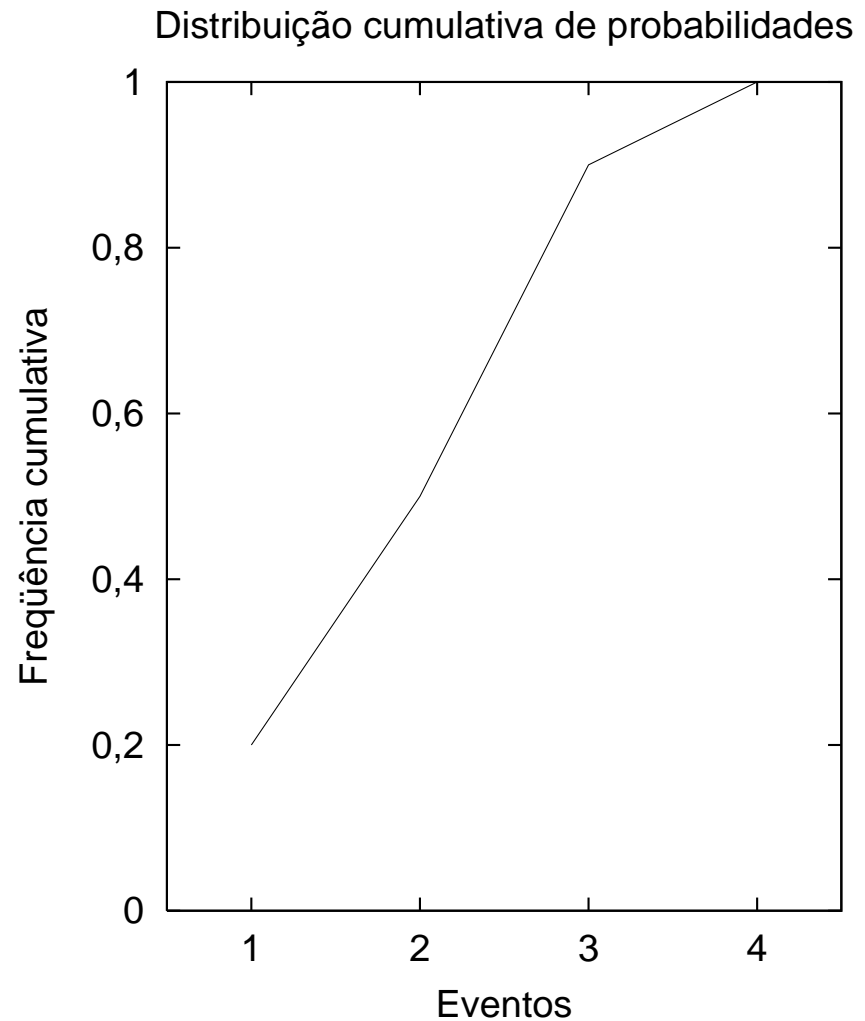
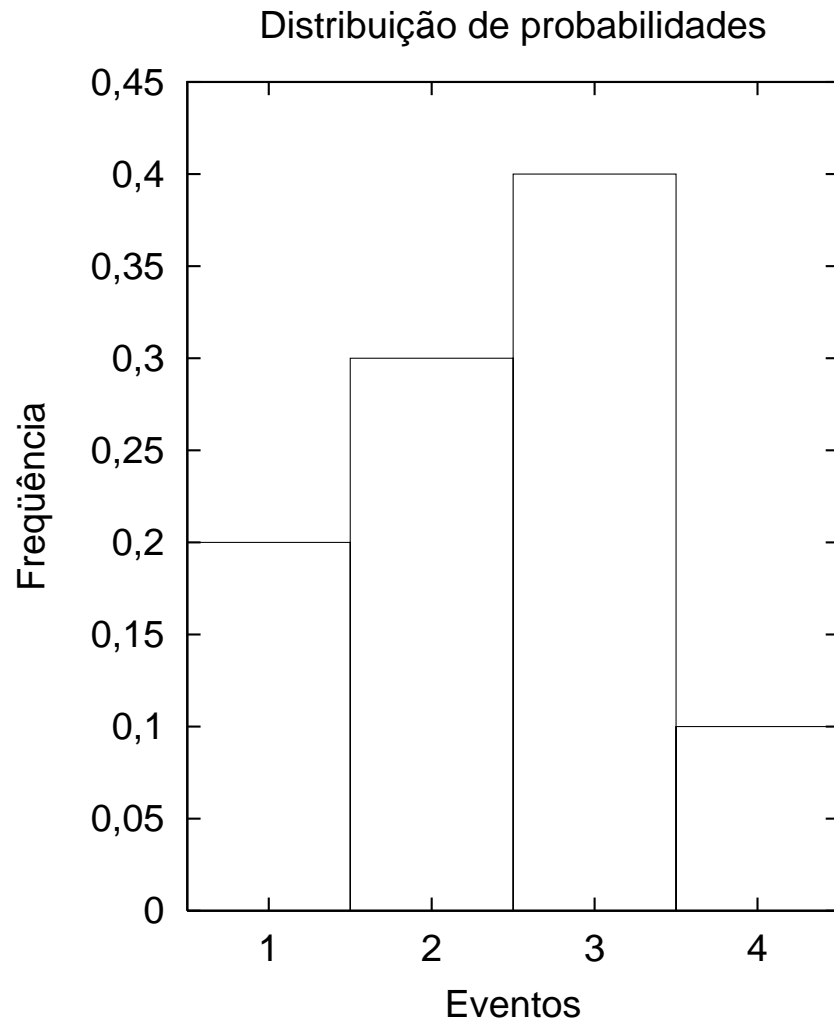


