

Eveline Alonso Veloso

Uma Hierarquia para Caracterização de *Live Streaming Media*

Dissertação apresentada ao Curso de Pós-Graduação em Ciência da Computação da Universidade Federal de Minas Gerais, como requisito parcial para a obtenção do grau de Mestre em Ciência da Computação.

Belo Horizonte

14 de agosto de 2002

Resumo

Este trabalho apresenta uma hierarquia para a caracterização de cargas de trabalho de *live streaming media* distribuídas através da *Internet*. Essa hierarquia baseia-se em três níveis de granularidades diferentes, correspondendo a transferências, sessões e clientes, respectivamente. Para a validação da hierarquia proposta, apresentamos a caracterização completa dos acessos aos objetos ao vivo disponibilizados por um servidor de *streaming media* muito popular no Brasil. Esses objetos receberam mais de três milhões de requisições e foram analisados durante um período de 28 dias. Os resultados encontrados dão suporte a duas conclusões importantes. A primeira delas é que a natureza das interações entre usuários e objetos é fundamentalmente diferente para objetos ao vivo e pré-armazenados. Acessos a objetos pré-armazenados são determinados pelo usuário, ao passo que, os acessos a objetos ao vivo são predominantemente dirigidos pelos próprios objetos. Essa diferença entre os papéis (ativo/passivo) de usuários e objetos conduz a contrastes bastante interessantes. Por exemplo, nossa análise mostra que o perfil de interesse dos usuários por um determinado objeto ao vivo pode ser modelado por uma lei de *Zipf*. Por outro lado, a popularidade dos objetos pré-armazenados segue uma lei de *Zipf*. Além disso, nossa análise revela que as durações das transferências de *streams* ao vivo são altamente variáveis e que esta variabilidade se deve não apenas às propriedades estruturais destes objetos, como tamanho e duração totais, mas também à aderência dos clientes ao *stream*. A segunda conclusão se baseia nas observações reportadas neste documento e que nos levam a acreditar que as características particulares das cargas de trabalho de *streaming media* ao vivo são, provavelmente, altamente dependentes da natureza do conteúdo ao vivo em exibição. Acreditamos que as fortes correlações temporais, observadas nos registros de acesso analisados no presente estudo, resultam da sincronização entre o conteúdo ao vivo e as características de seu acesso. Baseados em nossas análises, apresentamos um modelo para geração de cargas de trabalho de *live streaming media* que incorpora diversos resultados de nossa caracterização.

Abstract

This work shows a hierarchy for the characterization of *live streaming media* content delivered over the Internet. This hierarchy consists on three different granular levels, corresponding to transfers, sessions and clients, respectively. Using this hierarchy we characterize the accesses to live streaming objects. These objects received over three million requests, spanning a 28-day period. Our findings support two important conclusions. First, we show that the nature of interactions between users and objects is fundamentally different for live versus stored objects. Accesses to stored objects are *user driven*, whereas accesses to live objects are, predominantly, *object driven*. This reversal of active/passive roles of users and objects leads to interesting contrasts. For instance, our analysis underscores a Zipf-like profile for user interest in a given live object, which is to be contrasted to the classic Zipf-like popularity of pre-stored objects. Also, our analysis reveal that transfer lengths are highly variable and that this variability is due not only to the structural properties of the objects, like total size or length, but also, to the stickiness of clients to a particular live object. Second, based on observations we made, we conjecture that the particular characteristics of live media access workloads are likely to be highly dependent on the nature of the live object being accessed. In our study, this dependence is clear from the strong temporal correlations we observed in the traces, which we attribute to the synchronizing impact of live content on access characteristics. Based on our analysis, we present a model for live media workload generation that incorporates many of our findings.

Agradecimentos

A Deus, pelas oportunidades que vem me proporcionando e por sua presença em todos os momentos de minha vida.

A meus pais, Eliane e Luiz Antônio, pelo ilimitado e constante amor. Agradeço as orações e peço desculpas pelas ausências. Palavras não são suficientes para expressar meu carinho, respeito e gratidão por vocês. Ao Danilo, por tornar a vida tão alegre e interessante. Agradeço por seu amor, encorajamento, companherismo e paciência. Agradeço também por saber compreender e respeitar meu trabalho, minhas decisões e minhas ausências.

A meu orientador, Virgílio Almeida, minha gratidão por seu incentivo constante e pela orientação. Ao professor Azer Bestavros, da *Boston University*, pela orientação. Ao professor Wagner Meira Jr., por também ter participado do desenvolvimento deste trabalho e, principalmente, por me auxiliar a compreender que mudanças também são necessárias. Trabalhar com estes pesquisadores foi um privilégio para mim. Tenho por eles grande admiração.

A meu irmão, Adriano Veloso, pela companhia e momentos de descontração. A Rodrigo Barra, Adriano César, Flávia, Tassni e Benício, pela amizade. Vocês se tornaram minha segunda família. Meu agradecimento pela compreensão, auxílio e pelos momentos tão agradáveis que, sem dúvida, tornaram o trabalho mais suave. Agradeço especialmente ao Rodrigo Barra por seu interesse e por sua ajuda essencial no trabalho de minha dissertação. A todos os colegas do Laboratório *e-speed*, do Departamento de Ciência da Computação da UFMG, por tornarem agradável e divertido o convívio diário.

Agradeço aos vários anônimos profissionais que fizeram os registros das cargas de acesso utilizadas neste trabalho, sem os quais o mesmo poderia não ter sido realizado. Sua contribuição foi inestimável.

Agradeço também ao CNPQ pelo apoio financeiro.

Sumário

Lista de Figuras	5
Lista de Tabelas	7
1 Introdução	8
1.1 Características de <i>Streams</i> Ao Vivo x Pré-armazenados	10
1.2 Organização da Dissertação	11
2 Trabalhos Relacionados	13
2.1 Caracterização do Acesso a Objetos de <i>Streaming Media</i>	13
2.2 Caracterização do Tráfego Gerado por Objetos de <i>Streaming Media</i>	15
3 Hierarquia para Caracterização de <i>Live Streaming Media</i> e Terminologia	17
3.1 Camada de Transferência	18
3.2 Camada de Sessão	18
3.3 Camada de Cliente	19
4 Estudo de Caso: <i>Reality Show</i>	21
4.1 Natureza da Carga de Trabalho	21
4.2 Estatísticas e Informações Básicas sobre os Históricos de Acesso	22
4.3 Características da Camada de Transferência	23
4.3.1 Número de Transferências Concorrentes	24
4.3.2 Tempo entre as Chegadas de Transferências Consecutivas	27
4.3.3 Distribuição da Duração das Transferências	29
4.3.4 Largura de Banda Utilizada pelas Transferências	30
4.4 Características da Camada de Sessão	31
4.4.1 Número de Sessões	31
4.4.2 Distribuição do Período <i>ON</i> de Transferência	33
4.4.3 Distribuição do Período <i>OFF</i> Ativo	33
4.4.4 Número de Transferências por Sessão	35
4.4.5 Tempo entre as Chegadas de Transferências Consecutivas dentro de uma mesma Sessão	35
4.5 Características da Camada de Cliente	36
4.5.1 Distribuição Topológica e Geográfica dos Clientes	36
4.5.2 Número de Clientes Concorrentes	37
4.5.3 Tempo entre as Chegadas de Clientes Consecutivos	40
4.5.4 Distribuição do Período <i>ON</i> de Sessão	42
4.5.5 Distribuição do Período <i>OFF</i> Inativo	43
4.5.6 Perfil de Interesse dos Clientes	44

5	Caracterização de Outro Objeto Ao Vivo	46
5.1	Natureza da Carga de Trabalho	46
5.2	Camada de Transferência	47
5.2.1	Distribuição da Duração das Transferências	47
5.3	Camada de Sessão	47
5.3.1	Número de Sessões	47
5.3.2	Distribuição do Período <i>OFF</i> Ativo	48
5.3.3	Número de Transferências por Sessão	49
5.4	Camada de Cliente	49
5.4.1	Número de Clientes Concorrentes	49
5.4.2	Tempo entre as Chegadas de Clientes Consecutivos	51
5.4.3	Distribuição do Período <i>OFF</i> Inativo	53
5.4.4	Perfil de Interesse dos Clientes	53
6	Geração de Cargas de Trabalho de <i>Live Streaming Media</i> Sintéticas	55
7	Conclusões e Trabalhos Futuros	58
	Bibliografia	60

Lista de Figuras

3.1	Relação entre as atividades do cliente e os períodos <i>ON</i> e <i>OFF</i> da transferência e da sessão.	18
4.1	Distribuição marginal do número de transferências concorrentes associadas ao conteúdo ao vivo	24
4.2	Comportamento temporal do número de transferências concorrentes, $transf(t)$, associadas aos objetos ao vivo	25
4.3	Comportamento temporal do número de transferências concorrentes, $transf(t)$, associadas aos objetos pré-armazenados	26
4.4	Complemento da distribuição acumulativa do iat entre transferências consecutivas	27
4.5	Comportamento temporal do iat entre transferências consecutivas	28
4.6	Distribuição marginal da duração das transferências associadas ao conteúdo ao vivo	29
4.7	Distribuição da largura de banda média utilizada	31
4.8	Comportamento temporal da largura de banda total utilizada por transferências associadas a objetos ao vivo	32
4.9	Relação entre o número de sessões distintas identificadas no <i>log</i> de acesso a objetos ao vivo e T_{off}	32
4.10	Relação entre o número de sessões distintas identificadas no <i>log</i> de acesso a objetos pré-armazenados e T_{off}	33
4.11	Distribuição marginal dos períodos <i>ON</i> de transferência associados ao conteúdo ao vivo	34
4.12	Histograma da duração do <i>think time</i> dos clientes	34
4.13	Distribuição marginal do número de transferências dentro de uma mesma sessão	35
4.14	Distribuição marginal do iat de transferências consecutivas dentro de uma mesma sessão	36
4.15	Diversidade da população de clientes	37
4.16	Distribuição das transferências pelos países	38
4.17	Distribuição marginal do número de clientes ativos acessando <i>streams</i> ao vivo	38
4.18	Comportamento temporal do número de clientes ativos acessando <i>streams</i> ao vivo	39
4.19	Comportamento temporal do número de clientes ativos acessando objetos pré-armazenados	40
4.20	Função de auto-correlação do número de clientes ativos ao longo do tempo	41
4.21	Complemento da distribuição acumulativa do tempo decorrido entre as chegadas de clientes consecutivos	41
4.22	Comportamento temporal do iat de clientes consecutivos acessando <i>streams</i> ao vivo	42
4.23	Distribuição marginal do período <i>ON</i> de sessão	43
4.24	Distribuição marginal do período <i>OFF</i> inativo	44
4.25	Perfil de interesse do cliente	45

5.1	Distribuição marginal da duração das transferências individuais associadas ao conteúdo ao vivo	47
5.2	Relação entre o número de sessões distintas identificadas e T_{off}	48
5.3	Distribuição marginal da duração do <i>think time</i> dos clientes	48
5.4	Distribuição acumulativa do número de transferências dentro de uma mesma sessão	49
5.5	Distribuição marginal do número de clientes concorrentes associados ao conteúdo ao vivo	50
5.6	Comportamento temporal do número de clientes concorrentes, $c(t)$, que acessam o objeto ao vivo	51
5.7	Distribuição marginal do <i>iat</i> de clientes consecutivos	52
5.8	Comportamento temporal do <i>iat</i> de clientes consecutivos	53
5.9	Complemento da distribuição acumulativa do período <i>OFF</i> inativo	54
5.10	Perfil de interesse do cliente	54

Lista de Tabelas

4.1	Estatísticas básicas dos históricos de acesso analisados neste capítulo . . .	22
6.1	Resumo das propriedades de distribuição das variáveis utilizadas para a geração de cargas de trabalho de <i>streaming media</i> ao vivo - características encontradas para os <i>streams</i> associado ao <i>reality show</i>	57
6.2	Resumo das propriedades de distribuição das variáveis utilizadas para a geração de cargas de trabalho de <i>streaming media</i> ao vivo - características encontradas para o <i>stream</i> associado ao programa de rádio	57

Capítulo 1

Introdução

A utilização da *Internet* como um canal para distribuição e transferência de mídia de fluxo contínuo tem aumentado consideravelmente nos últimos anos. São exemplos de objetos de fluxo contínuo, também chamados de *streaming media*, ou simplesmente *streams*, programas de notícias, eventos esportivos, *trailers* de filmes, músicas, programas educacionais e de entretenimento, entre outros. Neste tipo de mídia, o usuário assiste e/ou escuta o *stream* de dados solicitado à medida que este é recebido na máquina cliente, não sendo necessário armazenar o objeto localmente.

Objetos de *streaming media* ainda impõem diversos desafios aos projetistas de sistemas. Dentre eles, podemos citar: primeiro, se comparados com outras aplicações tradicionais na *Internet*, tais como correio eletrônico e navegação por páginas *Web*, objetos de *streaming media* podem necessitar de taxas de transferência de dados maiores e consumir uma largura de banda mais significativa. Segundo, o protocolo da camada de transporte mais freqüentemente utilizado para transmitir objetos de *streaming media* é o UDP - *user datagram protocol*. Assim, a tarefa de realizar controle de congestionamento na rede passa a ser de responsabilidade da camada de aplicação. Terceiro, o tráfego gerado tende a ser em rajadas e a apresentar alta sensibilidade a atrasos. Quarto, objetos de *streaming media* requerem significativamente mais espaço de armazenamento do que objetos *Web* clássicos, aumentando, potencialmente, os requisitos de armazenamento dos servidores de *streaming media* e de *proxy caches* e os requisitos de largura de banda da rede por onde os objetos serão transmitidos. Esse fato motiva o desenvolvimento de políticas mais complexas e eficientes de substituição de objetos nas *caches*, que armazenam uma cópia local do objeto. Quinto, objetos de *streaming media* apresentam longas durações, se comparadas com o tempo de transferência de objetos *Web* tradicionais. Além disso, são observadas múltiplas requisições simultâneas pelo mesmo objeto. Esses fatos introduzem a oportunidade de utilização de técnicas de *multicast*, para reduzir a utilização da rede para a transmissão de objetos de *streaming media* populares. Todas essas características evidenciam a neces-

sidade de elaboração de estudos mais detalhados sobre objetos de *streaming media* e de desenvolvimento de técnicas de distribuição e transferência de *streams* mais eficazes. A caracterização e geração de cargas de acesso a objetos de *streaming media* sintéticas são, então, de fundamental importância para a avaliação de tais técnicas.

Nos últimos anos, um pequeno número de estudos foi realizado com o objetivo de caracterizar cargas de trabalho de *streaming media*. Dentre eles, podemos citar [1, 2, 3, 11, 18] e [21]. No entanto, pelo que conhecemos, todos esses trabalhos se concentraram em objetos de *streaming media* pré-armazenados, isto é, aqueles que foram previamente gravados e armazenados no servidor de *streaming media* para, só então, poderem ser assistidos pelos usuários tais como *trailers* de filmes, músicas e programas educacionais. Nenhum deles considerou a caracterização, em detalhes, de objetos ao vivo, também conhecidos como objetos *live*, ou seja, objetos em que o usuário assiste e/ou escuta aquilo que está acontecendo, em tempo real, como programas de rádio e televisão transmitidos através da *Internet*, entrevistas, eventos esportivos, etc. Esse documento apresenta uma hierarquia para a caracterização desse último tipo de mídia. Para validar a hierarquia proposta, mostramos a caracterização completa dos acessos a *streams* ao vivo disponibilizados por um servidor de *streaming media* muito popular no Brasil. Esses objetos receberam milhões de requisições, originadas de milhares de clientes e foram transmitidos pela *Internet* como complemento a um *reality show* exibido como programa de televisão no Brasil.

A caracterização de mídia de fluxo contínuo, ao vivo, distribuída através da *Internet*, além de ser um tópico bastante interessante é provavelmente de extrema importância dada a crescente utilização da *Internet* como canal para distribuição e transferência de conteúdo ao vivo em complemento a outros canais de difusão como o rádio e a televisão. A *Internet* permite que objetos de *streaming media* ao vivo sejam disponibilizados por indivíduos comuns por não necessitar da infra-estrutura e das técnicas de editoração indispensáveis em outros meios de difusão. Por exemplo, um usuário pode fixar uma câmera e transmitir as imagens capturadas por ela, através da *Internet*, como a visão do goleiro em um jogo de futebol, uma aula ou palestra, etc. Se comparada com outros canais de difusão de mídia, apenas a arquitetura da *Internet* permite a distribuição e transferência de conteúdo de *streaming media* ao vivo dessa forma. Assim, a ausência da necessidade de complexas e enormes infra-estruturas para difusão de mídia contínua e de controle de editoração tem proporcionado o crescimento da *Internet* como um complemento dos difusores tradicionais de informação.

A caracterização de cargas de trabalho é um ingrediente importante na avaliação de desempenho e planejamento de capacidade de sistemas de computação em geral. Quando nos referimos ao planejamento adequado da capacidade de infra-estruturas de distribuição de mídia de fluxo contínuo ao vivo como servidores, rede, *CDNs* - *Content Distribution Networks* -, etc, esse tópico torna-se particularmente crítico. Se a carga em um recurso

associado à distribuição de conteúdo pré-armazenado como um servidor, por exemplo, atingir um determinado limite a partir do qual esse recurso não é mais capaz de cumprir sua função adequadamente, esse pode optar por simplesmente rejeitar novas requisições. Essa solução, baseada em controle de admissão, pode ser aceitável em ambientes de distribuição de mídia armazenada, já que o usuário que teve sua requisição rejeitada pode reenviá-la mais tarde. No entanto, quando nos referimos a objetos de *streaming media* ao vivo, rejeitar a requisição de um usuário equivale a negar a esse usuário acesso à mídia ao vivo, uma vez que o valor desse tipo de conteúdo está justamente em sua *liveliness*. Nesse ambiente, se a requisição do usuário for rejeitada em um determinado instante e esse mesmo usuário retornar mais tarde, o conteúdo ao vivo que o usuário receberá como resposta a essa segunda requisição será totalmente diferente daquele que teria sido enviado em resposta à primeira tentativa. Assim, controle de admissão não é uma alternativa viável para provedores de conteúdo (ou seus *proxies*, como *CDNs*) quando estes disponibilizam para seus clientes *streams* ao vivo. Assim, o planejamento de capacidade baseado em um adequado, acurado e aperfeiçoado entendimento das características da carga de trabalho torna-se uma necessidade. Um fato que ilustra tal importância refere-se à experiência de milhares de clientes que, em janeiro de 1999, frustraram-se ao tentar assistir, ao vivo, através da *Internet*, o amplamente divulgado desfile da coleção de inverno da *VictoriaSecret.com*.

1.1 Características de *Streams* Ao Vivo x Pré-armazenados

As características da carga de trabalho de *streams* ao vivo são provavelmente fundamentalmente diferentes daquelas apresentadas pelas cargas de trabalho de objetos pré-armazenados. Cargas de trabalho de objetos de *streaming media* ao vivo provavelmente exibem correlações temporais, por exemplo, diárias ou semanais, fortes, que podem não estar presentes, ou pelo menos não tão significativamente, em cargas de trabalho de mídia armazenada. Além disso, o conjunto de operações VCR que podem ser executadas sobre objetos de *streaming media* pré-armazenados simplesmente não está disponível, quando nos referimos a conteúdo ao vivo. Mais importante ainda, as correlações entre diversas variáveis podem ser significativamente diferentes para objetos ao vivo e pré-armazenados. Por exemplo, consideremos a possível correlação existente entre a duração do intervalo de tempo durante o qual um usuário está assistindo a um *stream* e a qualidade, percebida por esse usuário, do *stream* em questão, que é resultante, entre outros fatores, da largura de banda da rede disponível. Para objetos pré-armazenados poderíamos esperar uma correlação positiva entre essas duas propriedades características da carga de trabalho, isto é, poderíamos esperar que os usuários apresentassem a tendência de parar de assistir ao

stream quando a qualidade de serviço (*QoS* - *Quality of Service*) se degradasse a um nível abaixo de um determinado patamar. Para *streams* ao vivo esta correlação pode ser muito mais fraca e/ou o patamar aceitável da *QoS* pode ser significativamente diferente uma vez que os usuários não possuem a opção de requisitar o conteúdo novamente, no futuro, como acontece com objetos pré-armazenados.