# Café: Caracterização da Carga de Trabalho

- Parâmetros

- Frequência das sessões
  - especifica o intervalo entre duas sessões consecutivas
- Composição das sessões
  - especifica os serviços acessados em uma sessão
- Tamanho das sessões
  - especifica o n'umero de requisições em cada sessão
- Frequência das requisições
  - especifica a taxa de chegada
- Parâmetros das requisições
  - especifica os custos associados a cada requisição

# Distribuição cumulativa de probabilidades



# Distribuição cumulativa de probabilidades

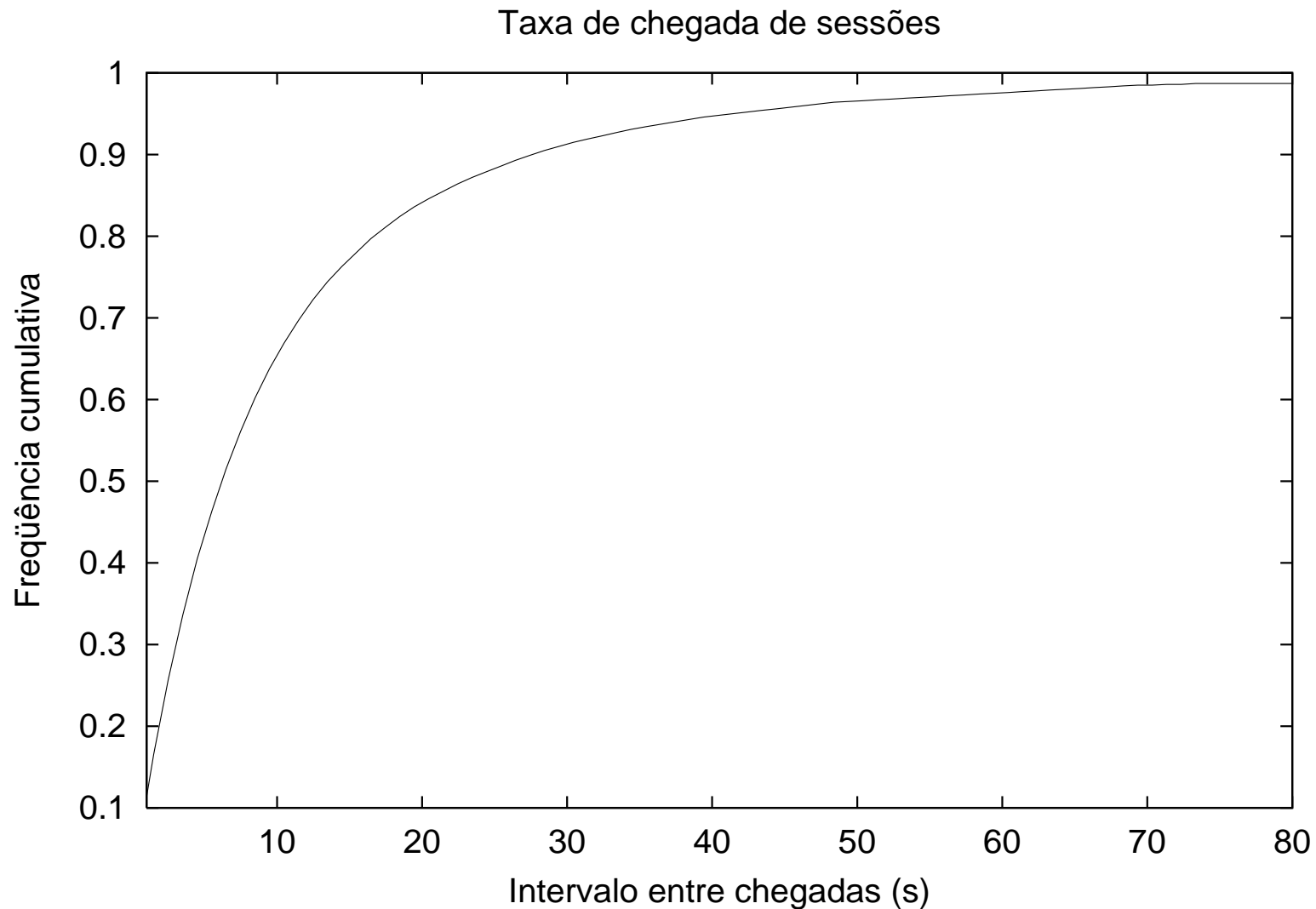
```
1  gera_cdf(eventos)
2      for (i=0; i < |eventos|; i++)
3          freq[eventos[i]] ++
4      ordena(freq)
5      distcum[0] = freq[0]
6      for (i=1; i < |freq|; i++)
7          distcum[i] = distcum[i-1] + freq[i]
8      return distcum
```

# Caracterização de carga de uma loja virtual

## Dados de entrada

- Data e hora do acesso
- Identificador de sessão
- Identificador do tipo de requisição
- Parâmetros da requisição, se for o caso.

# Processo de Chegada das Sessões



# Composição das sessões: Sessão Comprador

<i>Requisição Origem</i>	<i>Requisição Destino</i>						
	Entrar	Iniciar	Buscar	Navegar	Ver	Alocar	Pagar
Entrar	0%	100%	0%	0%	0%	0%	0%
Iniciar	0%	0%	100%	0%	0%	0%	0%
Buscar	0%	0%	33%	0%	67%	0%	0%
Navegar	0%	0%	0%	0%	0%	0%	0%
Ver	0%	0%	0%	0%	33%	67%	0%
Alocar	0%	0%	50%	0%	0%	0%	50%
Pagar	0%	0%	0%	0%	0%	0%	0%



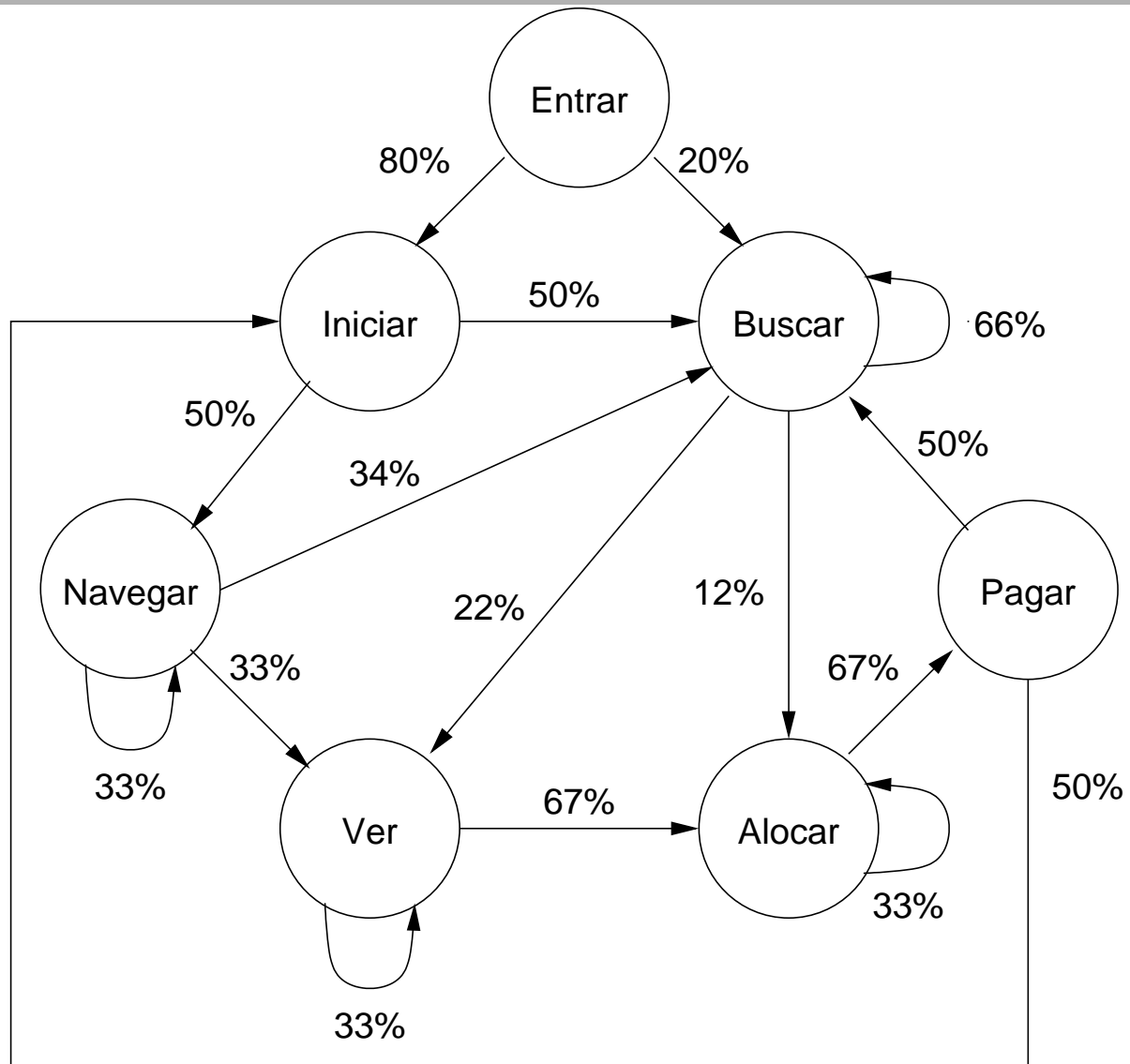
# Composição das sessões: Sessão Não-Comprador