As diferenças entre os padrões de acesso a objetos ao vivo e pré-armazenados mencionadas acima derivam fundamentalmente das divergências entre os papéis que usuários e objetos desempenham em cada caso. Os acessos a conteúdo armazenado são determinados pelo usuário: ele indica o objeto que deve ser acessado e o instante em que esse acesso deve ocorrer, ou seja, os acessos a esse tipo de mídia são diretamente influenciados pelas preferências dos usuários. Por outro lado, os acessos à mídia de fluxo contínuo, ao vivo, são predominantemente dirigidos pelo próprio objeto: eles são diretamente influenciados por aspectos relacionados à natureza do objeto, como o horário em que o *stream* ao vivo está disponível, as atividades particulares exibidas em cada instante, etc. Em tal ambiente, os usuários desempenham um papel menos ativo e, a forma como podem interagir com os objetos é amplamente limitada. Na verdade, eles podem apenas juntar-se a ou abandonar a audiência do *stream* ao vivo, o qual, por sua vez, desempenha função mais ativa que a apresentada por objetos armazenados.

1.2 Organização da Dissertação

O restante desse documento está organizado da seguinte maneira: no Capítulo 2, apresentamos uma visão geral dos trabalhos relacionados ; no Capítulo 3, propomos e descrevemos, em detalhes, uma hierarquia de três níveis de granularidades diferentes para a caracterização de cargas de trabalho de *streaming media* ao vivo. Ainda nesse capítulo, introduzimos a terminologia que será adotada no restante da dissertação. Essa hierarquia foi aplicada na caracterização completa de *streams* ao vivo transmitidos, através da *Internet*, como complemento a um *reality show* exibido como programa de televisão. No Capítulo 4, apresentamos informações e estatísticas básicas relacionadas aos registros de acesso coletados para esses objetos e mostramos também o resultado de sua caracterização nos três níveis de granularidades diferentes correspondendo às características das transferências (Seção 4.3), às propriedades das sessões (Seção 4.4) e ao processo de chegada e comportamento dos clientes (Seção 4.5). Aplicamos ainda a hierarquia proposta nesse documento na caracterização de outro objeto de fluxo contínuo de dados ao vivo associado a um programa de rádio. Mostramos, no Capítulo 5, algumas propriedades interessantes encontradas. No Capítulo 6, apresentamos um modelo para geração de cargas de trabalho de *streaming media* ao vivo sintéticas, que incorpora diversos resultados de nossa caracterização. Finalmente, concluímos esse trabalho com o Capítulo 7, onde apresentamos um

resumo de nossas conclusões e algumas direções para trabalhos futuros.

Capítulo 2

Trabalhos Relacionados

A caracterização de cargas de trabalho é fundamental para a geração de cargas de trabalho sintéticas, porém realistas. As caracterização e geração de cargas de trabalho sintéticas de mídias, que não apresentam um fluxo contínuo de dados, como, por exemplo, páginas *Web* tradicionais, têm sido o tópico de diversos estudos. Dentre eles, podemos destacar [4, 5, 6, 8, 7, 9, 13, 14, 16, 23] e [22]. Todos esses trabalhos têm contribuído para o aperfeiçoamento de nosso entendimento sobre a natureza e as propriedades dos padrões de acesso a conteúdo armazenado, que não corresponde a objetos com um fluxo contínuo de dados. São exemplos de observações importantes reportadas nesses estudos: a distribuição da popularidade de objetos desse tipo segue uma lei de *Zipf* [25], a distribuição do tamanho desses documentos apresenta cauda-pesada e o padrão de acesso a esse tipo de mídia apresenta evidências de localidade de referência e auto-similaridade. A discussão mais detalhada das várias características observadas em cargas de trabalho de mídia não-contínua, apesar de constituir-se um tópico bastante interessante, não enquadra-se no escopo e contexto do presente estudo. Assim, no restante deste capítulo, restringimo-nos a cobrir trabalhos relacionados à caracterização de cargas de trabalho de *streaming media*.

2.1 Caracterização do Acesso a Objetos de *Streaming*

Media

Nos últimos anos, alguns estudos caracterizaram cargas de trabalho de objetos de *streaming media* pré-armazenados. Um resumo desses esforços é apresentado a seguir.

Acharya e Smith, em [2], caracterizaram o acesso dos usuários a vídeos disponibilizados na *Web*. Suas análises revelaram a existência de fortes evidências de localidade temporal no acesso a esse tipo de mídia. Além disso, os acessos estudados apresentaram também localidade geográfica, isto é, um pequeno número de máquinas clientes foi responsável pela

maioria das requisições. Os autores observaram ainda que a distribuição de popularidade dos vídeos apresentava uma inclinação que não podia ser modelada por uma lei de *Zipf*. Adicionalmente, aproximadamente metade das requisições representavam acessos parciais e, freqüentemente, esses acessos eram interrompidos ainda no início do vídeo.

Padhye e Kurose, em [21] estudaram os padrões das interações dos usuários com um servidor de mídia contínua, em um sistema *MANIC*. Nesse trabalho a sessão de um usuário foi definida como sendo uma seqüência de períodos *ON* (quando esse usuário está efetivamente recebendo o *stream*) alternados com períodos *OFF* (quando nenhuma mídia está sendo transmitida a esse usuário). Os autores caracterizaram a duração das sessões e as atividades dos usuários dentro de uma mesma sessão. As distribuições dos períodos *ON* e *OFF* têm a aparência de uma distribuição cauda-pesada sendo modeladas por distribuições Gamma ou Lognormal. Além disso, foi observado que os usuários não assistiam ao vídeo seqüencialmente, executando operações para avanço ou retrocesso do *stream*.

Em [11], foi analisada uma carga de trabalho de objetos de *streaming media* composta pelas requisições realizadas pelos clientes que utilizavam os roteadores de borda que servem à Universidade de Washington para conectarem-se à *Internet*. Esse trabalho se concentrou na determinação de características como a distribuição do tamanho dos objetos requisitados, a distribuição da popularidade dos servidores e objetos acessados, padrões de compartilhamento de objetos e a utilização da largura de banda. Os autores reportaram que muitos dos objetos de *streaming media* requisitados eram pequenos, mas que, no entanto, uma pequena porcentagem das requisições foi responsável por quase a metade da quantidade de *bytes* transmitidos. Adicionalmente, a distribuição de popularidade dos objetos foi modelada por uma lei de *Zipf*. Além disso, as análises realizadas mostraram que as requisições exibiam um alto grau de localidade temporal durante os períodos de pico de carga. Utilizando a carga de trabalho coletada os autores estudaram a eficácia de técnicas de *caching* e *multicast* para reduzir a largura de banda utilizada pelos objetos de *streaming media* e obtiveram resultados interessantes e animadores.

Almeida et al., em [3], analisaram as cargas de trabalho de dois servidores de *streaming media* com fins educacionais. Durante períodos quando a taxa de chegada de requisições era aproximadamente estacionária, o processo de chegada de sessões de clientes em um desses servidores foi identificado como sendo aproximadamente de *Poisson* e o tempo entre as chegadas de requisições consecutivas no outro servidor foi modelado por uma distribuição de Pareto. Os autores modelaram a distribuição da popularidade dos objetos de *streaming media* pré-armazenados analisados por uma concatenação de duas leis de *Zipf*. Frequentemente, os diversos segmentos dos *streams* não eram acessados de maneira igualitária e, para objetos menos populares, seus segmentos iniciais apresentavam maior probabilidade de serem requisitados. Além disso, foi reportado que a distribuição da duração efetivamente acessada dos objetos pré-armazenados depende da duração total do objeto. Para

objetos com maior duração, essa distribuição apresenta freqüentemente uma aparência de cauda-pesada. Foi indicada ainda, nesse trabalho, a existência de um alto grau de interatividade do usuário na carga analisada, o que limita a eficácia de técnicas de *multicast* para reduzir os requisitos de largura de banda do servidor.

Nenhum trabalho que considerasse, em detalhes, a caracterização dos acessos a *streams* ao vivo foi encontrado. Por isso, acreditamos que o presente estudo seja o primeiro a tratar o tópico profundamente.

2.2 Caracterização do Tráfego Gerado por Objetos de *Streaming Media*

Muitos estudos concentraram-se em aspectos de baixo nível do acesso a objetos de *streaming media*, como taxa de perda de pacotes, atrasos e protocolos da camada de transporte utilizados. Dentre esses trabalhos podemos destacar os seguintes:

Mena e Heidemann, em [18], examinaram o tráfego originado de um serviço de áudio muito popular na *Internet* que utilizava o programa *RealAudio*. O foco desse trabalho foi identificar as características da carga de trabalho no nível de rede, como a distribuição do tamanho dos pacotes e outras características do fluxo de pacotes. Os autores descobriram que o tráfego gerado por mídia de fluxo contínuo apresenta uma tendência à utilização do UDP como protocolo da camada de transporte. Além disso, eles observaram significativa regularidade no tamanho dos pacotes de tráfego de áudio e nos padrões de taxa de transferência de dados, como o tempo entre chegadas de pacotes consecutivos, o que pode ser útil como ferramenta para identificar fluxo contínuo de áudio de servidores na *Internet*. Baseados nesse estudo recentemente Lan e Heidemann, em [10], identificaram as propriedades estruturais do tráfego contínuo de áudio gerado pelo *RealAudio*, desenvolveram e validaram um modelo de simulação elaborado na camada de aplicação.

Em [17], Loguinov e Radha analisaram várias métricas de desempenho da rede, incluindo taxa de perda de pacotes, atrasos, *jitter*, reordenação de pacotes, dentre outras. Particularmente, os resultados encontrados pelos autores sugerem que a taxa de perda de pacotes na *Internet* apresenta uma natureza em rajadas. Além disso, as distribuições da duração do *burst* de perda de pacotes e do tempo de transferência dos pacotes apresentam uma aparência de cauda-pesada.

Wang et al., em [24], analisaram o tráfego gerado por vários servidores de *streaming media* na *Internet* que utilizavam o *RealVideo* para servir diversos usuários distribuídos geograficamente. Os autores se concentraram predominantemente na taxa de *frames* e na influência da largura de banda do cliente. Eles descobriram que vídeos típicos transmitidos por servidores *RealVideo* alcançam alta qualidade, com taxa média de 10 *frames/s* ou mais.

O desempenho do vídeo é mais influenciado pela largura de banda da conexão do cliente com a *Internet*, mas, quando o cliente apresenta conexão de alta velocidade, o gargalo do desempenho do vídeo se aproxima do servidor.

Capítulo 3

Hierarquia para Caracterização de *Live Streaming Media* e Terminologia

Requisições por mídia de fluxo contínuo, ao vivo, são recebidas pelos servidores de *streaming media* de maneira *interleaved*. Para entendermos as características desse tipo de carga de trabalho, assim como as estruturas implícitas existentes nas interações entre usuários e serviços de *streaming media* ao vivo, definimos um modelo hierárquico para a caracterização dessa carga de trabalho ([19]). Nesse modelo, a carga de trabalho de *streaming media* ao vivo é tratada em uma hierarquia de camadas. Na camada que ocupa o nível mais baixo da hierarquia, o servidor de *streaming media* recebe requisições originadas de diversos clientes. No nível seguinte, requisições associadas a um mesmo cliente são agrupadas em sessões. Finalmente, na camada mais alta da hierarquia, sessões pertencentes a um mesmo cliente são agrupadas de maneira a identificar seu padrão comportamental.

Neste documento, utilizamos o termo objetos de *streaming media* ao vivo, ou simplesmente objetos ao vivo, ao referirmo-nos a *streams* ao vivo ou fluxo contínuo de dados ao vivo, cuja existência é definida pela duração do evento associado (um programa de rádio ou televisão transmitido através da *Internet*, um evento esportivo, uma entrevista, um desfile de moda, etc). Caracterizamos o acesso a tais objetos em três níveis de abstração ou camadas com granularidades diferentes, correspondendo a transferências, sessões e clientes. Dentro de cada camada, conduzimos uma análise das propriedades estatísticas e de distribuição das diversas variáveis relevantes dentro do nível de abstração considerado. Nosso objetivo é analisar cada camada individualmente para obter a caracterização dos processos de chegada significativos da camada (tempo entre chegadas de requisições consecutivas - *iat*, nível de concorrência), padrões de acesso (períodos *ON* e *OFF*, correlações temporais) e outras propriedades (utilização da largura de banda, distribuição de popularidade). A seguir, definimos a semântica associada a cada nível da hierarquia de caracterização e

introduzimos a terminologia adotada no restante da dissertação.

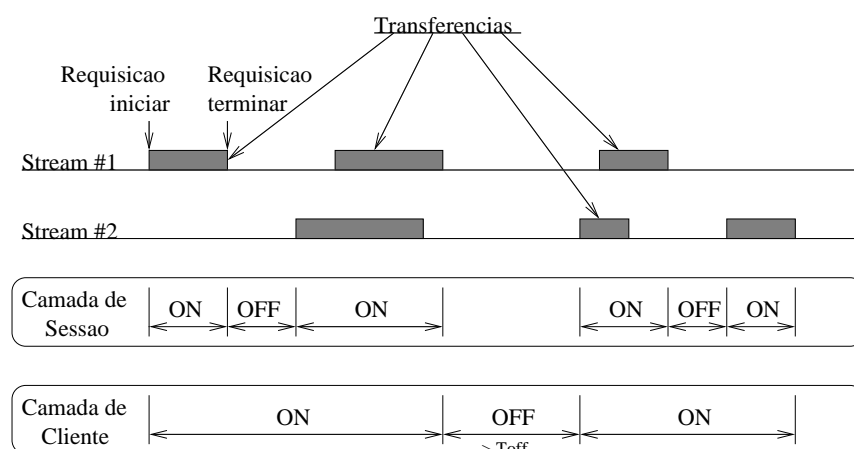


Figura 3.1: Relação entre as atividades do cliente e os períodos *ON* e *OFF* da transferência e da sessão.

3.1 Camada de Transferência

Esta camada ocupa o nível mais baixo da hierarquia para caracterização de *streaming media* ao vivo proposta e se concentra nas transferências individuais de fluxos contínuos de dados ao vivo, que resultam de ações específicas desempenhadas pelo cliente, como mostra a Figura 3.1. Uma transferência é o resultado de um par de requisições enviadas pelo cliente. Especificamente, para objetos ao vivo, essas requisições solicitam que o objeto comece e, eventualmente, pare de ser transmitido para o cliente. Para objetos de *streaming media* pré-armazenados, outras requisições podem incluir funcionalidades VCR como pausa, avanço (*fast-forward*) e retrocesso (*rewind*) do *stream*. Neste nível de abstração, caracterizamos a duração de transferências individuais, o tempo entre chegadas (*interarrival time*, ou simplesmente *iat*) de solicitações consecutivas por transferências do objeto ao vivo, o número de transferências concorrentes ao longo do tempo, a utilização da largura de banda, assim como correlações temporais de algumas variáveis.

3.2 Camada de Sessão

Agrupando transferências individuais associadas a um mesmo cliente, formamos sessões, que são caracterizadas nessa camada. Definimos uma sessão de um cliente como o intervalo de tempo durante o qual este está ativamente engajado em requisitar e receber partes dos objetos ao vivo que compõem um mesmo serviço, (por exemplo, um mesmo programa de televisão transmitido através da *Internet*) de forma que a duração de qualquer período em

que não ocorra nenhuma transferência de dados do servidor para o cliente não exceda um determinado patamar (*threshold* T_{off}). De acordo com essa definição, uma sessão de um cliente é uma sequência alternada de períodos de atividade e de silêncio, sendo que, por definição, esse último deve ser menor que T_{off} . Assim, uma sessão é caracterizada por intervalos quando ocorre transferência de dados do servidor para o cliente, que denominaremos período *ON* da transferência, cuja duração corresponde à união das durações das transferências simultâneas associadas a esse cliente, e por períodos de inatividade, chamados de *OFF* ativo. Durante o período *ON* da transferência, o cliente recebe partes de um ou mais objetos ao vivo associados a um mesmo serviço, como, por exemplo, diferentes visões de um mesmo programa. Durante intervalos de *OFF* ativo, nenhum objeto ao vivo está sendo transmitido ao cliente. Períodos de *OFF* ativo correspondem ao *think time* do cliente, ou períodos de inatividade dentro de uma mesma sessão. A Figura 3.1 mostra como as atividades do cliente resultam em diferentes períodos *ON* e *OFF* dentro de uma mesma sessão. Neste nível de abstração, iremos caracterizar as propriedades relevantes observadas dentro de uma mesma sessão como, a duração de períodos *ON* da transferência, a duração de períodos *OFF* ativos, o *iat* entre transferências de uma mesma sessão e o número de transferências pertencentes a uma mesma sessão.

3.3 Camada de Cliente

Esta é a camada que ocupa o nível mais alto da hierarquia para caracterização de *streaming media* ao vivo proposta e se concentra nas características e comportamento da população de clientes. Cada cliente é identificado através de um campo (*player ID*) presente em todo registro armazenado no histórico, ou *log*, de acessos. Esse campo apresenta uma chave, globalmente única, associada ao *software* utilizado pelo usuário para acessar objetos de *streaming media*. Devemos ressaltar que um cliente nesse caso, corresponde aproximadamente a um usuário individual. Exceções ocorrem apenas em casos em que o mesmo *software* cliente é utilizado por vários usuários que compartilham a mesma máquina cliente. O comportamento do cliente é caracterizado por uma sequência alternada de períodos de atividade, durante os quais o cliente apresenta sessões ativas, e períodos de silêncio. Cada período de atividade possui a duração da sessão ativa correspondente e será denominado *ON* de sessão. Já os períodos de silêncio serão denominados *OFF* inativos e correspondem a intervalos de tempo entre sessões consecutivas de um mesmo cliente. Por definição, cada período *OFF* inativo deve ser maior ou igual a T_{off} . As características que serão consideradas nessa camada incluem o número de clientes simultâneos acessando o conteúdo ao vivo (ou seja, o nível de concorrência) ao longo do tempo, o *iat* entre eles, a duração dos períodos de atividade (*ON* de sessão) e de inatividade (*OFF* inativo) dos clientes, suas distribuições geográfica e topológica, além do interesse do cliente no conteúdo ao vivo (em

relação a todos os outros usuários) e a frequência de acesso desse cliente (medida pelo número de sessões e transferências associadas a ele).

Caracterizar a carga de trabalho nesses níveis distintos de abstração, nos permite concentrar nossa análise nos diferentes elementos que interagem nesse tipo de ambiente, como clientes e objetos. Além disso, essa caracterização hierárquica pode ser utilizada para capturar mudanças no comportamento do cliente e mapear os efeitos dessas mudanças nas camadas mais abaixo no modelo hierárquico, isto é, nos níveis de sessão e de transferência. Finalmente, esse modelo em camadas nos possibilita desenvolver um processo para geração de cargas de trabalho de *streaming media* ao vivo sintéticas, como mostrado no Capítulo 6).

Capítulo 4

Estudo de Caso: *Reality Show*

A hierarquia para caracterização de *streaming media* ao vivo, proposta no Capítulo 3, foi utilizada para caracterizar, em detalhes, os acessos a um servidor de *streaming media* que disponibilizou, através da *Internet*, objetos de fluxo contínuo de dados ao vivo, associados a um *reality show* exibido na televisão brasileira. Nesse capítulo, descrevemos a natureza e apresentamos algumas estatísticas e informações básicas sobre a carga de trabalho analisada, assim como os resultados da caracterização e algumas possíveis explicações.

Sempre que for possível e relevante, iremos comparar os resultados encontrados para a caracterização dos acessos aos objetos ao vivo associados ao *reality show* com os obtidos na caracterização dos acessos a objetos de *streaming media* pré-armazenados, como *clips* musicais, *trailers* de filme, entrevistas, etc.

4.1 Natureza da Carga de Trabalho

Conseguimos o histórico, de quase um mês, dos acessos a um servidor de *streaming media* ao vivo muito popular operado por um dos dez maiores provedores de conteúdo do Brasil. Este servidor, que utilizava o *software* da *Microsoft* (*Windows Media Server* [12]) para disponibilizar mídia de fluxo contínuo ao vivo através da *Internet*, permitia que os usuários assistissem a um ou dois objetos de *streaming media* ao vivo associados a um *reality show* muito popular, exibido no início de 2002, por aproximadamente 90 dias, na televisão brasileira. A todo instante, cada um desses dois *streams* ao vivo transmitiam o áudio e o vídeo capturados pelas dúzias de câmeras espalhadas no ambiente dos competidores do *reality show*. Além disso, cada um desses dois objetos ao vivo era disponibilizado durante 24 horas por dia, durante todo o período de duração da competição.