<i>Requisição Origem</i>	<i>Requisição Destino</i>						
	Entrar	Iniciar	Buscar	Navegar	Ver	Alocar	Pagar
Entrar	0%	100%	0%	0%	0%	0%	0%
Iniciar	0%	0%	0%	100%	0%	0%	0%
Buscar	0%	0%	60%	0%	40%	0%	0%
Navegar	0%	0%	0%	0%	100%	0%	0%
Ver	0%	0%	100%	0%	0%	0%	0%
Alocar	0%	0%	0%	0%	0%	0%	0%
Pagar	0%	0%	0%	0%	0%	0%	0%

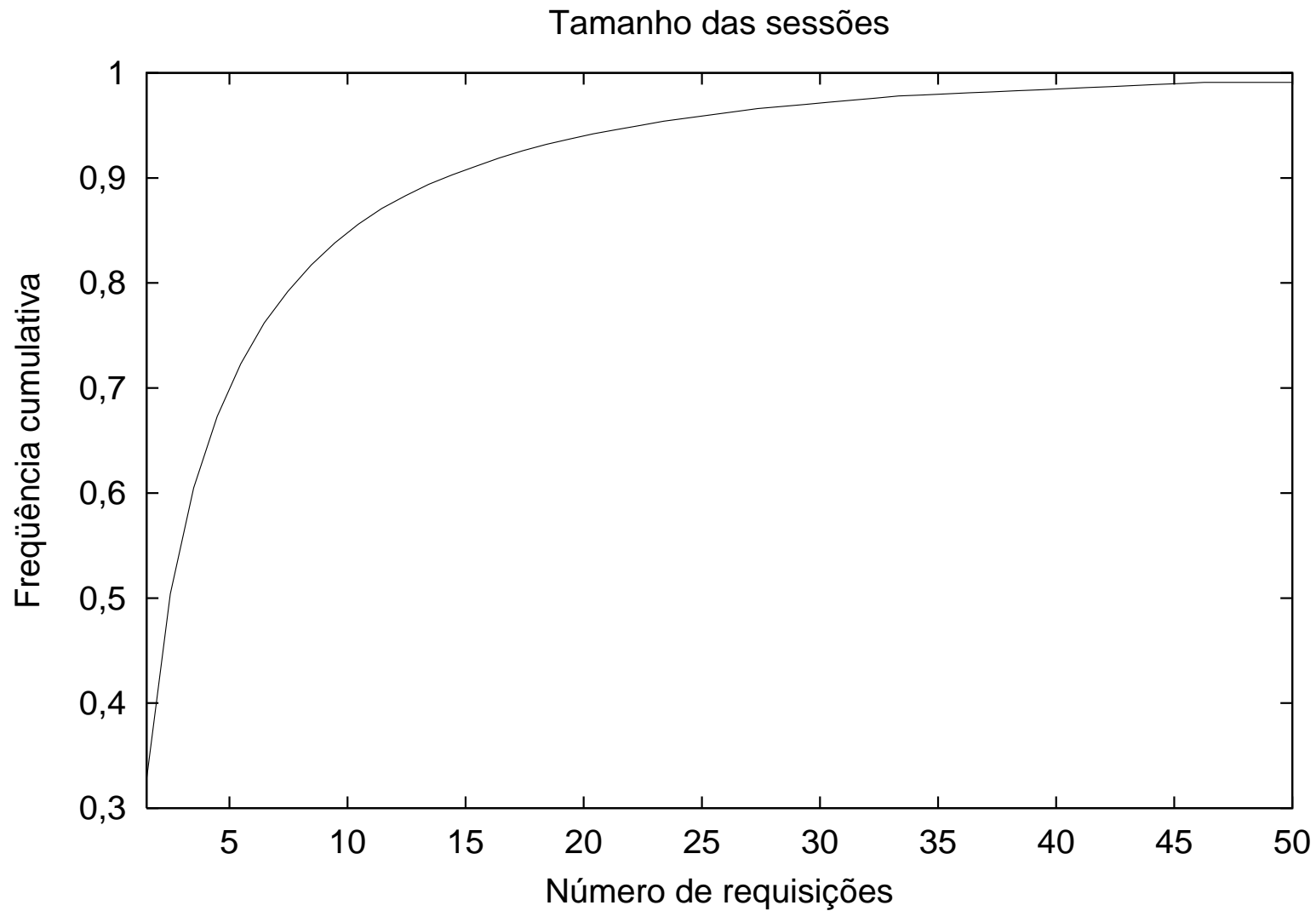
# Composição das sessões: Sessão Composta

<i>Requisição Origem</i>	<i>Requisição Destino</i>						
	Entrar	Iniciar	Buscar	Navegar	Ver	Alocar	Pagar
Entrar	0%	100%	0%	0%	0%	0%	0%
Iniciar	0%	0%	50%	50%	0%	0%	0%
Buscar	0%	0%	47%	0%	53%	0%	0%
Navegar	0%	0%	0%	0%	100%	0%	0%
Ver	0%	0%	50%	0%	17%	33%	0%
Alocar	0%	0%	50%	0%	0%	0%	50%
Pagar	0%	0%	0%	0%	0%	0%	0%

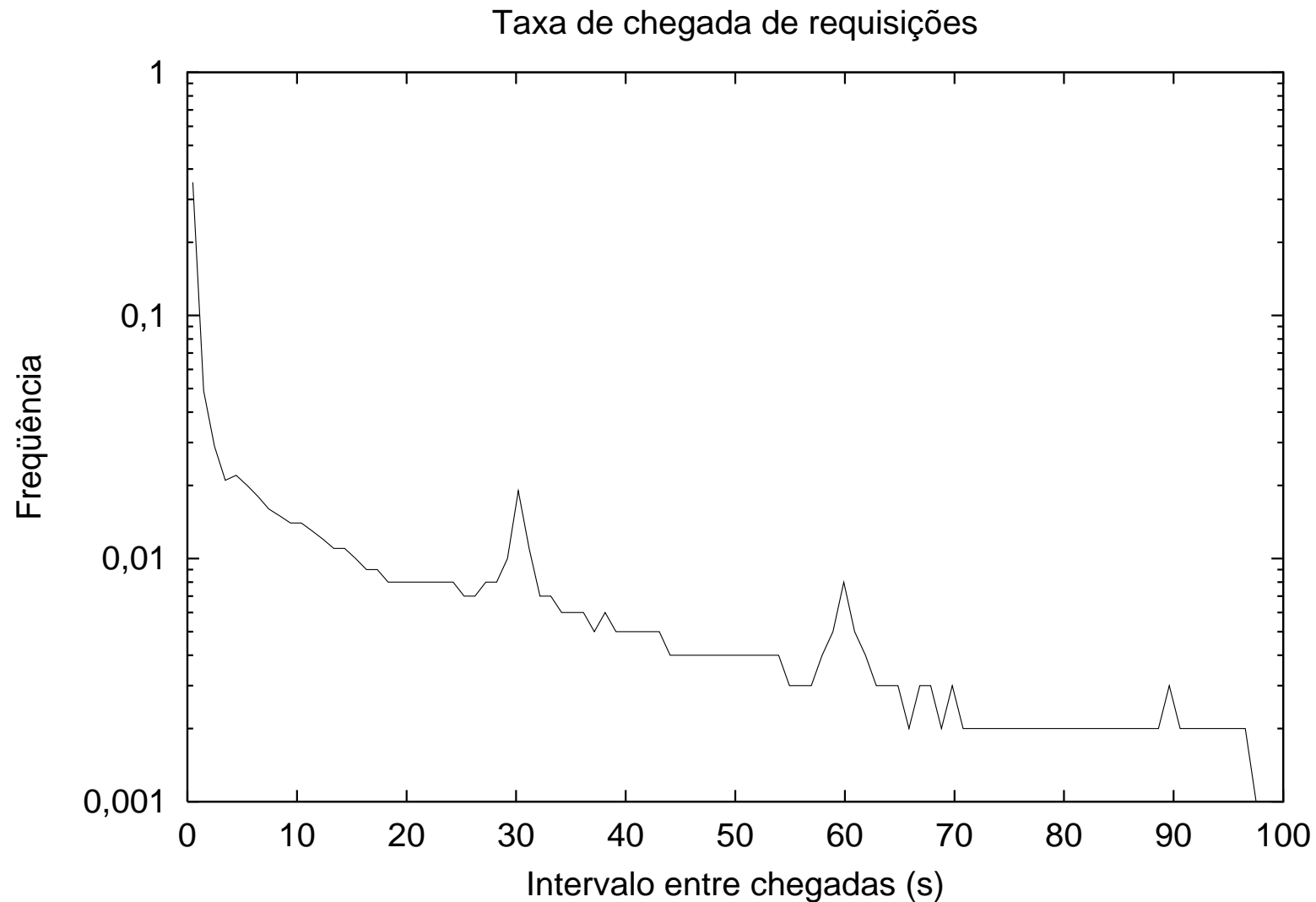
# CBMG de uma loja virtual.