Analisamos também o histórico de aproximadamente 19 dias, dos acessos a um servidor de *streaming media* que disponibilizava objetos pré-armazenados, tais como *trailers* de

filmes, *clips* musicais, entrevistas, etc. Essa última análise foi realizada apenas com o objetivo de comparar, quando julgamos interessante, os padrões e características de acesso de *streams* ao vivo e pré-armazenados.

4.2 Estatísticas e Informações Básicas sobre os Históricos de Acesso

A Tabela 4.1 apresenta as estatísticas básicas dos históricos de acesso analisados nesse capítulo.

	Objetos ao vivo	Objetos pré-armazenados
Período de <i>log</i> analisado	28 dias, 2002	19 dias, 2002
# total de objetos distintos	2	27821
# total de IPs clientes	364184	104181
# total de clientes distintos	690650	111836
# total de transferências	> 3 milhões	> 700000
# total de sessões	> 1400000	148056
Quantidade total de conteúdo servido	> 4,65 <i>TBytes</i>	> 310 <i>GBytes</i>

Tabela 4.1: Estatísticas básicas dos históricos de acesso analisados neste capítulo

Em ambos os históricos de acesso analisados, o servidor *Windows Media* foi configurado de forma a permitir o registro completo de todas as atividades dos clientes. Cada registro armazenado no *log* de acesso identifica um único par requisição do cliente-resposta do servidor. Apesar do servidor *Windows Media* suportar tanto transmissões *unicast* quanto *multicast*, somente transferências *unicast* foram permitidas. As seguintes informações são apresentadas em cada um dos registros armazenados nos históricos de acessos:

1. Identificação do cliente: endereço IP da máquina cliente; identificador, globalmente único, associado ao *software* cliente utilizado para acessar *streams*.
2. Especificação do ambiente do cliente: sistema operacional, versão do sistema operacional, processador da máquina cliente, etc.
3. Identificação do objeto de *streaming media* requisitado: URI do *stream* solicitado.
4. Data e horário (em segundos) quando o registro foi gravado no *log* de acesso.
5. Estatísticas da transferência: duração da transferência (em segundos), taxa de perda de pacotes, largura de banda média utilizada, etc.
6. Estatísticas sobre a carga do servidor: utilização da CPU do servidor, nível de concorrência.

7. Outras informações como: código de resposta HTTP, URI de onde a requisição se originou, etc.

O leitor pode consultar a documentação dos serviços *Windows Media* [12] se estiver interessado em conhecer maiores detalhes sobre as informações disponíveis nos *logs* de acesso.

Já que as informações armazenadas no *log* de acesso do servidor, relacionadas ao horário e à duração das transferências apresentam a granularidade de um segundo, freqüentemente seriam medidos intervalos de tempo de *zero* segundos para características como *iats* de transferências e de clientes, duração dos períodos *ON* de transferências e de sessões, duração dos períodos *OFF* ativo e inativo, etc. Assim, para permitir a apresentação de tais medidas em uma escala logarítmica, neste documento optamos por utilizar a função $\lfloor t + 1 \rfloor$ para representar um intervalo de tempo de t segundos.

Como ficará evidente nas próximas seções deste capítulo, existem períodos em que o número de clientes simultâneos acessando o conteúdo disponibilizado pelo servidor de *streaming media* é muito grande. Assim, uma questão importante é determinar se as características encontradas durante a caracterização da carga de trabalho foram influenciadas pelas capacidade e carga do sistema. Dada a natureza das interações entre os clientes e o sistema, um servidor sobrecarregado pode reduzir a velocidade das atividades dos clientes, ou mesmo, fazer com que os clientes desistam de acessar o conteúdo por ele disponibilizado. Se alguma dessas situações ocorrer, provocará algum impacto em nossa caracterização, influenciando, por exemplo, os *iats* de transferências e de clientes, o nível de concorrência ou a duração dos períodos *ON* de transferências e de sessões. Para garantir que as características apresentadas no restante desse capítulo não são afetadas por uma possível sobrecarga do servidor, analisamos os *logs* de acesso e descobrimos que períodos de sobrecarga do servidor são extremamente raros. Especificamente, extraímos todas as medições da utilização do processador do servidor, como reportado nos históricos de acessos, e calculamos sua média, em *bins* de um segundo. Os resultados indicam que a utilização do servidor ficou abaixo de 10% em mais de 99% do tempo. Analogamente, a carga do servidor ficou abaixo de 10% em mais de 99% das transferências registradas nos *logs* de acesso.

4.3 Características da Camada de Transferência

Nesta camada, estamos interessados em caracterizar a carga de trabalho do servidor de *streaming media* ao vivo no nível de granularidade de transferências individuais. Como foi definido no Capítulo 3, uma transferência é o resultado de um par de requisições enviadas pelo cliente ao servidor de *streaming media*. Especificamente para objetos ao vivo, essas requisições solicitam que o objeto comece e, eventualmente, pare de ser transmitido para

o cliente.

4.3.1 Número de Transferências Concorrentes

Em qualquer instante de tempo t , existe um determinado número de transferências ativas, $transf(t)$, entre o servidor de *streaming media* e uma certa quantidade de clientes. Este nível de concorrência pode ser utilizado como uma estimativa da carga do servidor no instante t .

A Figura 4.1 mostra a distribuição marginal do número de transferências simultâneas, associadas ao conteúdo ao vivo analisado, ao longo de toda a duração estudada do *log* de acesso. Essa característica pode ser modelada por uma distribuição Exponencial (ae^{-ax}) com parâmetro $a = 0.0029 \pm 1.48e^{-05}$. O coeficiente de correlação encontrado para essa aproximação foi $R^2 = 0.97$.

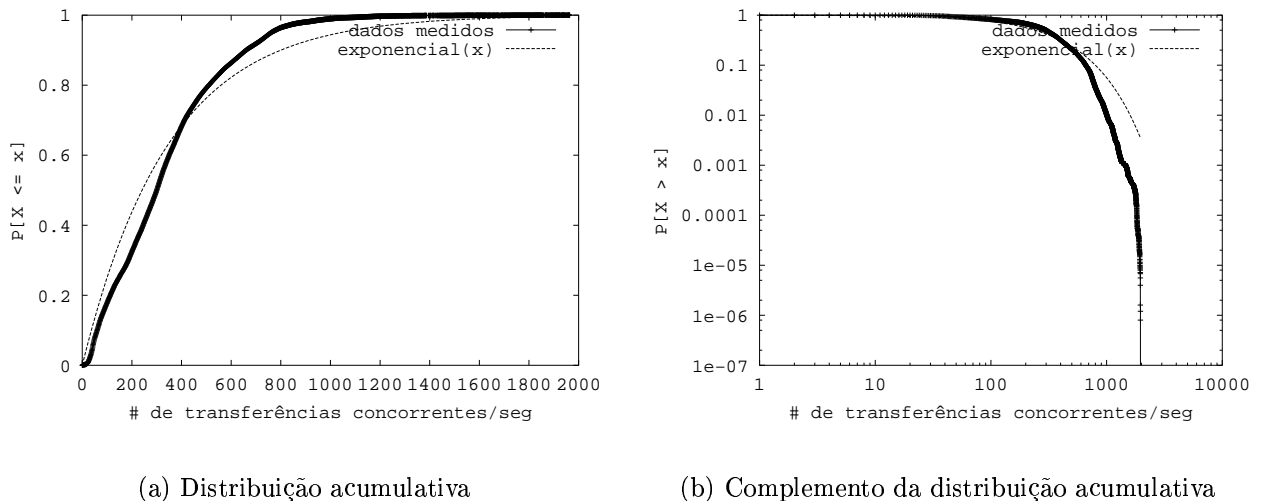
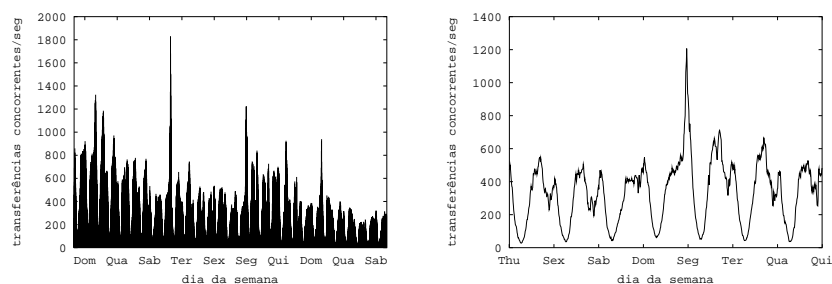


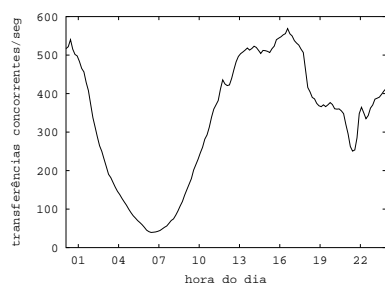
Figura 4.1: Distribuição marginal do número de transferências concorrentes associadas ao conteúdo ao vivo

Na Figura 4.2 apresentamos o comportamento temporal do número de transferências ativas associadas aos objetos ao vivo, $transf(t)$. A Figura 4.2(a) apresenta o número médio de transferências concorrentes por segundo, computado em *bins* de 10 minutos, ao longo de todo o período analisado do histórico de acesso. Nessa figura, podemos observar grande variabilidade no número de transferências simultâneas. Por sua vez, a Figura 4.2(b) evidencia o comportamento periódico de $transf(t)$, apresentando $transf(t \bmod p)$, em que p é uma semana. A Figura 4.2(c) mostra $transf(t \bmod p)$, em que t é um instante de tempo localizado em dia útil de semana e p é um dia. De forma análoga, as Figuras 4.2(d) e (e) apresentam $transf(t \bmod p)$ em que p é um dia e t é um instante de tempo localizado, respectivamente, em um sábado e em um domingo.

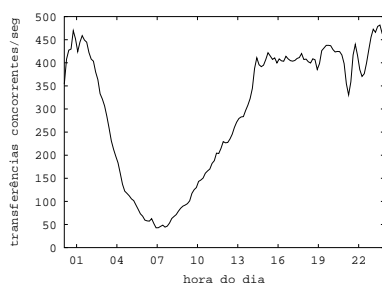


(a) Ao longo de todo o período analisado

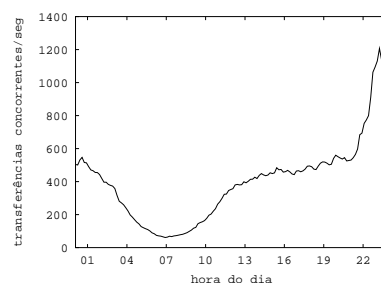
(b) $transf(t \bmod p)$, para uma semana



(c) $transf(t \bmod p)$, para dias úteis



(d) $transf(t \bmod p)$, para os sábados



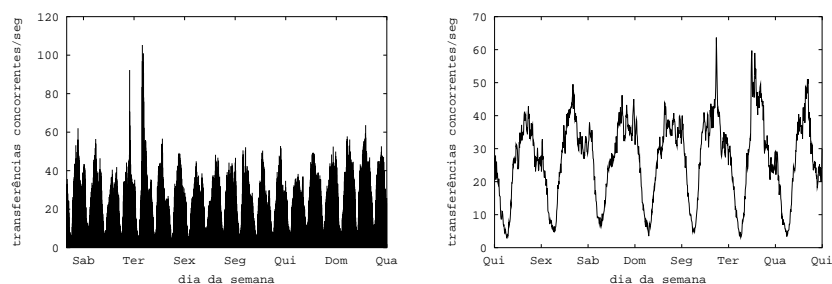
(e) $transf(t \bmod p)$, para os domingos

Figura 4.2: Comportamento temporal do número de transferências concorrentes, $transf(t)$, associadas aos objetos ao vivo

A Figura 4.3 apresenta as figuras correspondentes para objetos de *streaming media* pré-armazenados.

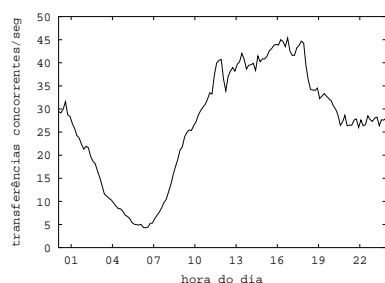
Devemos notar que diversos fatores podem contribuir para a grande variabilidade observada no número de transferências simultâneas associadas aos objetos ao vivo, incluindo atividades específicas que acontecem no *reality show* associado aos objetos de *streaming media* ao vivo, assim como efeitos diurnos do conteúdo transmitido e da população de clientes. As Figuras 4.2(b) e (e) mostram que o período das 21 às 24 horas de domingo apresenta um número consideravelmente maior de transferências concorrentes. Esse efeito não é observado nas figuras correspondentes para objetos pré-armazenados (Figuras 4.3(b) e (e), respectivamente). Como esse horário coincide com o momento em que um participante do *reality show* era eliminado, acreditamos que esse aumento no número de transferências ativas associadas aos objetos ao vivo, ocorra devido ao conteúdo ao vivo sendo transmitido.

As Figuras 4.2(c) e (d) evidenciam uma diminuição no número de transferências simultâneas no período entre as 21 e 22 horas dos dias de semana e no sábado. Mais uma vez, não observamos essa característica nas figuras correspondentes (Figuras 4.3(c) e (d))

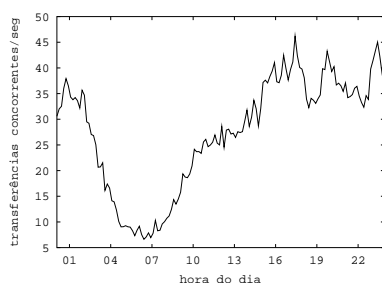


(a) Ao longo de todo o período analisado

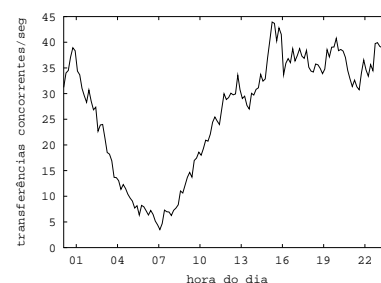
(b) $transf(t \bmod p)$, para uma semana



(c) $transf(t \bmod p)$, para os dias úteis



(d) $transf(t \bmod p)$, para os sábados



(e) $transf(t \bmod p)$, para os domingos

Figura 4.3: Comportamento temporal do número de transferências concorrentes, $transf(t)$, associadas aos objetos pré-armazenados

para objetos de *streaming media* pré-armazenados. Esse horário coincide com o momento em que um programa diário, associado ao *reality show*, era exibido na televisão. Esse programa de televisão exibia um resumo das atividades praticadas durante o dia pelos participantes da competição e não apresentava o mesmo conteúdo dos objetos ao vivo sendo transmitidos, no mesmo horário, através da *Internet*. Assim, acreditamos que a diminuição do número de transferências ativas nesse intervalo de tempo ocorra, mais uma vez, em virtude do conteúdo ao vivo sendo transmitido. Provavelmente, o cliente considerava mais interessante assistir ao programa de televisão.

Podemos observar também que o sistema, associado à transmissão dos objetos de *streaming media* ao vivo, apresenta um número consideravelmente menor de transferências concorrentes no período entre 1 e 11 horas. Esse efeito é analogamente observado nos acessos a objetos pré-armazenados. Assim, acreditamos que essa característica ocorra devido a fatores relacionados ao conteúdo ao vivo sendo disponibilizado (nesse período, os participantes do *reality show* estavam dormindo, o que provavelmente não era interessante para os espectadores) e a aspectos associados à população de clientes.

Por último, podemos observar que a Figura 4.2(c) apresenta uma diminuição no número de transferências simultâneas por volta das 18 horas. Essa mesma característica é observada na figura correspondente para objetos pré-armazenados (Figura 4.3(c)). Esse horário coincide com o término do expediente de trabalho diário e, por isso, provavelmente essa característica ocorre devido a fatores da população de clientes.

4.3.2 Tempo entre as Chegadas de Transferências Consecutivas

Seja $t(j)$ o instante de chegada da j -ésima transferência no histórico de acessos. Assim, $arrivalTransfer(j) = t(j+1) - t(j)$ denota o tempo decorrido entre as chegadas das transferências j e $j+1$.

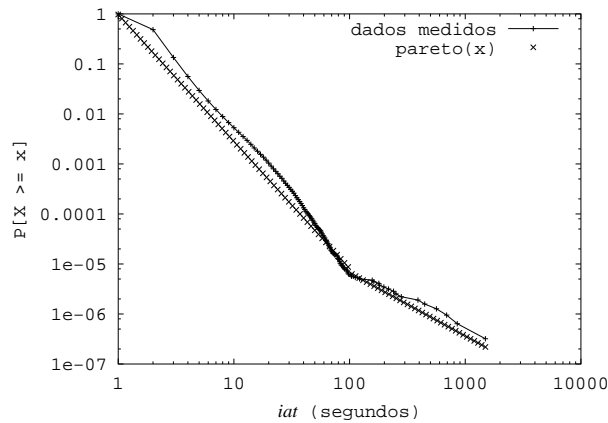


Figura 4.4: Complemento da distribuição acumulativa do iat entre transferências consecutivas

A Figura 4.4 apresenta o complemento da função de distribuição acumulativa de $arrivalTransfer(j)$ para acessos a objetos ao vivo. Essa figura sugere que a distribuição dos $iats$ de transferências consecutivas apresenta uma natureza cauda-pesada, com dois comportamentos distintos na cauda. Assim, aproximamos os $iats$ computados de acordo com o histórico de acessos analisado, de distribuições de Pareto (ab^a/x^{a+1}). O primeiro comportamento evidenciado na Figura 4.4 engloba os $iats$ de até 100 segundos e apresenta parâmetros $a = 2.5$ e $b = 0.98$. Por sua vez o segundo comportamento na cauda apresenta parâmetros $a = 1.26$ e $b = 0.008$ e compreende os $iats$ maiores que 100 segundos. Acreditamos que esses dois regimes correspondem a dois processos distintos de geração de requisições dos clientes, associados a requisições enviadas durante intervalos de tempo populares e impopulares, respectivamente.

Na Figura 4.5, mostramos o comportamento temporal do iat de solicitações consecutivas por transferências dos objetos ao vivo. O processo de chegada de requisições é claramente não estacionário, como evidencia a Figura 4.5(a) que apresenta o iat médio,

percebido no instante t ($iatTransf(t)$), ao longo de toda a duração analisada do *log* de acesso. Para estudarmos a natureza periódica do *iat* de requisições consecutivas, calculamos $iatTransf(t \bmod p)$, em que p é uma semana, e o resultado é mostrado na Figura 4.5(b). As Figuras 4.5(c), (d) e (e) apresentam $iatTransf(t \bmod p)$, em que p é um dia e t corresponde a um instante de tempo localizado, respectivamente, em um dia útil de semana, em um sábado e em um domingo. Esses gráficos foram obtidos calculando-se o tempo médio entre chegadas de requisições durante períodos consecutivos de 10 minutos.