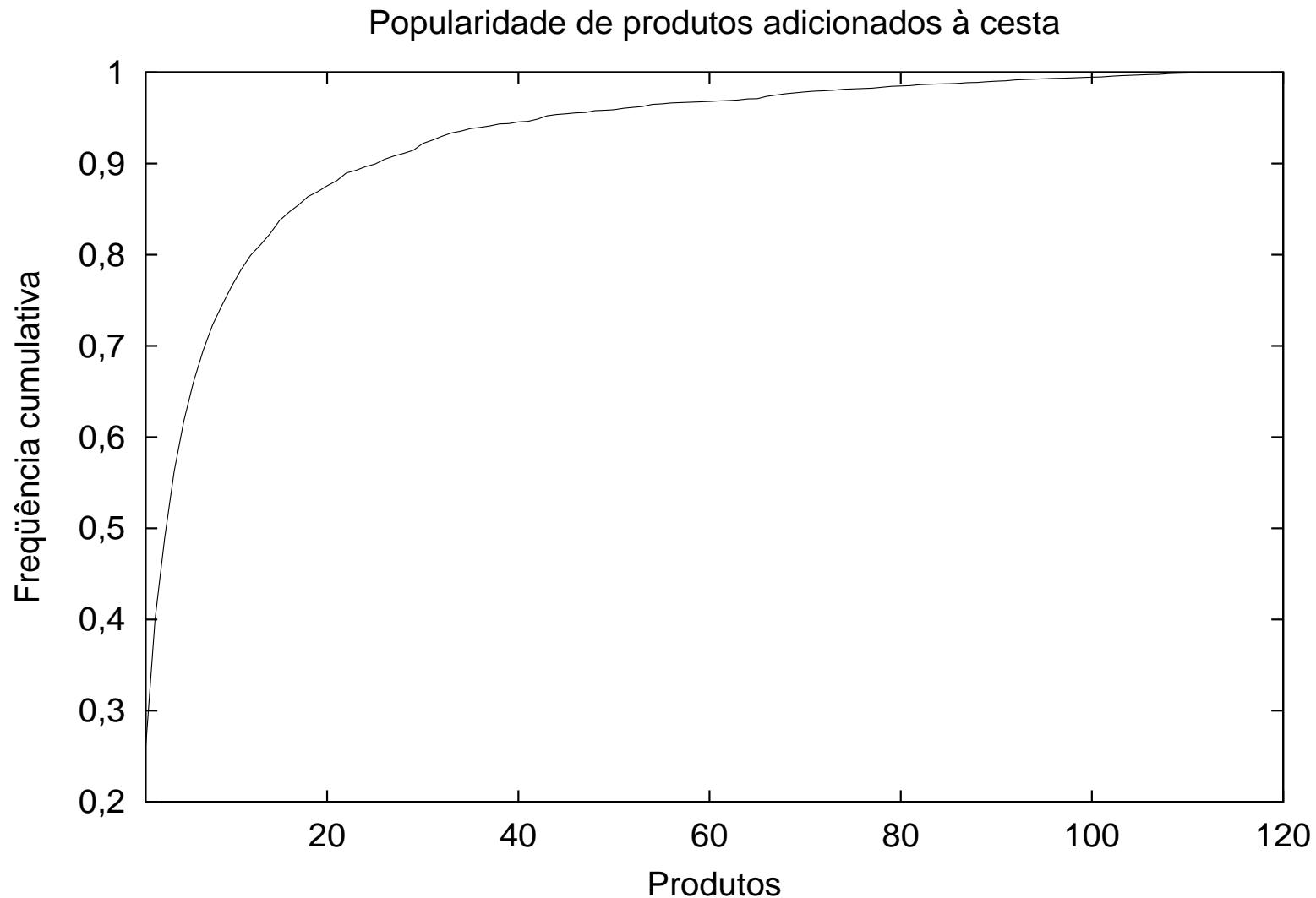
# Tamanho das sessões



# Processo de chegada das requisições



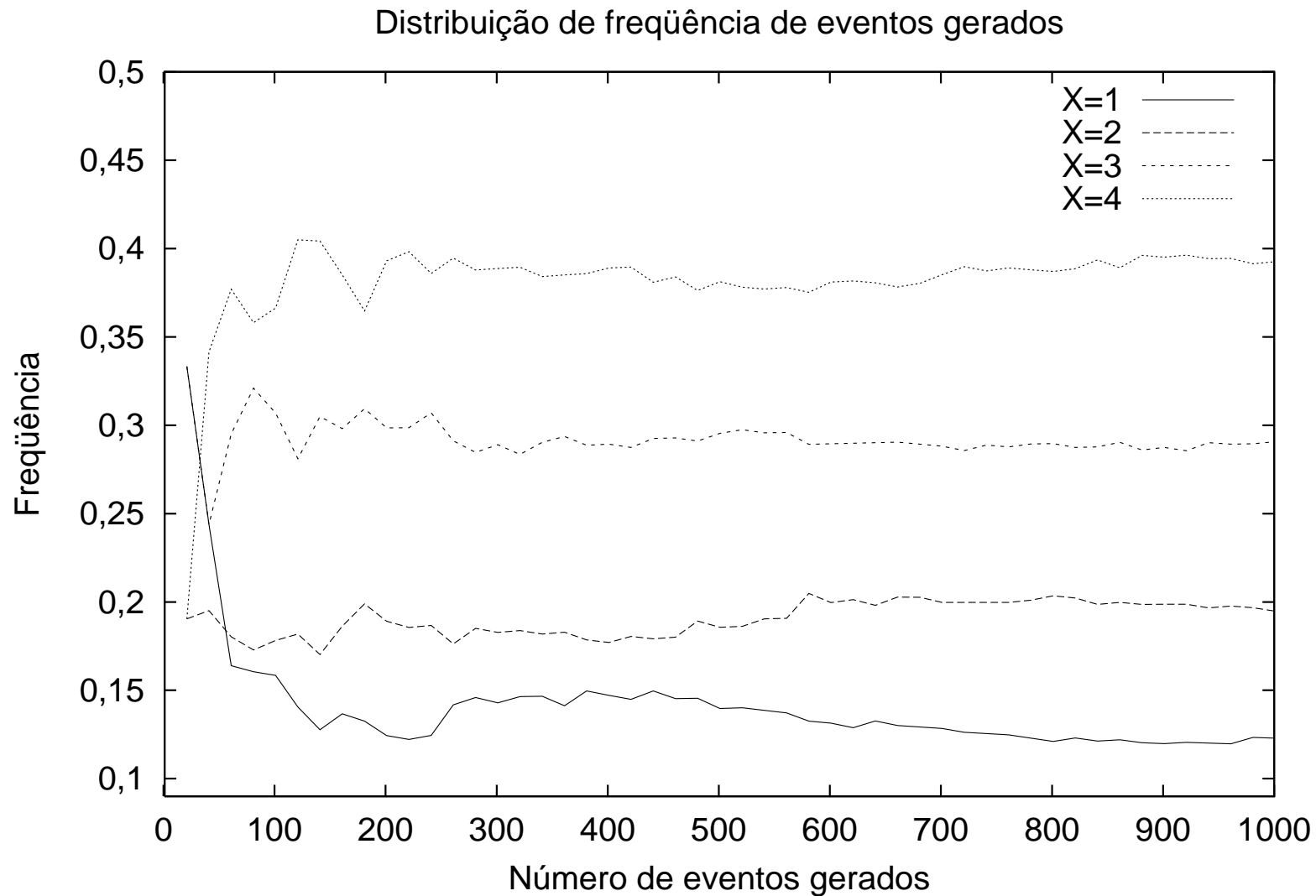
# Parâmetros das requisições



# Geração de carga de trabalho – DCP

```
1  gera_valor(n,ev,lim)
2      pr = random(0,1)
3      i = 0;
4      while (i<n && pr > lim[i]) i++;
5      if (i<n)
6          return ev[i];
7      else
8          return ev[n-1];
```

# Geração de carga de trabalho – DCP





# Geração de carga de trabalho a partir de CBMG

<i>Requisição corrente</i>	<i>Próxima requisição</i>						
	Entrar	Iniciar	Buscar	Navegar	Ver	Alocar	Pagar
Entrar	-	0,80	1,00	-	-	-	-
Iniciar	-	-	0,50	1,00	-	-	-
Buscar	-	-	0,66	-	0,88	1,00	-
Navegar	-	-	0,34	0,67	1,00	-	-
Ver	-	-	-	-	0,33	1,00	-
Alocar	-	-	-	-	-	0,33	1,00
Pagar	-	0,50	1,00	-	-	-	-

# Geração de carga de trabalho a partir de CBMG

```
1  gera_req(reqcorrente,mattrans)
2      pr = random(0,1)
3      i = 0;
4      while (i<numreq && pr > mattrans[reqcorrente][i]) i++;
5      if (i<n)
6          return mattrans[reqcorrente][i];
7      else
8          return mattrans[reqcorrente][n-1];
```

# Geração de carga de trabalho a partir de CBMG

<i>Requisição</i>	<i>Número aleatório</i>	<i>Requisição corrente</i>	<i>Próxima requisição</i>
1	0,041630	Entrar	Iniciar
2	0,454492	Iniciar	Buscar
3	0,834817	Buscar	Ver
4	0,335986	Ver	Alocar
5	0,565489	Alocar	Pagar
6	0,001767	Pagar	Iniciar
7	0,187590	Iniciar	Buscar
8	0,990434	Buscar	Alocar
9	0,750497	Alocar	Pagar
10	0,366274	Pagar	Iniciar

# Geração de carga de servidores de comércio eletrônico

## Parâmetros

1. nsess
2. freqsess
3. cbmgssess
4. tamsess
5. freqreq
6. paramreq

# Geração de carga de servidores de comércio eletrônico

```
1  gera_carga(nsess,freqsess,cbmgssess,tamsess,freqreq,paramreq)
2      inicioúltimasessão = 0
3      for (s=0; s<nsess; s++)
4          iníciosessão = inicioúltimasessão+gera_valor(freqsess)
5          tamsessão = gera_valor(tamsess)
6          reqcorr = Entrar
7          inícioúltimareq = iníciosessão
8          for (r=0; r<tamsessão; r++)
9              req[r].tipo = gera_req(reqcorr,cbmgssess)
10             req[r].início = inícioúltimareq
11             if (r>0)
12                 req[r].início += gera_valor(freqreq)
13             if (temporâmetro(req[r].tipo))
14                 req[r].param = gera_valor(paramreq)
15             reqcorr = req[r].tipo
16             inícioúltimareq = req[r].início
17             salva_sessão(iníciosessão,tamsessão,req)
18             inicioúltimasessão = iníciosessão
```

# Jambreiro

## Análise de Desempenho

- Componentes de software
  - Servidor Web Apache 1.3.9
  - Loja (servidor de transações)
  - SGBD MySQL 3.22.22

# Jambreiro

## Análise de Desempenho

- Serviços básicos
  - Página inicial
  - Navegação
  - Busca
  - Seleção
  - Colocação na cesta
  - Compra
  - Arquivo

# Jambreiro

## Nível Funcional

- Métricas de desempenho
  - Custo de processamento
    - tempo de utilização efetiva do processador em useg, delimitado pelo início e fim de operações, para determinar candidatos a otimização
    - Tempo decorrido
    - intervalo cronológico para execução de serviço ou segmento de código, sendo uma aproximação da latência percebida pelo cliente



### Caracterização de Carga

- geração baseada em *traces*
- dados divididos em sessões
  - parâmetros das requisições
  - intervalos entre requisições
- Experimentos:
  - frequência de sessões
  - granulação dos intervalos entre requisições
  - conteúdo realista da resposta

# Jambreiro

## Nível de Execução

- Instrumentação
  - código fonte dos componentes foi instrumentado
  - informação gerada por requisição tratada
- primitivas de sistema
  - gettimeofday
  - getrusage
- escalabilidade:
  - intrusão não depende da natureza dos dados

# Jambreiro

## Nível de Execução

---

- Escopo
  - local e global
  - cada componente gera um perfil local por requisição
  - perfis globais são gerados com base em identificadores únicos de requisição

# Jambreiro

## Nível de Execução

### Dimensões

- Componente:
  - permite localizar problemas de desempenho
- Serviço:
  - permite identificar serviços de maior custo
- Fases:
  - serviços são divididos em fases, permitindo determinar as tarefas de maior impacto

# Jambreiro

## Nível de Execução

---

- Monitor WWW
  - Interpretação
  - Controle de acesso
  - Execução
- Monitor do servidor de transações
  - Receber
  - Processar
  - Requisitar dados
  - Responder

# Jambreiro

## Nível de Execução

---

- Monitor do servidor de banco de dados
  - Insert
  - Update
  - Select
  - Connect
  - Quit

# Jambreiro Exemplos

- Custo de conexões seguras
  - aumento constante: 0.6 seg
  - impacto maior em serviços mais simples
  - custo aumenta com quantidade de dados transmitidos

- Módulos Apache vs. CGI
  - medimos ganhos da ordem de 7
  - taxa de utilização do servidor WWW não ultrapassou 10
  - o servidor de banco de dados é o ponto de contenção
  - ganhos apenas em termos de latência, pois nível de multiprogramação se manteve



- Avaliação da escalabilidade
  - variação do número de clientes simultâneos
  - página inicial e arquivo são escaláveis
  - busca não é escalável tendo em vista a saturação do servidor de banco de dados