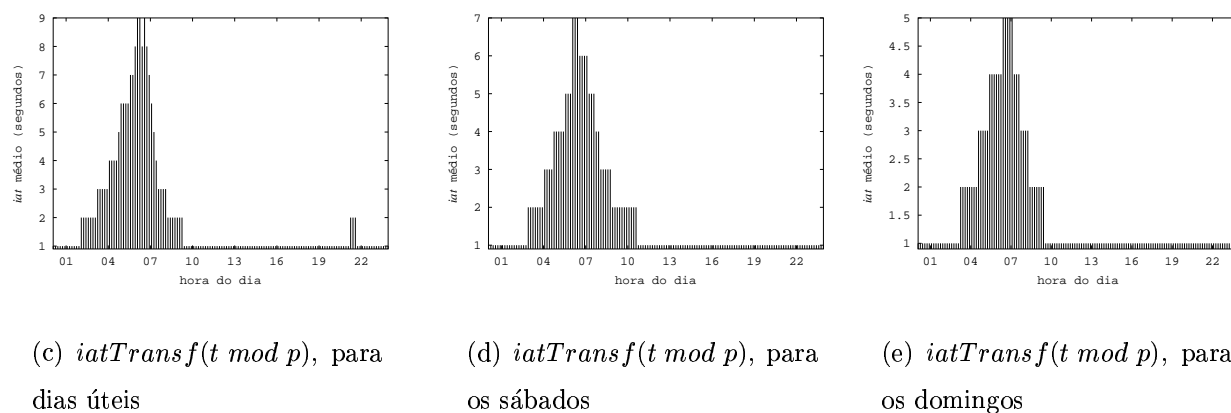
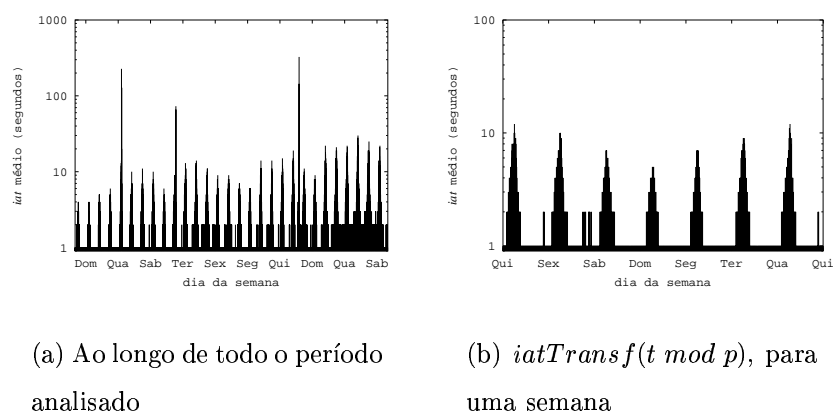


Figura 4.5: Comportamento temporal do *iat* entre transferências consecutivas

Como mostrado na Figura 4.5(b), o *iat* de transferências apresenta algumas variações de acordo com o dia da semana, com os finais de semana apresentando *iats* médios um pouco menores do que os observados nos dias úteis. No entanto, como mostrado nas Figuras 4.5(c), (d) e (e), o comportamento diário é a principal fonte de variabilidade, com o período de 2 às 10 apresentando *iats* mais longos. Acreditamos que essa característica ocorra devido a fatores relacionados à população de clientes e ao conteúdo ao vivo disponibilizado, pois esse período coincide com o horário em que os participantes do *reality show* dormem, despertando, portanto, pouco interesse nos clientes.

Além disso, como evidencia a Figura 4.5(c), o *iat* médio percebido nos dias úteis, no

período entre as 21 e 22 horas, apresenta um valor um pouco maior. Como esse período coincide com o horário em que o programa diário associado ao *reality show*, era exibido na televisão, acreditamos que essa característica ocorra devido ao conteúdo ao vivo sendo transmitido através da *Internet*.

4.3.3 Distribuição da Duração das Transferências

Nesta seção analisamos a duração de transferências individuais de objetos ao vivo. É importante observar que a duração de transferências individuais não necessariamente corresponde ao período *ON* de transferência, caracterizado na Seção 4.4.2, uma vez que esse último pode resultar da sobreposição de várias transferências individuais, como mostrado na Figura 3.1. Dada a natureza de tempo real de transmissões ao vivo, a caracterização da duração de transferências individuais em segundos, é bastante apropriada. Converter as características para *bytes* transmitidos seria uma função da taxa de transferência, que será analisada na Seção 4.3.4.

Seja $l(j)$ a duração, em segundos, da j -ésima transferência no registro de acessos. As Figuras 4.6(a) e (b) apresentam, respectivamente, o histograma e a distribuição acumulativa de $l(j)$. Aproximamos os valores computados para $l(j)$, de acordo com o *log* de acesso, de uma distribuição Lognormal ($p(x) = (\sqrt{2\pi}Sx)^{-1}e^{-(\ln(x)-M)^2/(2S^2)}$), com parâmetros $M = 4.292 \pm 0.0012$ e $S = 1.282 \pm 0.0010$. O coeficiente de correlação encontrado para essa aproximação foi $R^2 = 0.99$.

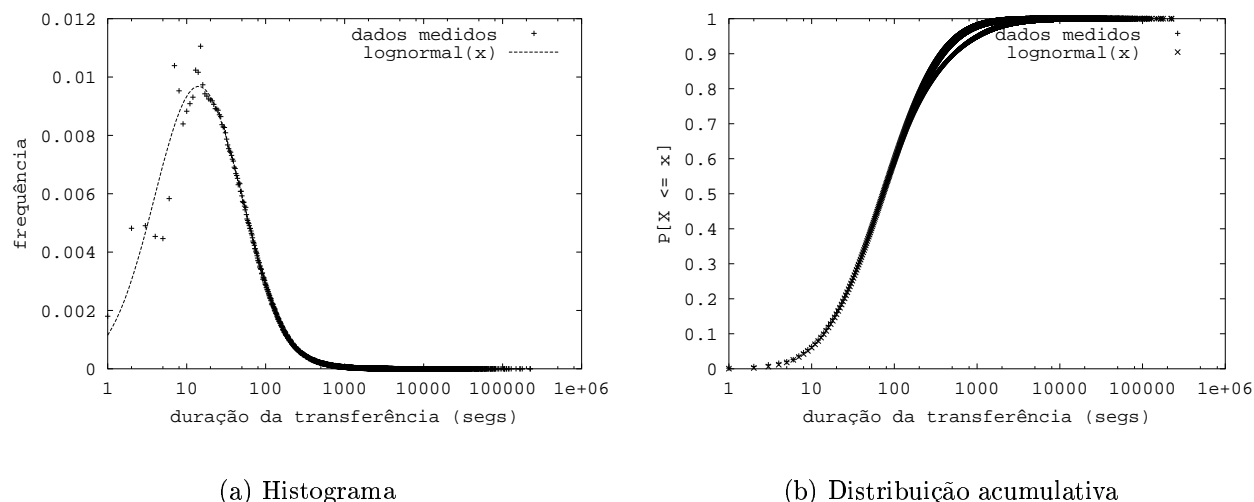


Figura 4.6: Distribuição marginal da duração das transferências associadas ao conteúdo ao vivo

A distribuição de tamanho de transferências *unicast*, através da *Internet*, tem sido extensivamente estudada na literatura, devido aos possíveis impactos que a distribuição de

tal variável pode causar nas características do tráfego gerado na rede e nos servidores. Em [13], Crovella e Bestavros argumentaram que as causas da auto-similaridade observada no tráfego na *Internet* poderiam estar relacionadas à natureza cauda-pesada de transferências individuais de arquivos, as quais, por sua vez, poderiam ter sua origem na natureza cauda-pesada da distribuição de tamanho dos arquivos disponíveis. Trabalhos mais recentes como [15] e [20], a respeito da verdadeira natureza da distribuição de tamanho de arquivos (se esta pode ser modelada por uma distribuição Lognormal, de Pareto ou *Double Pareto*), evidenciam a importância de caracterizar de maneira acurada, e compreender as verdadeiras causas das características das distribuições de duração das transferências.

Para cargas de trabalho de mídia de fluxo contínuo de dados ao vivo, a grande variabilidade da distribuição da duração de transferências individuais é bastante interessante, pois não resulta da distribuição de tamanho dos objetos disponíveis, mas da disposição do cliente em assistir ao objeto ao vivo sendo transmitido. Devemos lembrar que, para objetos ao vivo, uma transferência é o resultado de um par de requisições enviadas pelo cliente ao servidor de *streaming media*. Especificamente, essas requisições solicitam que o objeto comece e, eventualmente, pare de ser transmitido para o cliente. Assim, a duração de uma transferência é o tempo decorrido entre essas duas ações do cliente. Desta forma, a alta variabilidade na duração de transferências individuais de *streams* deve-se ao comportamento do cliente, e não às características do tamanho do objeto. É importante observar que, para objetos de *streaming media* pré-armazenados, tanto a distribuição de tamanho dos objetos, quanto a interatividade do cliente, contribuem na determinação da duração das transferências.

4.3.4 Largura de Banda Utilizada pelas Transferências

Nesta seção analisamos a largura de banda média utilizada pelas transferências associadas a objetos de *streaming media* ao vivo. O valor reportado no *log* de acesso estudado corresponde à largura de banda média efetivamente utilizada pela conexão entre o cliente e o servidor de *streaming media*. Assim, esse valor é determinado pela velocidade do equipamento que o cliente utiliza para conectar-se à *Internet*, pelas condições da rede e do servidor de *streaming media* e, finalmente, pela taxa de codificação do *stream*.

A Figura 4.7 apresenta a distribuição da largura de banda média (em *bits/segundo*) experimentada por todas as transferências presentes no histórico de acessos. A Figura 4.7(a) mostra o histograma da largura de banda média utilizada e evidencia, claramente, duas situações distintas. A primeira é exemplificada pelos picos, na porção mais à direita desta figura, que correspondem a valores de largura de banda *client-bound*, que são determinados primariamente pelas velocidades das conexões dos clientes com a *Internet*. Podemos identificar, por exemplo, diversas velocidades de *modem*, DSL, *cable modem*, etc. A segunda

situação é exemplificada por valores muito menores de largura de banda média, que aparecem na parte mais à esquerda da distribuição e que correspondem a valores de largura de banda *congestion-bound*, que resultam de recursos, extremamente limitados, da rede e/ou do servidor de *streaming media*. A Figura 4.7(b) apresenta a distribuição acumulativa da largura de banda média utilizada. Essa figura sugere que aproximadamente 10% de todas as transferências foram *congestion-bound*. Como foi observado na Seção 4.2, momentos em que o servidor de *streaming media* mostrou-se sobrecarregado foram extremamente raros, indicando que esta não é a causa principal das transferências *congestion-bound*.

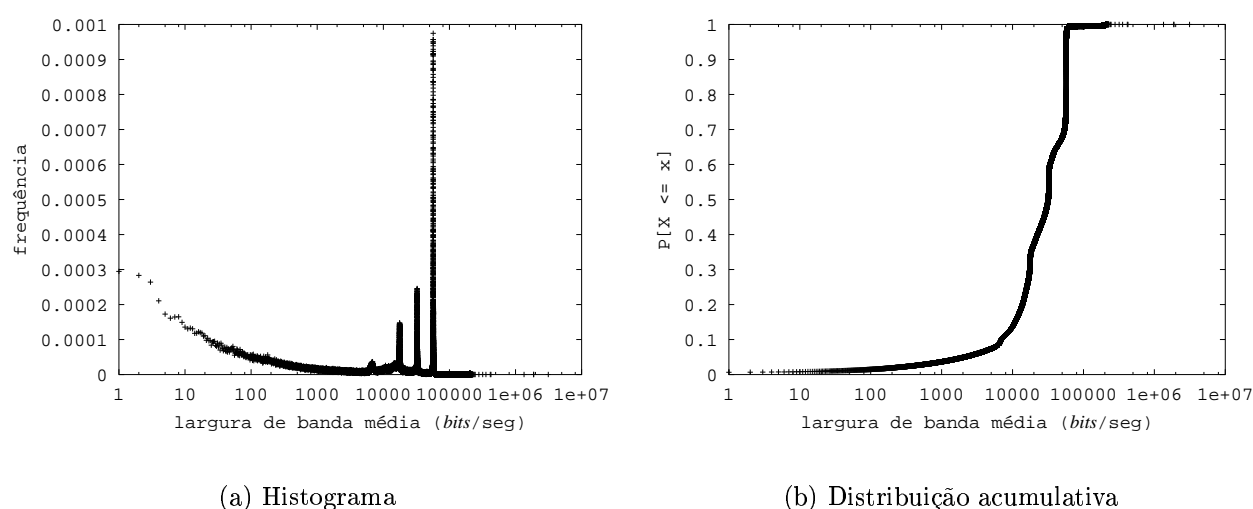


Figura 4.7: Distribuição da largura de banda média utilizada

A Figura 4.8 apresenta o comportamento temporal da largura de banda total utilizada pelas transferências associadas a *streams* ao vivo. Nesta figura, mostramos a quantidade total de largura de banda utilizada, calculada em *bins* de 10 minutos, ao longo de todo o período analisado do *log* de acesso.

4.4 Características da Camada de Sessão

Nesta seção, apresentamos as diversas características relevantes observadas dentro de uma mesma sessão, incluindo períodos *ON* de transferência, períodos *OFF* ativos, *iat* de transferências associadas a uma mesma sessão e o número de transferências pertencentes a uma mesma sessão.

4.4.1 Número de Sessões

Já que o histórico de acessos não identifica, explicitamente, os limites de uma dada sessão, o número de sessões identificadas depende do valor que escolhermos para o patamar

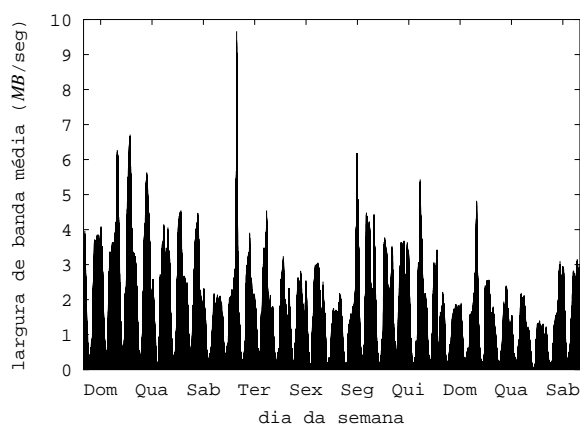


Figura 4.8: Comportamento temporal da largura de banda total utilizada por transferências associadas a objetos ao vivo

T_{off} , que representa o *timeout* de sessão. A Figura 4.9 apresenta a relação entre o número de sessões distintas identificadas no *log* de acesso a objetos ao vivo e o valor escolhido para o patamar T_{off} . Esta relação demonstra que o número de sessões identificadas não se altera drasticamente para $T_{off} > 3600$ segundos. Além disso, observamos claramente um pico, na distribuição de períodos *OFF* inativos, em torno de 3600 segundos.¹ Assim, no restante deste capítulo, utilizamos o patamar $T_{off} = 3600$ segundos para identificar diferentes sessões por objetos ao vivo.

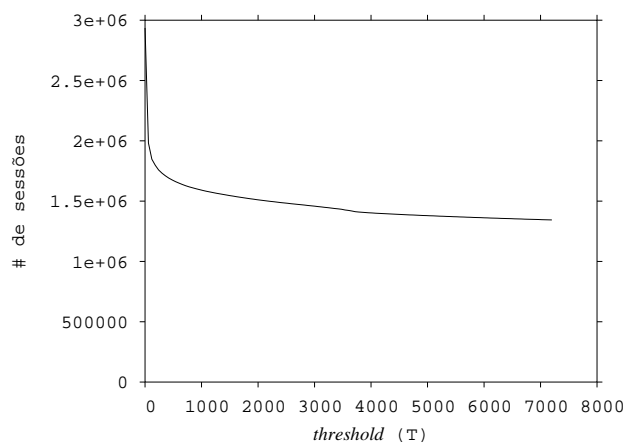


Figura 4.9: Relação entre o número de sessões distintas identificadas no *log* de acesso a objetos ao vivo e T_{off}

A Figura 4.10 demonstra esse mesmo estudo para o histórico de acessos a objetos pré-armazenados. Como podemos observar, o número de sessões associadas à transmissão de objetos pré-armazenados não se altera consideravelmente para $T_{off} > 1800$ segundos. Dessa forma, no restante desse documento, utilizamos o patamar $T_{off} = 1800$ segundos

¹Esta característica está representada na Figura 4.24.

para identificar diferentes sessões associadas a objetos pré-armazenados.

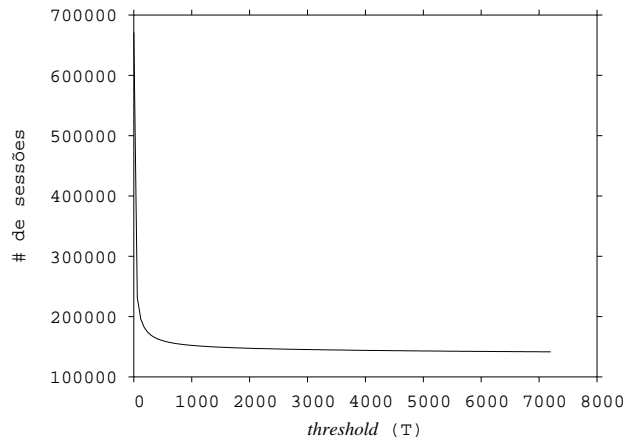


Figura 4.10: Relação entre o número de sessões distintas identificadas no *log* de acesso a objetos pré-armazenados e T_{off}

4.4.2 Distribuição do Período *ON* de Transferência

Nesta seção analisamos o período *ON* de transferência. É importante observar que o período *ON* de transferência não corresponde, necessariamente, à duração de transferências individuais, caracterizada na Seção 4.3.3, uma vez que pode resultar da sobreposição de várias transferências individuais associadas a um mesmo cliente, como mostrado na Figura 3.1. Assim, a duração do período *ON* de transferência corresponde à união das durações das transferências individuais simultâneas associadas a um mesmo cliente. As Figuras 4.11(a) e (b) apresentam, respectivamente, o histograma e a distribuição acumulativa do período *ON* de transferência. Aproximamos os valores computados para o período *ON* de transferência de acordo com o *log* de acesso, de uma distribuição Log-normal ($p(x) = (\sqrt{2\pi}Sx)^{-1}e^{-(\ln(x)-M)^2/(2S^2)}$), com parâmetros $M = 4.291 \pm 0.0011$ e $S = 1.278 \pm 0.0009$. O coeficiente de correlação encontrado para essa aproximação foi $R^2 = 0.99$. Não é surpreendente que a distribuição do período *ON* de transferência seja muito semelhante à distribuição da duração de transferências individuais, uma vez que é obtido através da união das durações das transferências individuais simultâneas associadas a um mesmo cliente.

4.4.3 Distribuição do Período *OFF* Ativo

Nesta seção caracterizamos a duração dos períodos de inatividade de um cliente dentro de uma sessão, também chamados de *OFF* ativo ou *think time*. Durante intervalos de *OFF* ativo, nenhum objeto ao vivo é transmitido ao cliente e a duração desses períodos de inatividade, por definição, deve ser menor que T_{off} .

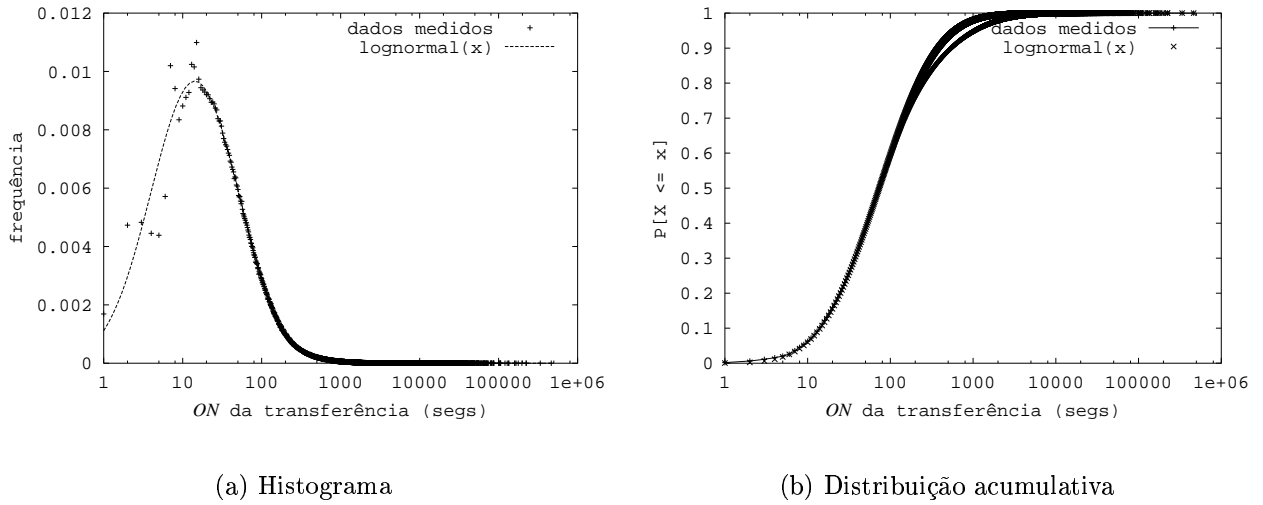


Figura 4.11: Distribuição marginal dos períodos *ON* de transferência associados ao conteúdo ao vivo

Sejam i e j duas transferências consecutivas que pertencem a uma mesma sessão, sendo que i iniciou-se no instante de tempo $t(i)$ e j , no instante $t(j)$. Além disso, devemos recordar que $l(i)$ representa a duração da transferência i . Assim, $off\ Ativo(i) = t(j) - (t(i) + l(i))$ representa o período *OFF* ativo entre as transferências i e j . A Figura 4.12 mostra o histograma dos períodos *OFF* ativos. Aproximamos os valores computados para o *think time* dos clientes, de uma distribuição Lognormal ($p(x) = (\sqrt{2\pi}Sx)^{-1}e^{-(\ln(x)-M)^2/(2S^2)}$), com parâmetros $M = 3.462 \pm 0.0053$ e $S = 0.887 \pm 0.0044$. O coeficiente de correlação encontrado para essa aproximação foi $R^2 = 0.90$.

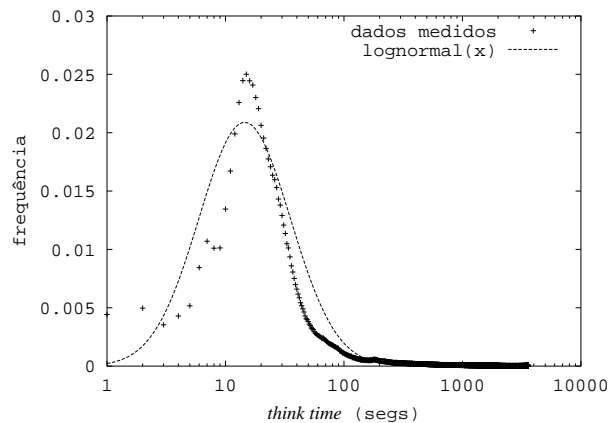


Figura 4.12: Histograma da duração do *think time* dos clientes

4.4.4 Número de Transferências por Sessão

O número de transferências dentro de uma mesma sessão evidencia a atividade continuada do cliente associado a essa sessão. A Figura 4.13 apresenta a distribuição marginal do número total de transferências dentro de cada sessão identificada no *log* de acesso. A distribuição resultante caracteriza um comportamento de natureza cauda-pesada, que pode ser modelado por uma distribuição de Pareto ($(ab^a)/x^{a+1}$), com parâmetros $a = 1.432 \pm 0.0098$ e $b = 0.624 \pm 0.0034$. O coeficiente de correlação encontrado para essa aproximação foi $R^2 = 0.99$.

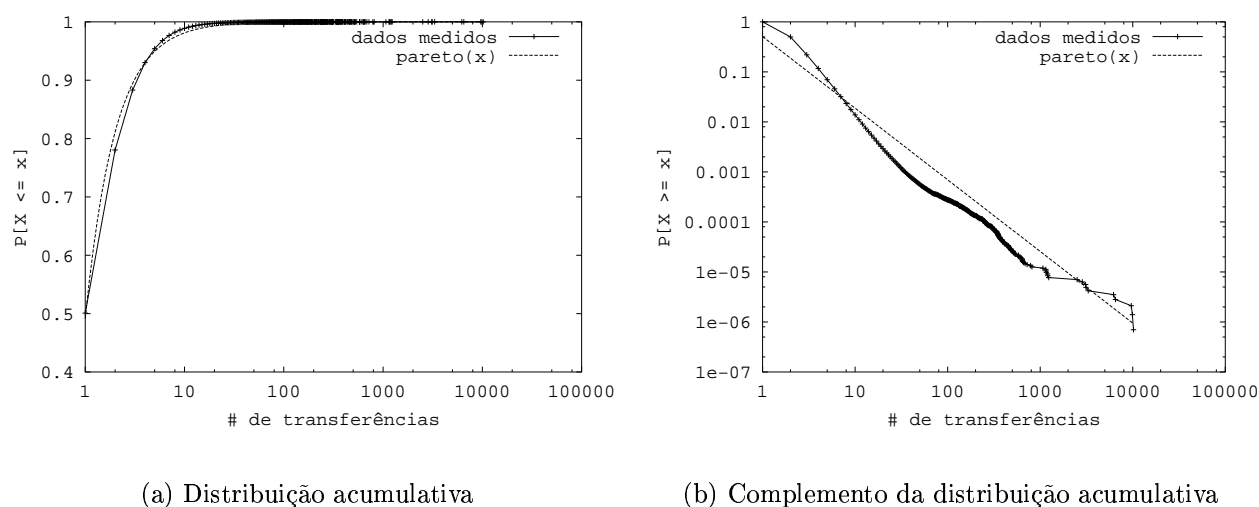


Figura 4.13: Distribuição marginal do número de transferências dentro de uma mesma sessão

Estudamos também a correlação entre o horário do dia e o número de transferências associadas a cada sessão, mas concluímos que a variabilidade no número de transferências por sessão não é fortemente associada a características temporais. Então, atribuímos essa variabilidade à natureza das interações dos clientes com o conteúdo ao vivo.

4.4.5 Tempo entre as Chegadas de Transferências Consecutivas dentro de uma mesma Sessão

A última variável que caracterizamos na camada de sessão corresponde ao *iat* de transferências consecutivas pertencentes a uma mesma sessão de cliente. A Figura 4.14 apresenta a distribuição marginal associada a essa característica.

A distribuição apresentada na Figura 4.14 pode ser modelada por uma distribuição Lognormal ($p(x) = (\sqrt{2\pi}Sx)^{-1}e^{-(\ln(x)-M)^2/(2S^2)}$) com parâmetros $M = 4.930 \pm 0.0036$ e $S = 1.264 \pm 0.0030$. O coeficiente de correlação encontrado para essas duas distribuições

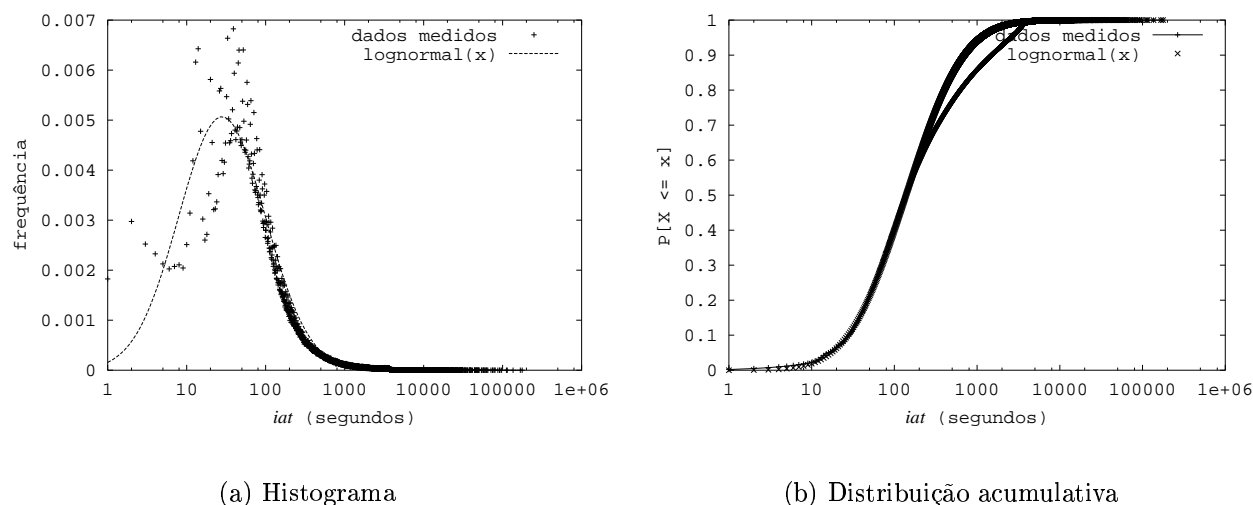


Figura 4.14: Distribuição marginal do *iat* de transferências consecutivas dentro de uma mesma sessão

foi $R^2 = 0.95$.

4.5 Características da Camada de Cliente

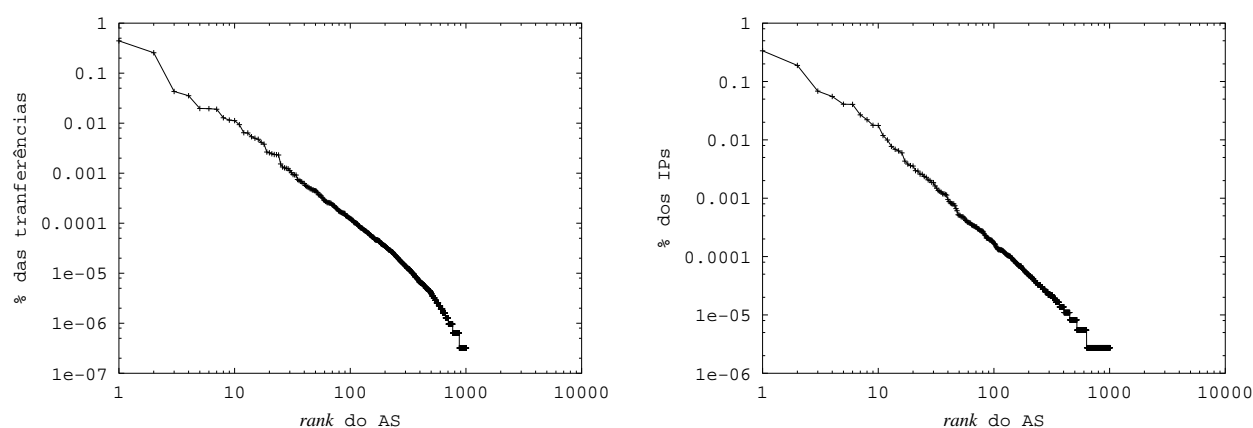
Nesta seção apresentamos diversas características relacionadas ao comportamento dos clientes, incluindo o número de clientes ativos ao longo do tempo, ou seja, o nível de concorrência observado na carga de trabalho analisada, a relação entre a frequência de acesso e o interesse do cliente no serviço de *streaming media* ao vivo, a duração de períodos *ON* de sessão e *OFF* inativo, assim como outras características relacionadas à população de clientes em geral.

4.5.1 Distribuição Topológica e Geográfica dos Clientes

Um importante atributo que é frequentemente questionado quando analisamos estudos sobre a caracterização de cargas de trabalho, relaciona-se com a representatividade da carga analisada. Os valores reportados na Tabela 4.1 demonstram que a carga de trabalho gerada em função dos acessos a *streams* ao vivo e caracterizada neste capítulo é bastante ampla, tanto em termos do número de clientes distintos identificados (através do identificador do *software* instalado na máquina do cliente para acessar objetos de *streaming media*), quanto do número de acessos realizados por esses clientes.

Utilizando o endereço IP do cliente em uma dada sessão, fomos capazes de mapear a população de clientes em mais de 1000 *Internet Autonomous Systems* (AS'es) diferentes, espalhados por diversos países. A Figura 4.15 mostra a popularidade de cada AS identifi-

cado em nossa carga de trabalho. A Figura 4.15(a) apresenta o *rank* do AS em função do número de transferências a ele associado. Essa distribuição pode ser modelada por uma lei de *Zipf* com parâmetro $\alpha = 1.496$. O coeficiente de correlação encontrado para essa aproximação foi $R^2 = 0.95$. A Figura 4.15(b) mostra a popularidade do AS em relação ao número de endereços IPs, presentes no *log* de acesso, a ele associados. Também neste caso, aproximamos a distribuição de popularidade dos AS'es a uma lei de *Zipf* com parâmetro $\alpha = 1.302$. O coeficiente de correlação dessa aproximação é $R^2 = 0.97$. Devemos ressaltar que não fomos capazes de identificar o AS associado a 5% dos endereços IPs encontrados no *log* de acesso.



(a) Distribuição das transferências pelos AS'es

(b) Distribuição dos endereços IPs pelos AS'es

Figura 4.15: Diversidade da população de clientes

A Figura 4.16 apresenta a distribuição de transferências pelos vários países de onde se originaram acessos aos *streams* ao vivo analisados. Uma lei de *Zipf* com parâmetro $\alpha = 4.334$ pode ser utilizada para modelar essa distribuição. O coeficiente de correlação entre essas curvas é $R^2 = 0.99$.

4.5.2 Número de Clientes Concorrentes

Em qualquer instante de tempo t , existe um determinado número de clientes, $c(t)$, considerados ativos, no sentido de que suas sessões ainda não terminaram. Este nível de concorrência pode ser utilizado como uma estimativa da popularidade do conteúdo específico que está sendo transmitido no instante t .

A Figura 4.17 apresenta a distribuição marginal de $c(t)$ ao longo de toda a duração analisada do histórico de acessos aos objetos de *streaming media* ao vivo. Essa característica pode ser modelada por uma distribuição Exponencial (ae^{-ax}) com parâmetro $a = 0.0019 \pm 6.951e^{-06}$. O coeficiente de correlação encontrado para essa aproximação foi $R^2 = 0.98$.

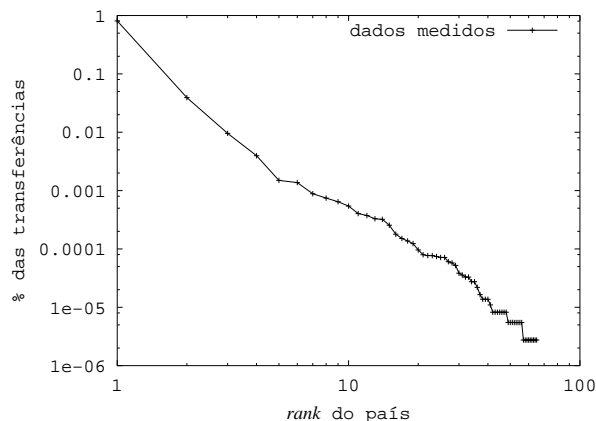
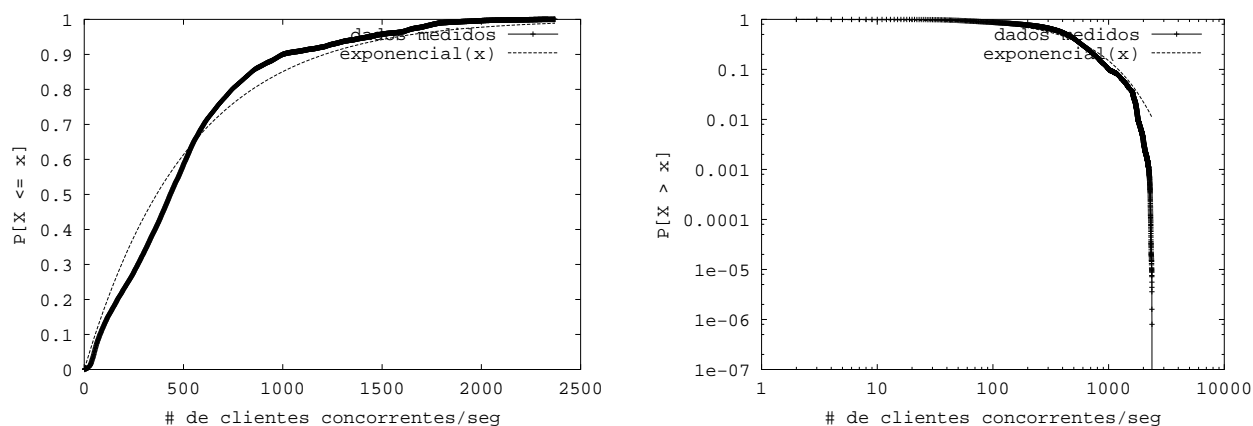


Figura 4.16: Distribuição das transferências pelos países



(a) Distribuição acumulativa

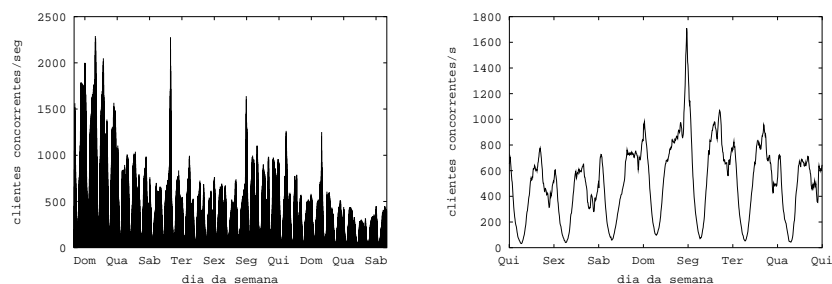
(b) Complemento da distribuição acumulativa

Figura 4.17: Distribuição marginal do número de clientes ativos acessando *streams* ao vivo

A Figura 4.18 apresenta o comportamento temporal do número de clientes ativos acessando os objetos de *streaming media* ao vivo. A Figura 4.18(a) apresenta o número médio de clientes concorrentes por segundo, computado em *bins* de 10 minutos, ao longo de todo o período analisado do histórico de acessos. Nessa figura, podemos observar grande variabilidade no número de clientes ativos. Por sua vez, a Figura 4.18(b) evidencia o comportamento periódico de $c(t)$, apresentando $c(t \bmod p)$, em que p é uma semana. A Figura 4.18(c) mostra $c(t \bmod p)$, em que t é um instante de tempo localizado em dia útil de semana e p é um dia. De forma análoga, as Figuras 4.18(d) e (e) apresentam $c(t \bmod p)$ em que p é um dia e t é um instante de tempo localizado, respectivamente, em um sábado e em um domingo.

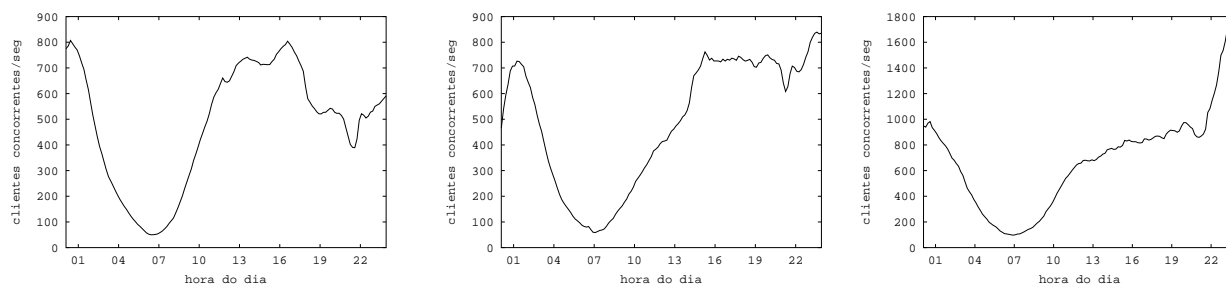
A Figura 4.19 apresenta as mesmas informações da Figura 4.18 associadas, no entanto, aos acessos a objetos de *streaming media* pré-armazenados.

Como já era esperado, essas distribuições são muito similares às observadas para o



(a) Ao longo de todo o histórico de acessos

(b) $c(t \bmod p)$, para uma semana



(c) $c(t \bmod p)$, para os dias úteis

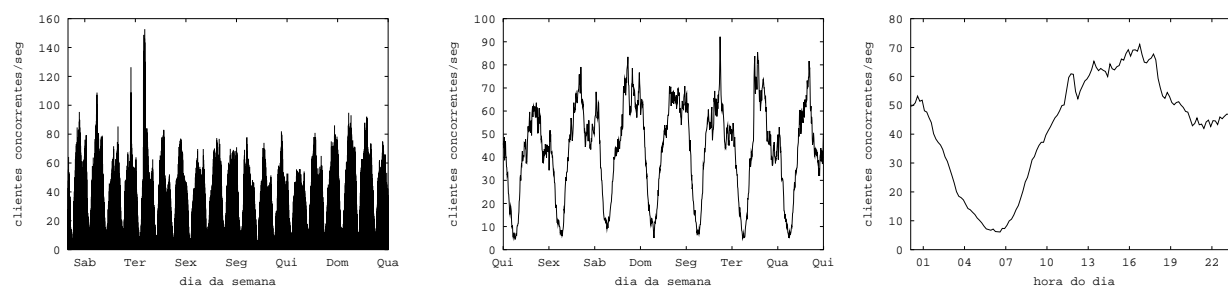
(d) $c(t \bmod p)$, para os sábados

(e) $c(t \bmod p)$, para os domingos

Figura 4.18: Comportamento temporal do número de clientes ativos acessando *streams* ao vivo

número de transferências concorrentes ao longo do tempo relacionadas ao acesso a *streams* ao vivo (Figura 4.2) e ao acesso a objetos de *streaming media* pré-armazenados (Figura 4.3). Assim como observamos na Seção 4.3.1, para o número de transferências concorrentes ao longo do tempo, podemos notar que existem diversos fatores que contribuem para a grande variabilidade observada no número de clientes ativos no sistema. Dentre esses fatores podemos destacar atividades específicas que ocorrem no *reality show* associado aos objetos ao vivo analisados como, por exemplo, a seleção de um participante para ser eliminado do jogo nas noites de domingo; a exibição, na televisão, entre as 21 e 22 horas de todos os dias de semana e do sábado, de um resumo de todas as atividades praticadas pelos participantes da competição durante o dia; e a transmissão de nenhuma atividade interessante dos competidores entre 1 e 11 horas de todos os dias. Além disso, aspectos relacionados à população de clientes também contribuem para a variabilidade observada.

Para quantificar a correlação temporal entre o número de clientes ativos nos vários momentos do dia e nos diversos dias de *log* analisados, calculamos a função de auto-correlação de $c(t)$, para vários deslocamentos d . A Figura 4.20 apresenta os resultados encontrados.



(a) Ao longo de todo o período analisado

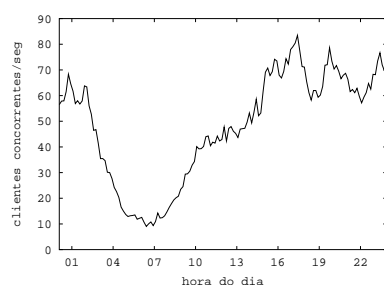
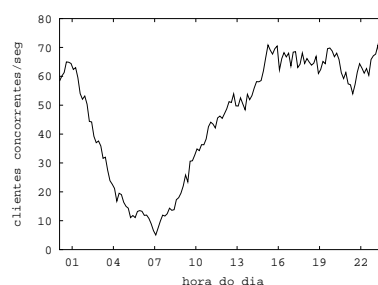
(b) $c(t \bmod p)$, para uma semana(c) $c(t \bmod p)$, para os dias úteis(d) $c(t \bmod p)$, para os sábados(e) $c(t \bmod p)$, para os domingos

Figura 4.19: Comportamento temporal do número de clientes ativos acessando objetos pré-armazenados

Podemos claramente observar nesta figura uma periodicidade diária, com picos em valores de deslocamento $d = 1440, 2880, \dots$ minutos, que são múltiplos de 1440, que é o número de minutos de um dia.

4.5.3 Tempo entre as Chegadas de Clientes Consecutivos

Seja $t(i)$ o instante de chegada da i -ésima sessão identificada no *log* de acessos analisado. Assim, o tempo decorrido entre as chegadas da i -ésima e da $(i + 1)$ -ésima sessões é dado por $arrivalClient(i) = t(i + 1) - t(i)$. As sessões i e $i + 1$ podem pertencer a clientes distintos. Claramente, $arrivalClient(i)$ é uma série temporal que descreve o *iat* de clientes consecutivos.

A Figura 4.21 apresenta o complemento da distribuição acumulativa de $arrivalClient(i)$, para acessos a objetos ao vivo. Essa figura sugere que a distribuição dos *iats* de clientes consecutivos apresenta uma natureza cauda-pesada, com dois comportamentos distintos na cauda, assim como foi observado para a distribuição dos *iats* de transferências consecutivas. Aproximamos os *iats* de clientes consecutivos de distribuições de Pareto (ab^a/x^{a+1}) .

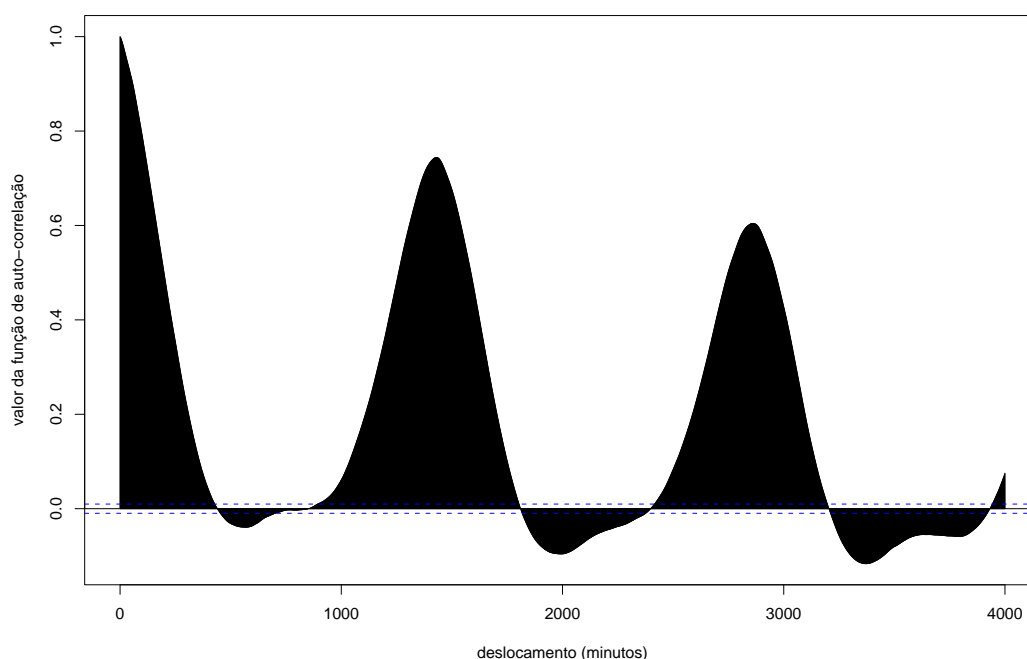


Figura 4.20: Função de auto-correlação do número de clientes ativos ao longo do tempo

O primeiro comportamento evidenciado na Figura 4.21 engloba os *iats* de até 200 segundos e apresenta parâmetros $a = 2.52$ e $b = 1.55$. O segundo comportamento na cauda apresenta parâmetros $a = 0.76$ e $b = 2.09872e^{-05}$ e compreende os *iats* maiores que 200 segundos. Acreditamos que esses dois regimes correspondem a dois processos distintos de chegada de clientes, associados a sessões iniciadas durante intervalos de tempo populares e impopulares, respectivamente.

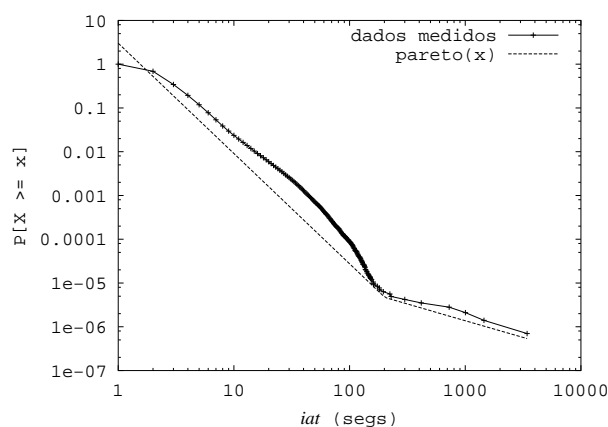


Figura 4.21: Complemento da distribuição acumulativa do tempo decorrido entre as chegadas de clientes consecutivos

Na Figura 4.22, mostramos o comportamento temporal do *iat* de clientes consecutivos

que acessam *streams* ao vivo. A Figura 4.22(a) apresenta o *iat* médio de clientes consecutivos, percebido no instante t ($iatClient(t)$), ao longo de toda a duração analisada do *log* de acessos. Por sua vez, as Figuras 4.22(b), (c) e (d) apresentam $iatClient(t \bmod p)$, em que p é um dia e t corresponde a um instante de tempo localizado, respectivamente, em um dia útil de semana, em um sábado e em um domingo. Esses gráficos foram obtidos calculando-se o tempo médio entre as chegadas de clientes durante períodos consecutivos de 10 minutos.

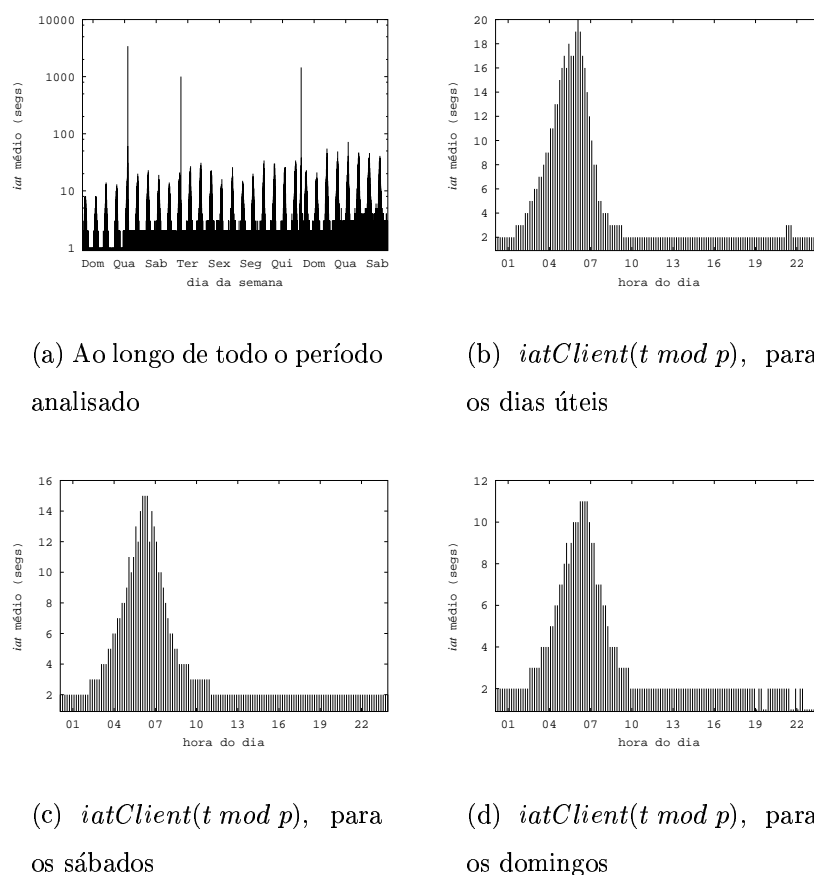


Figura 4.22: Comportamento temporal do *iat* de clientes consecutivos acessando *streams* ao vivo

Assim como foi observado para o tempo decorrido entre as chegadas de transferências consecutivas, o *iat* de clientes consecutivos apresenta grande variabilidade. Mais uma vez, acreditamos que essa característica ocorra devido a fatores relacionados à população de clientes e ao conteúdo ao vivo disponibilizado.

4.5.4 Distribuição do Período *ON* de Sessão

Seja $o(i)$ a duração, em segundos, da i -ésima sessão identificada no *log* de acessos. Assim, $on(i)$ denota o período *ON* da sessão i . A Figura 4.23 apresenta a distribuição marginal

de $on(i)$, para todas as sessões identificadas no histórico de acessos analisado. Essa distribuição foi aproximada de uma distribuição Lognormal ($p(x) = (\sqrt{2\pi}Sx)^{-1}e^{-(\ln(x)-M)^2/(2S^2)}$) com parâmetros $M = 5.193 \pm 0.0014$ e $S = 1.440 \pm 0.0011$. O coeficiente de correlação encontrado para essa aproximação foi $R^2 = 0.99$.

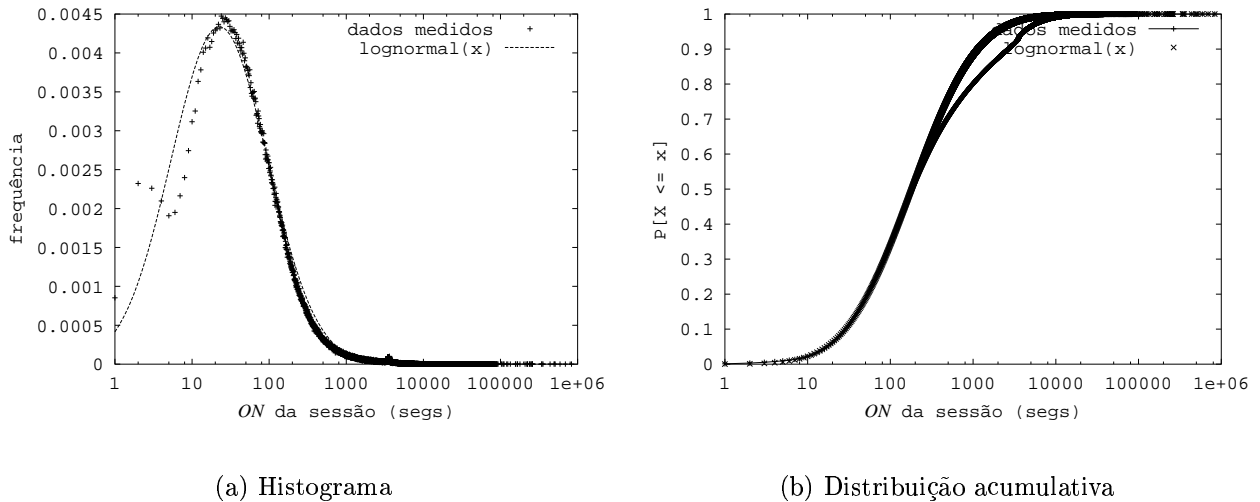


Figura 4.23: Distribuição marginal do período ON de sessão

A Figura 4.23 indica que os períodos ON de sessão são altamente variáveis. Com o intuito de determinar se esta variabilidade devia-se fundamentalmente à natureza das interações dos clientes com o conteúdo ao vivo sendo transmitido ou se estas variações estavam relacionadas a correlações temporais (como observamos para o nível de concorrência de clientes), caracterizamos a correlação entre o período ON de sessão e o momento do dia em que a sessão havia se iniciado. No entanto, observamos correlações muito fracas entre estas duas variáveis. Este resultado sugere que a alta variabilidade observada para o período ON de sessão não se deve a comportamentos temporais, mas é uma propriedade derivada da natureza das interações dos clientes com os *streams* ao vivo.

4.5.5 Distribuição do Período OFF Inativo

Nesta seção caracterizamos a duração dos períodos de inatividade de um cliente, entre duas de suas sessões consecutivas, também chamados de períodos OFF inativos. Durante intervalos de OFF inativo, o cliente não apresenta nenhuma sessão ativa e a duração desses períodos de inatividade, por definição, deve ser maior que T_{off} .

Sejam i e j duas sessões consecutivas que pertencem a um mesmo cliente, sendo que i iniciou-se no instante de tempo $t(i)$ e j , no instante $t(j)$. Além disso, devemos recordar que $on(i)$ representa a duração, ou período ON , da sessão i . Assim, $offInativo(i) = t(j) - (t(i) + on(i))$ representa o período de inatividade do cliente entre as sessões i e j .

A Figura 4.24 mostra a distribuição marginal dos períodos *OFF* inativos para todas as sessões identificadas no histórico de acessos. Aproximamos os valores computados para o *OFF* inativo dos clientes, de uma distribuição Exponencial (ae^{-ax}), com parâmetro $a = 5.025e^{-06} \pm 2.161e^{-09}$. O coeficiente de correlação encontrado para essa aproximação foi $R^2 = 0.92$.

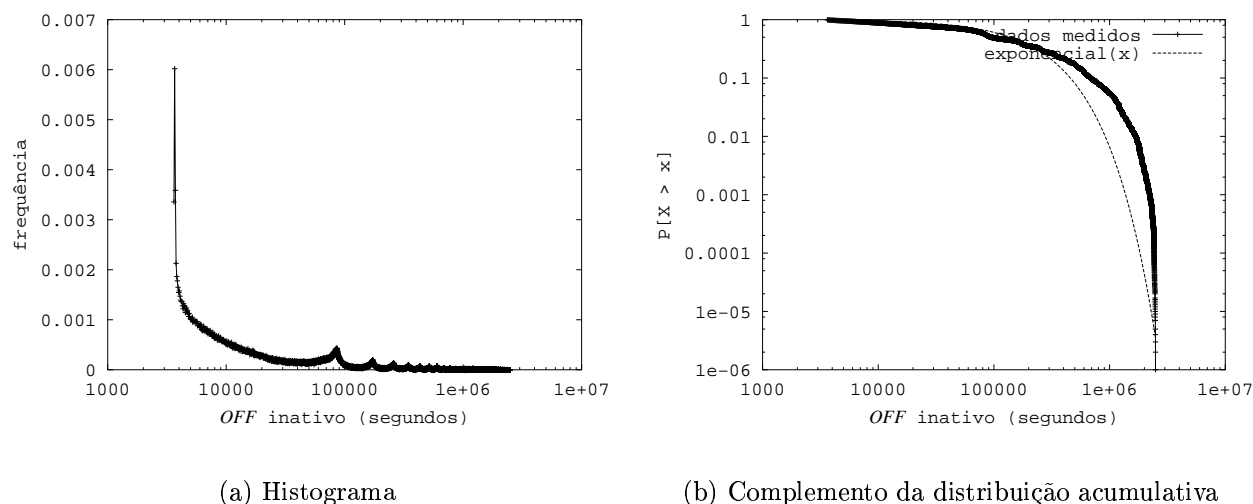


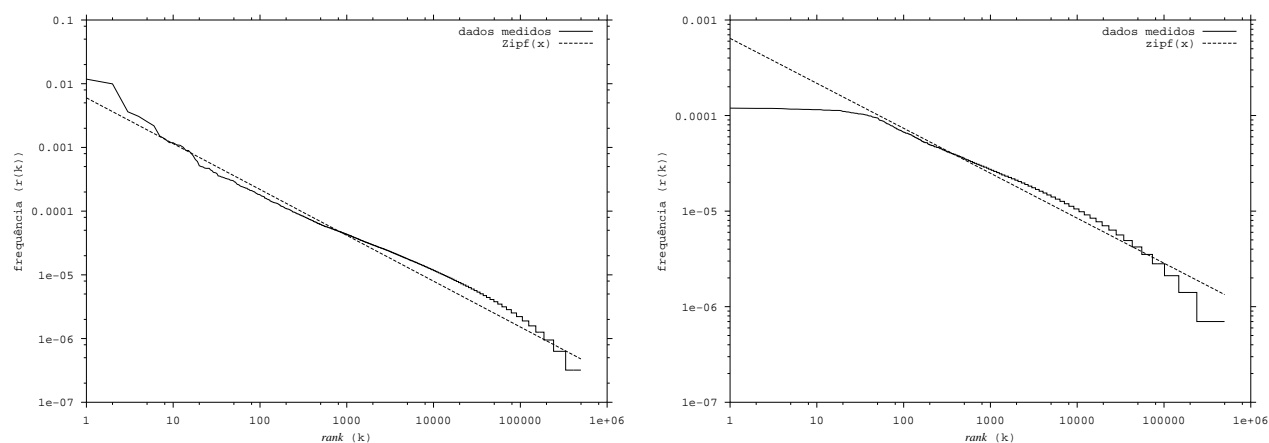
Figura 4.24: Distribuição marginal do período *OFF* inativo

A Figura 4.24(a) mostra que a distribuição de períodos *OFF* inativos parece formar ondulações em torno de valores específicos como 1 hora, 1 dia, 2 dias, 3 dias, etc. Esta propriedade evidencia a variabilidade do interesse dos clientes no *stream* ao vivo sendo disponibilizado, isto é, observamos clientes que revisitam o programa a cada hora, diariamente, a cada dois dias, etc.

Apesar da distribuição Exponencial se aproximar da distribuição da duração de períodos *OFF* inativos encontrada, ela não é capaz de modelar estas ondulações características do comportamento dos clientes.

4.5.6 Perfil de Interesse dos Clientes

Ao longo de todo o período de *log* analisado, cada cliente (re)visita o conteúdo ao vivo um número qualquer de vezes. Seja k o *rank* do cliente em termos do número de transferências (ou sessões) associadas a esse cliente no histórico de acessos. A Figura 4.25(a) mostra a relação entre o número de transferências para um cliente, em resposta a requisições enviadas por esse cliente ao servidor de *streaming media* ao vivo, (no eixo y) e o *rank* deste cliente (no eixo x). Por sua vez, a Figura 4.25(b) apresenta a relação entre o número de sessões de um cliente (no eixo y) e o *rank* deste cliente (no eixo x). Estas relações podem ser modeladas por leis de *Zipf* com parâmetros $\alpha = 0.9197$ e $\alpha = 0.4704$, respectivamente.

(a) *Rank* do cliente X # de transferências(b) *Rank* do cliente X # de sessões**Figura 4.25:** Perfil de interesse do cliente

Uma forma de interpretar esta relação é utilizar o número de transferências ou sessões associadas a um cliente, como uma medida do interesse do cliente no conteúdo ao vivo. Note-se que, para conteúdo armazenado (objetos de *streaming media* pré-armazenados ou páginas *Web* tradicionais), é comum determinar a popularidade de um objeto em termos do número de vezes que este objeto foi requisitado pelos vários clientes. No contexto de *streams* ao vivo, a caracterização da popularidade do objeto, dessa forma, não tem muito significado, uma vez que os clientes não podem acessar, repetidas vezes, o mesmo conteúdo. Assim, é mais interessante estimar o interesse de um dado cliente no conteúdo ao vivo em termos do número de vezes que esse cliente acessou as diversas partes do objeto ao vivo.

Capítulo 5

Caracterização de Outro Objeto Ao Vivo

Neste capítulo, apresentamos a caracterização de outro objeto de *streaming media* ao vivo. No entanto, a presente caracterização não será realizada utilizando-se o mesmo nível de detalhes usado para caracterizar os acessos aos *streams* ao vivo associados a um *reality show* exibido como programa de televisão no Brasil (Capítulo 4). Neste capítulo, as características presentes em uma camada da hierarquia de caracterização que podiam ser inferidas das propriedades de outras camadas não foram apresentadas.

5.1 Natureza da Carga de Trabalho

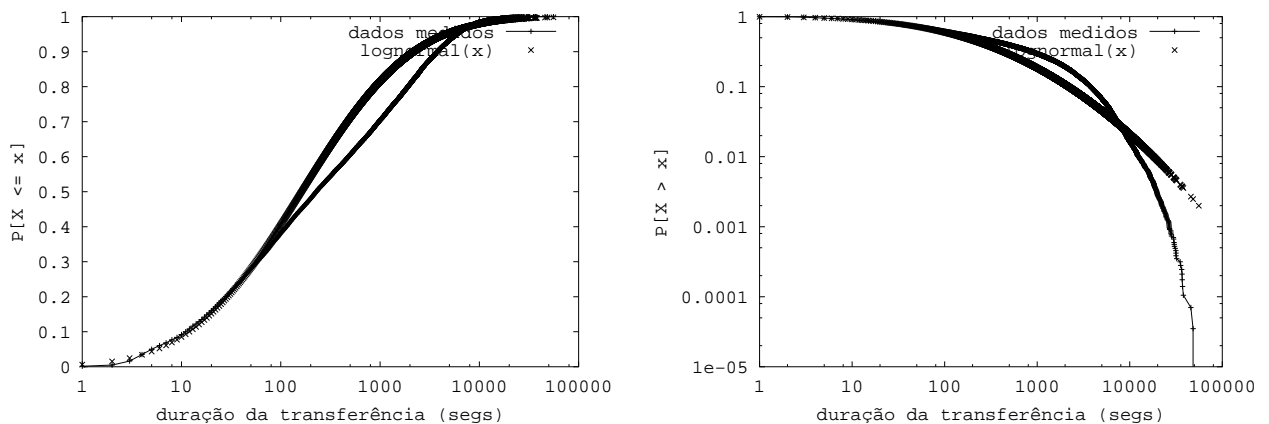
Analisamos neste capítulo, os acessos a um servidor de *streaming media* ao vivo que disponibiliza, através da *Internet*, o conteúdo transmitido também por uma estação de rádio de audiência nacional. Assim, o conteúdo ao vivo distribuído através da *Internet* apresenta uma programação rigorosamente definida. Basicamente, o objeto ao vivo está associado a programas jornalísticos e esportivos. Devemos ressaltar ainda que, o *stream* ao vivo é transmitido 24 horas por dia, durante todos os dias da semana, sendo que a programação dos finais de semana apresenta algumas alterações em relação aos programas exibidos durante os dias úteis de semana. O histórico de acessos analisado foi coletado no início do ano de 2002 e tem duração de 19 dias.

5.2 Camada de Transferência

5.2.1 Distribuição da Duração das Transferências

Nesta seção analisamos a duração de transferências individuais do objeto ao vivo transmitido através da *Internet* por uma estação de rádio.

Como definido na Seção 4.3.3, $l(j)$ denota a duração, em segundos, da j -ésima transferência no histórico de acessos. As Figuras 5.1(a) e (b) apresentam, respectivamente, a distribuição acumulativa e o complemento da distribuição acumulativa de $l(j)$. Aproximamos os valores computados para $l(j)$ de acordo com o *log* de acesso, de uma distribuição Lognormal ($p(x) = (\sqrt{2\pi}Sx)^{-1}e^{-(\ln(x)-M)^2/(2S^2)}$), com parâmetros $M = 5.085 \pm 0.0115$ e $S = 2.029 \pm 0.0088$. O coeficiente de correlação encontrado para essa aproximação foi $R^2 = 0.88$.



(a) Distribuição acumulativa

(b) Complemento da distribuição acumulativa

Figura 5.1: Distribuição marginal da duração das transferências individuais associadas ao conteúdo ao vivo

5.3 Camada de Sessão

5.3.1 Número de Sessões

A Figura 5.2 apresenta a relação entre o número de sessões distintas identificadas no *log* de acessos analisado e o valor escolhido para o patamar T_{off} , que representa o *timeout* de sessão. Esta relação demonstra que o número de sessões identificadas não se altera drasticamente para $T_{off} > 3600$ segundos. Assim, no restante deste capítulo, utilizamos o patamar $T_{off} = 3600$ segundos para identificar diferentes sessões.

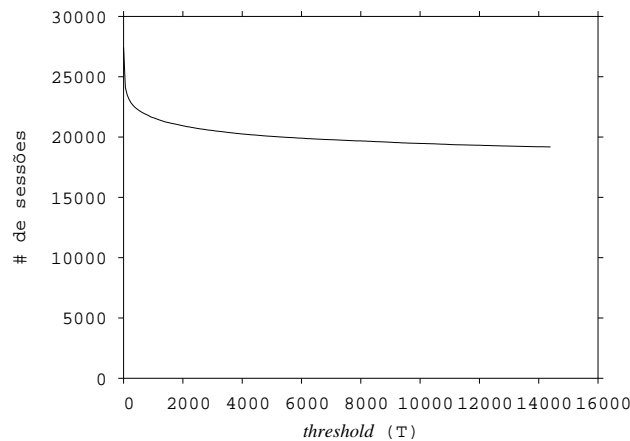
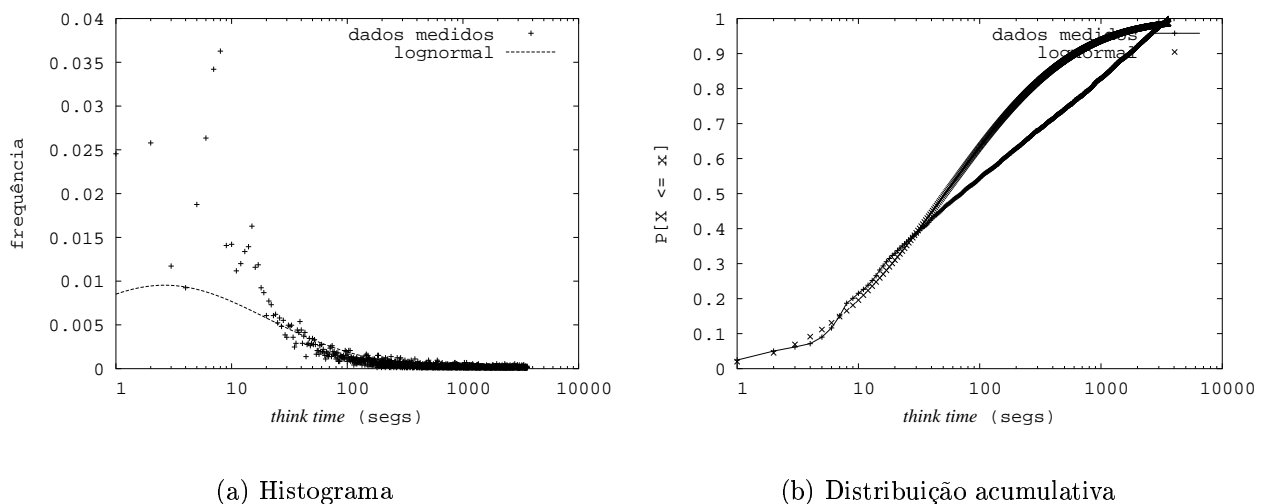


Figura 5.2: Relação entre o número de sessões distintas identificadas e T_{off}

5.3.2 Distribuição do Período *OFF* Ativo

Nesta seção caracterizamos a duração dos períodos de inatividade de um cliente, dentro de uma sessão. Por definição, a duração dos períodos *OFF* ativos deve ser menor que T_{off} .

Como foi especificado na Seção 4.4.3, $offAtivo(i) = t(j) - (t(i) + l(i))$ representa a duração do período *OFF* ativo do cliente entre as transferências i e j , que pertencem a uma mesma sessão. Além disso, devemos recordar que i e j são transferências consecutivas dentro de uma sessão. A Figura 5.3 mostra a distribuição marginal da duração do *think time* dos clientes.



(a) Histograma

(b) Distribuição acumulativa

Figura 5.3: Distribuição marginal da duração do *think time* dos clientes

Aproximamos os valores computados para $offAtivo(i)$ de acordo com o *log* de acessos, de uma distribuição Lognormal ($p(x) = (\sqrt{2\pi}Sx)^{-1}e^{-(\ln(x)-M)^2/(2S^2)}$), com parâmetros $M = 3.949 \pm 0.0246$ e $S = 1.924 \pm 0.0217$. O coeficiente de correlação encontrado para essa aproximação foi $R^2 = 0.81$.

5.3.3 Número de Transferências por Sessão

O número de transferências dentro de uma mesma sessão de um cliente evidencia a interação desse cliente com o objeto de *streaming media* ao vivo. A Figura 5.4 apresenta a distribuição acumulativa do número total de transferências dentro de cada sessão identificada no *log* de acessos. A distribuição resultante caracteriza um comportamento de natureza cauda-pesada, que pode ser modelado por uma distribuição de Pareto (ab^a/x^{a+1}), com parâmetros $a = 1.689 \pm 0.0218$ e $b = 0.391 \pm 0.0056$. O coeficiente de correlação encontrado para essa aproximação foi $R^2 = 0.998$.

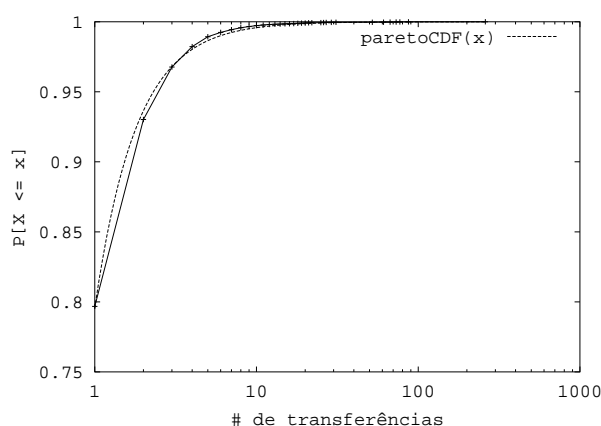


Figura 5.4: Distribuição acumulativa do número de transferências dentro de uma mesma sessão

Estudamos também a correlação entre o número de transferências associadas a cada sessão e o horário do dia em que esta sessão iniciou-se, mas concluímos que a variabilidade no número de transferências por sessão não é fortemente associada a características temporais. Então, atribuímos essa variabilidade à natureza das interações dos clientes com o conteúdo ao vivo.

5.4 Camada de Cliente

5.4.1 Número de Clientes Concorrentes

A Figura 5.5 mostra a distribuição marginal do número de clientes concorrentes, associados ao conteúdo ao vivo analisado neste capítulo, ao longo de toda a duração estudada do *log* de acessos. Essa característica pode ser modelada por uma distribuição Exponencial (ae^{-ax}) com parâmetro $a = 0.046 \pm 0.0004$. O coeficiente de correlação encontrado para essa aproximação foi $R^2 = 0.99$.

Na Figura 5.6 apresentamos o comportamento temporal do número de clientes ativos associados ao objeto ao vivo, em cada instante de tempo t , $c(t)$. A Figura 5.6(a) apresenta o

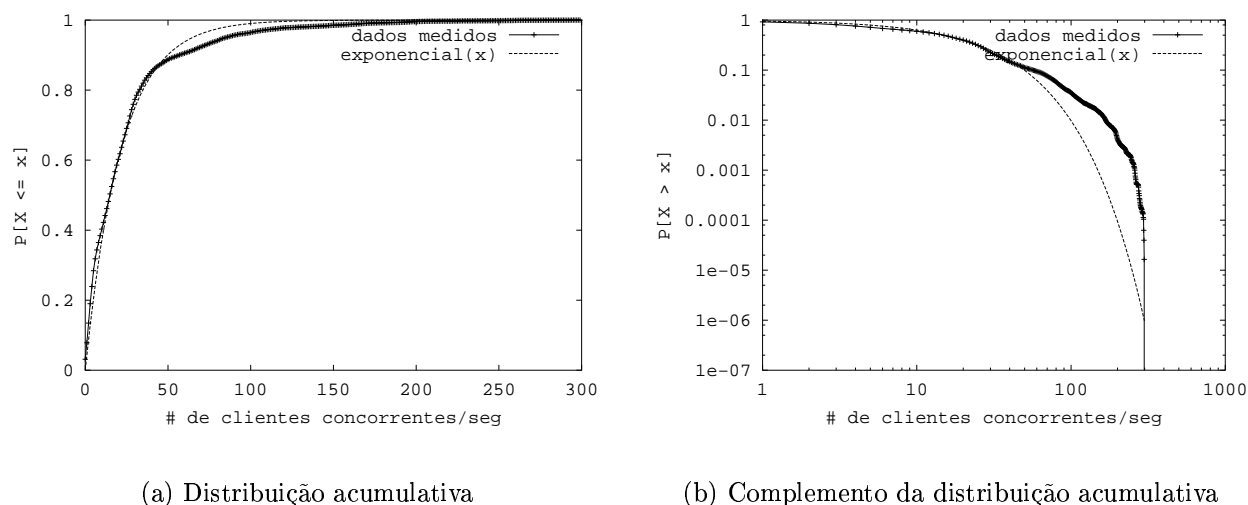


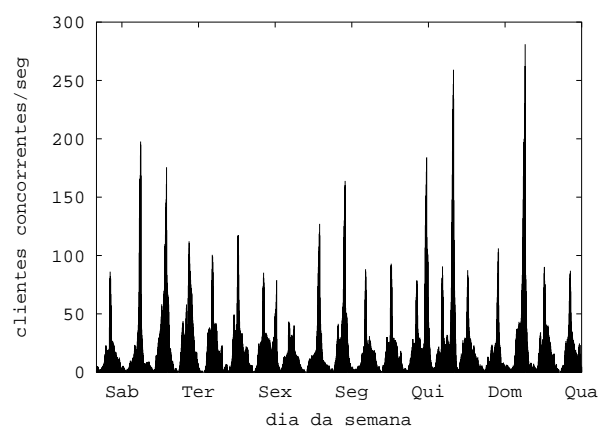
Figura 5.5: Distribuição marginal do número de clientes concorrentes associados ao conteúdo ao vivo

número médio de clientes concorrentes por segundo, computado em *bins* de 10 minutos, ao longo de todo o período analisado do histórico de acessos. Nessa figura, podemos observar grande variabilidade no número de clientes ativos. As Figuras 5.6(b), (c) e (d) evidenciam o comportamento periódico de $c(t)$, apresentando $c(t \bmod p)$, em que p é um dia e t é um instante de tempo localizado em dia útil de semana, em um sábado e em um domingo, respectivamente.

Devemos notar que diversos fatores podem contribuir para a grande variabilidade observada no número de clientes concorrentes que acessam o objeto ao vivo, incluindo o conteúdo do programa ao vivo que está sendo transmitido, assim como aspectos relacionados à população de clientes.

Na Figura 5.6(c), podemos observar um aumento considerável no número de clientes concorrentes no período das 16 às 18 horas. Nesse horário, de acordo com a programação definida para o *stream* ao vivo, são transmitidas partidas de futebol dos principais campeonatos realizados no Brasil. Por sua vez, a Figura 5.6(d) apresenta um grande aumento no número de clientes ativos no período das 16 às 19 horas. Ainda de acordo com a programação divulgada para o conteúdo do objeto de *streaming media* ao vivo, nesse horário são transmitidos jogos de futebol ou entrevistas e reportagens sobre fatos relacionados a este esporte. Assim, acreditamos que a variabilidade observada no número de clientes ativos dessas figuras ocorra devido ao conteúdo ao vivo disponibilizado e o interesse que este desperta nos espectadores.

Também na Figura 5.6(b) podemos observar variações, ao longo do tempo, no número de clientes ativos. Como, por exemplo, os dois maiores picos de acesso observados nesta figura ocorreram no horário de transmissão de programas esportivos, acreditamos que a



(a) Ao longo de todo o período analisado

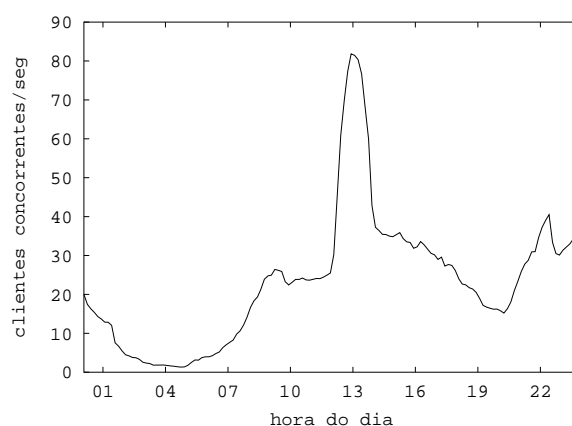
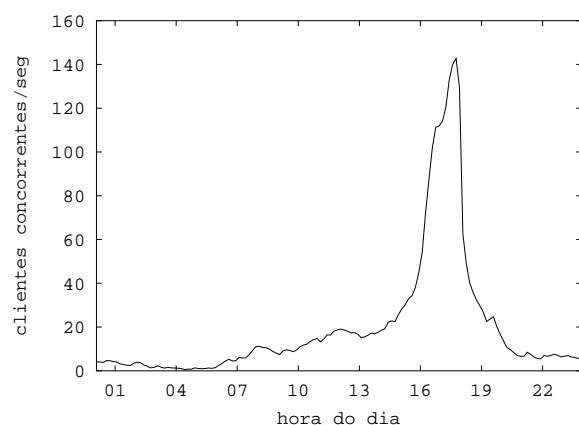
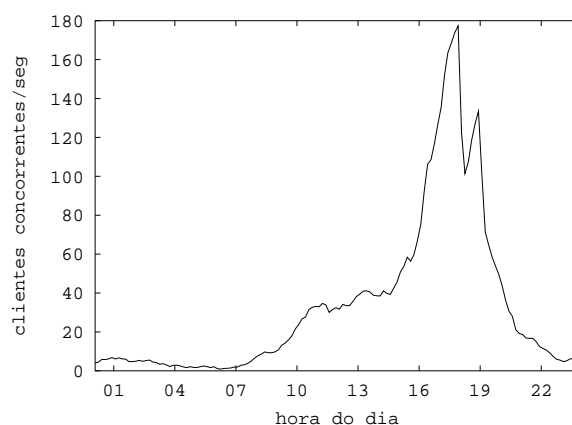
(b) $c(t \bmod p)$, para dias úteis(c) $c(t \bmod p)$, para os sábados(d) $c(t \bmod p)$, para os domingos

Figura 5.6: Comportamento temporal do número de clientes concorrentes, $c(t)$, que acessam o objeto ao vivo

variabilidade no número de clientes concorrentes, ao longo do tempo, seja um indicativo de que alguns programas despertam maior interesse nos clientes. Assim, mais uma vez, acreditamos que o conteúdo ao vivo que está sendo transmitido influencia fortemente o nível de concorrência dos clientes.

5.4.2 Tempo entre as Chegadas de Clientes Consecutivos

Como instanciado na Seção 4.5.3, $arrivalClient(j)$ denota o tempo decorrido entre as chegadas dos clientes j e $j + 1$.

A Figura 5.7 apresenta a distribuição marginal de $arrivalClient(j)$, para toda a duração analisada do *log* de acessos. Aproximamos os *iats* computados de acordo com o histórico

de acessos analisado, de uma distribuição Lognormal ($p(x) = (\sqrt{2\pi}Sx)^{-1}e^{-(\ln(x)-M)^2/(2S^2)}$) com parâmetros $M = 3.578 \pm 0.0116$ e $S = 1.518 \pm 0.0010$. O coeficiente de correlação encontrado para essa aproximação foi $R^2 = 0.97$.

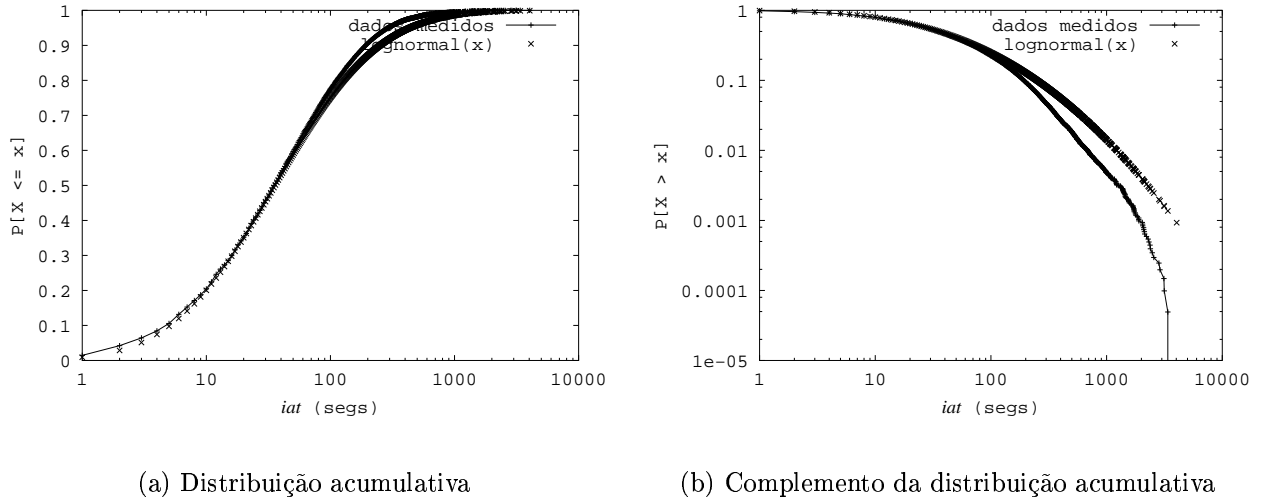


Figura 5.7: Distribuição marginal do *iat* de clientes consecutivos

Na Figura 5.8, mostramos o comportamento temporal do *iat* de clientes consecutivos que acessam o objeto ao vivo. A Figura 5.8(a) apresenta o *iat* médio de clientes consecutivos, percebido no instante t ($iatClient(t)$), ao longo de toda a duração analisada do *log* de acessos. Por sua vez, as Figuras 5.8(b), (c) e (d) apresentam o $iatClient(t \bmod p)$, em que p é um dia e t corresponde ao instante de tempo localizado, respectivamente, em um dia útil de semana, em um sábado e em um domingo. Esses gráficos foram obtidos calculando-se o tempo médio entre chegadas de clientes durante períodos consecutivos de 10 minutos.

O *iat* de clientes consecutivos apresenta algumas variações ao longo do tempo. Como mostrado nas Figuras 5.8(b), (c) e (d), o comportamento diário é a principal fonte de variabilidade, com o período de 0 às 10 horas apresentando *iats* muito mais longos. Acreditamos que essa característica ocorra devido a fatores relacionados à população de clientes e ao conteúdo ao vivo disponibilizado.

Além disso, como evidenciam as Figuras 5.8(c) e (d), o *iat* médio de clientes consecutivos percebido, no sábado, no período das 16 às 18 horas, e no domingo, no período das 16 às 19 horas, respectivamente, apresenta um valor um pouco menor. Como esses períodos coincidem com o horário de transmissão de jogos de futebol ou programas esportivos muito populares, acreditamos que essa característica ocorra devido ao conteúdo ao vivo transmitido.

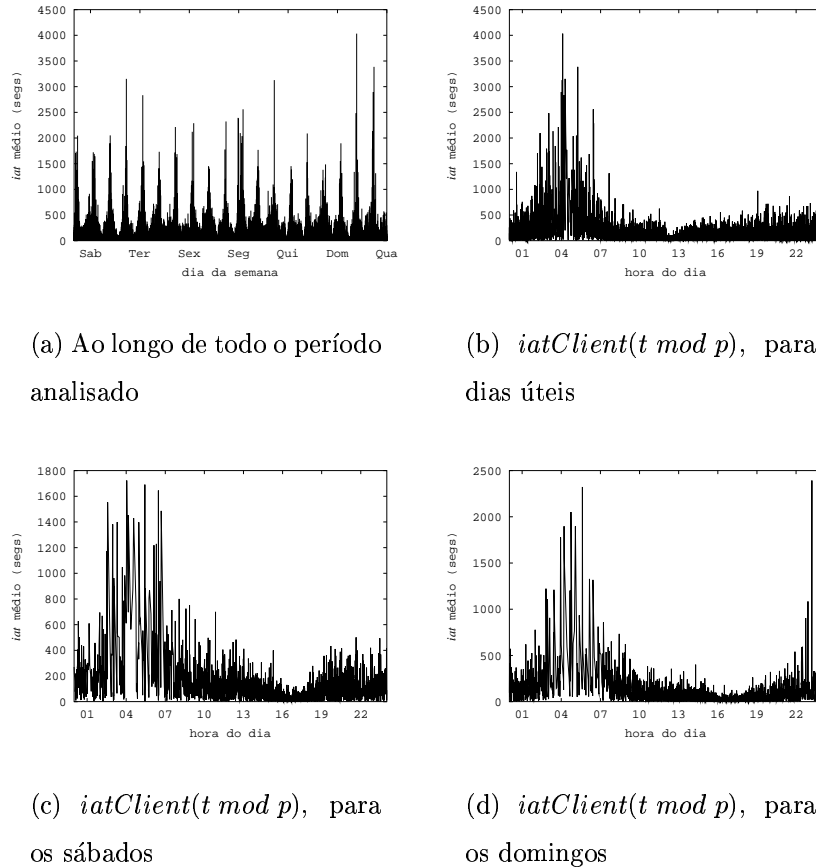


Figura 5.8: Comportamento temporal do iat de clientes consecutivos

5.4.3 Distribuição do Período *OFF* Inativo

Nesta seção caracterizamos a duração dos períodos de inatividade de um cliente, entre duas de suas sessões consecutivas. Por definição, a duração dos períodos *OFF* inativos deve ser maior que T_{off} .

Como foi instanciado na Seção 4.5.5, o período de inatividade do cliente entre as suas sessões consecutivas i e j é dado por $offInativo(i) = t(j) - (t(i) + on(i))$. A Figura 5.9 mostra o complemento da distribuição acumulativa dos períodos *OFF* inativos, para todas as sessões identificadas no *log* de acessos. Aproximamos os valores computados para o *OFF* inativo dos clientes de uma distribuição Exponencial (ae^{-ax}), com parâmetro $a = 6.008e^{-6} \pm 1.775e^{-8}$. O coeficiente de correlação encontrado para essa aproximação foi $R^2 = 0.96$.

5.4.4 Perfil de Interesse dos Clientes

Cada cliente pode (re)visitar o conteúdo ao vivo um número qualquer de vezes, ao longo do tempo. Assim, seja k o *rank* do cliente em termos do número de transferências

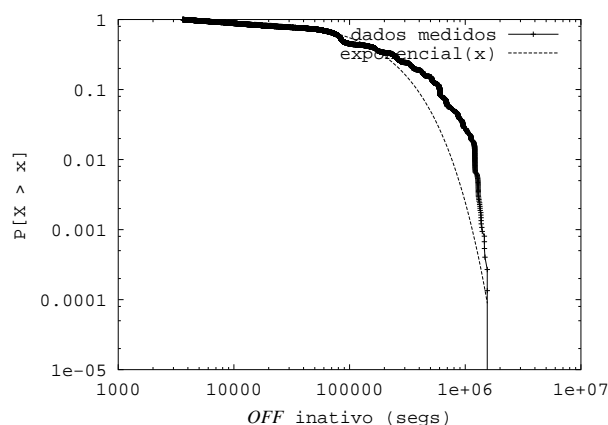
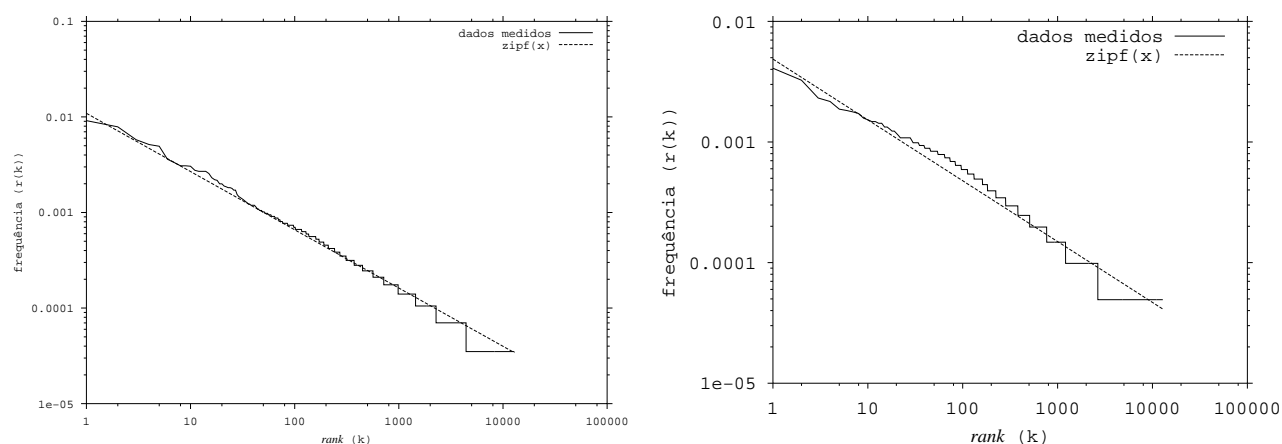


Figura 5.9: Complemento da distribuição acumulativa do período *OFF* inativo

(ou sessões) associadas a esse cliente no histórico de acessos. A Figura 5.10 (a) mostra a relação entre o número de transferências para um cliente e o *rank* deste cliente. Por sua vez, a Figura 5.10(b) apresenta a relação entre o número de sessões de um cliente e o *rank* deste cliente. Estas relações podem ser modeladas por leis de *Zipf* com parâmetros $\alpha = 0.6094$ e $\alpha = 0.5044$, respectivamente.



(a) *Rank* do cliente X # de transferências

(b) *Rank* do cliente X # de sessões

Figura 5.10: Perfil de interesse do cliente

O número de transferências ou sessões associadas a um cliente, pode ser utilizada para quantificar o interesse do cliente no conteúdo ao vivo.

Capítulo 6

Geração de Cargas de Trabalho de

Live Streaming Media Sintéticas

Como discutimos anteriormente, a caracterização de cargas de trabalho de *streaming media* ao vivo é fundamental para a geração de cargas de trabalho sintéticas e parametrizáveis. Neste capítulo, descrevemos como os resultados da caracterização hierárquica podem ser utilizados para gerar cargas de trabalho de mídia de fluxo contínuo de dados ao vivo.

Durante a definição da hierarquia de caracterização proposta (Capítulo 3), consideramos diversas variáveis, nas diferentes camadas. Muitas dessas variáveis não são independentes. A distribuição do número de clientes ativos, por exemplo, pode ser obtida a partir das distribuições do *iat* de clientes consecutivos e da duração do período *ON* de sessão. Permitir algumas redundâncias na caracterização de uma carga de trabalho é bastante interessante, pois permite entender vários aspectos dos padrões de acesso e a maneira como as mudanças em uma propriedade da carga de trabalho se refletem em outras características dessa carga. No entanto, quando necessitamos utilizar os resultados da caracterização para gerar cargas de trabalho sintéticas, precisamos escolher quais variáveis serão utilizadas na geração do *log* sintético.

Neste capítulo, apresentamos um modelo para a geração de cargas de trabalho sintéticas baseado em um subconjunto das variáveis analisadas nas diversas camadas da hierarquia de caracterização proposta no Capítulo 3. É importante observarmos que o modelo apresentado a seguir não é único. Outros modelos poderiam ser elaborados considerando-se outras variáveis.

A seguir, apresentamos as diferentes propriedades das camadas da hierarquia de caracterização que colaboram para a elaboração do modelo para a geração de cargas de trabalho de *streaming media* ao vivo sintéticas.

Chegada de Clientes: Para gerarmos o comportamento dos clientes que acessam o objeto de *streaming media* ao vivo, é necessário especificar suas sessões, e finalmente as transferências que ocorrem dentro de cada sessão. Para isso, precisamos determinar o momento em que as sessões dos clientes se iniciam e quais clientes estão associados a cada sessão. Para determinar o momento quando uma sessão começa, utilizamos a distribuição do *iat* de clientes consecutivos e observamos a natureza periódica dessa variável. Para determinar o cliente associado a cada sessão, usamos o perfil de interesse do cliente no *stream* ao vivo, medido em termos do número de sessões a que este cliente está associado, e a distribuição da duração do período *OFF* inativo.

Tamanho da Sessão: Para especificarmos as propriedades observadas dentro de cada sessão, é necessário determinar quantas transferências ocorreram dentro de cada sessão. Para a geração desta característica, utilizamos a distribuição do número de transferências por sessão.

Transferências: Para gerarmos as transferências associadas a cada sessão, é preciso determinar quando cada transferência se inicia e por quanto tempo esta transferência permanece ativa. Por definição, a primeira transferência de uma sessão começa no instante em que esta sessão tem início. O instante em que as outras transferências dentro da sessão devem se iniciar (se existir alguma) pode ser determinado utilizando-se a distribuição do *iat* de transferências pertencentes a uma mesma sessão. Alternativamente, esse instante pode ser também determinado através do uso da distribuição de períodos *OFF* ativos. A duração de cada transferência dentro de uma sessão é determinada utilizando-se a distribuição da duração de transferências individuais.

A Tabela 6.1 indica o subconjunto de variáveis utilizado em nosso modelo de geração de cargas de trabalho de *streaming media* ao vivo sintéticas. Esta tabela também apresenta as propriedades de distribuição dessas variáveis, de acordo com o resultado encontrado durante a caracterização hierárquica da carga de trabalho gerada pelos acessos aos objetos ao vivo associados à transmissão de um *reality show* exibido como programa de televisão no Brasil.

Por sua vez, a Tabela 6.2 mostra as propriedades de distribuição dessas variáveis, de acordo com os resultados encontrados através da caracterização hierárquica da carga de trabalho associada à transmissão, através da *Internet*, de um programa de rádio de audiência nacional.

É importante observarmos que muitas das características das cargas de trabalho de *streaming media* ao vivo apresentam forte correlação com o conteúdo ao vivo sendo transmitido. Por exemplo, a periodicidade observada na carga de trabalho gerada pelos acessos aos objetos ao vivo associados ao *reality show* exibido como programa de televisão pro-

Variável	Distribuição	Parâmetros
iat de clientes consecutivos	Pareto	$x < 200 : a = 2.52 e b = 1.56$ $x \geq 200 : a = 0.76 e b = 2.10e^{-5}$
Perfil de interesse do cliente (em sessões)	<i>Zipf</i>	$\alpha = 0.9197$
<i>OFF</i> inativo	Exponencial	$a = 5.025e^{-6}$
# de transferências por sessão	Pareto	$a = 1.43 e b = 0.62$
iat de transferências de uma mesma sessão	Lognormal	$M = 4.93eS = 1.26$
duração de transferências individuais	Lognormal	$M = 4.29eS = 1.28$

Tabela 6.1: Resumo das propriedades de distribuição das variáveis utilizadas para a geração de cargas de trabalho de *streaming media* ao vivo - características encontradas para os *streams* associado ao *reality show*

Variável	Distribuição	Parâmetros
iat de clientes consecutivos	Lognormal	$M = 3.58eS = 1.52$
Perfil de interesse do cliente (em sessões)	<i>Zipf</i>	$\alpha = 0.504$
<i>OFF</i> inativo	Exponencial	$a = 6.008e^{-6}$
# de transferências por sessão	Pareto	$a = 1.69 e b = 0.39$
<i>OFF</i> ativo	Lognormal	$M = 3.95eS = 1.92$
duração de transferências individuais	Lognormal	$M = 5.09eS = 2.03$

Tabela 6.2: Resumo das propriedades de distribuição das variáveis utilizadas para a geração de cargas de trabalho de *streaming media* ao vivo - características encontradas para o *stream* associado ao programa de rádio

vavelmente é bastante diferente daquela demonstrada em cargas de trabalho associadas à exibição de partidas de futebol. Assim, as características encontradas para as cargas de trabalho de *streaming media* ao vivo podem variar, de acordo com a categoria do conteúdo ao vivo sendo transmitido. No entanto, a hierarquia para a caracterização de *streaming media* ao vivo proposta e o modelo para geração de cargas de trabalho sintéticas podem ser utilizados na determinação das propriedades dos diversos tipos de cargas de trabalho associadas a objetos ao vivo.

Capítulo 7

Conclusões e Trabalhos Futuros

Este trabalho apresenta uma hierarquia para a caracterização de cargas de trabalho de *streaming media* ao vivo distribuídas através da *Internet*. Essa hierarquia baseia-se em três níveis de granularidades diferentes, correspondendo a transferências, sessões e clientes, respectivamente. Aplicamos essa hierarquia na caracterização de dois tipos de objetos de *streaming media* ao vivo distintos. O primeiro deles corresponde a *streams* ao vivo transmitidos, através da *Internet*, em complemento a um *reality show* exibido como programa de televisão. O segundo, está associado a um objeto que retransmite, através da *Internet*, um programa de rádio de audiência nacional. Nossa caracterização identificou algumas propriedades interessantes de objetos ao vivo, nas três camadas da hierarquia:

Camada de Transferência:

- A duração de transferências individuais segue, aproximadamente, uma distribuição Lognormal. Atribuímos a grande variabilidade observada na duração dessas transferências à disposição do cliente em assistir ao *stream* ao vivo, e não apenas às propriedades estruturais desses objetos, como tamanho e duração totais.
- A largura de banda média utilizada pelas transferências é primariamente determinada pela velocidade de conexão do cliente com a *Internet*. Aproximadamente 10% das transferências presentes na carga de trabalho gerada pelos acessos aos objetos relacionados ao *reality show*, foram limitadas por recursos insuficientes de rede.

Camada de Sessão:

- A duração do período *OFF* ativo segue aproximadamente uma distribuição Lognormal.
- O número de transferências dentro de uma mesma sessão parece poder ser modelado por uma distribuição de *Pareto*.

Camada de Cliente:

- A carga dos servidores de *streaming media* analisados, medida em termos do nível de concorrência no sistema, exibe fortes correlações temporais e a distribuição do número de clientes ativos pode ser modelada por uma distribuição Exponencial.
- O *iat* de clientes consecutivos que acessam o *stream* ao vivo transmitido em complemento a um *reality show* exibido na televisão segue, aproximadamente, distribuições de Pareto. Já o tempo entre chegadas de clientes consecutivos que requisitam o *stream* associado ao programa de rádio, parece seguir uma distribuição Lognormal.
- O perfil de interesse dos clientes em objetos ao vivo pode ser modelado por uma lei de *Zipf*.
- A duração do período *OFF* inativo segue aproximadamente uma distribuição Exponencial.

Além disso, podemos observar que a natureza das interações entre usuários e objetos é fundamentalmente diferente para objetos ao vivo e pré-armazenados. Acessos a objetos pré-armazenados são determinados pelo usuário, ao passo que, os acessos a objetos ao vivo são predominantemente dirigidos pelos próprios objetos. Essa diferença entre os papéis (ativo/passivo) de usuários e objetos conduz a contrastes bastante interessantes. Acreditamos que as características particulares das cargas de trabalho de *streaming media* ao vivo são, provavelmente, altamente dependentes da natureza do conteúdo ao vivo sendo exibido. Além disso, acreditamos que as fortes correlações temporais, observadas nos registros de acesso analisados no presente estudo, resultam da sincronização entre o conteúdo ao vivo e as características de seu acesso.

Baseados em nossas análises, apresentamos um modelo para geração de cargas de trabalho de *streaming media* ao vivo que incorpora diversos resultados de nossa caracterização.

Pretendemos, em trabalhos futuros, caracterizar outros tipos de *streams* ao vivo e determinar suas propriedades de distribuição.

No presente estudo, não avaliamos o impacto que o congestionamento na rede, medido em função do número de pacotes perdidos ou na quantidade de conexões canceladas, teria nas características dos padrões de acesso dos clientes. A investigação dessa questão também constitui-se objeto de trabalhos futuros que pretendemos abordar.

Bibliografia

- [1] S. Acharya and B. Smith. An experiment to characterize videos stored on the Web. In *Proceedings of the ACM/SPIE Conference on Multimedia Computing and Networking (MMCN)*, 1998.
- [2] S. Acharya, B. Smith, and P. Parns. Characterizing user access to video on the World Wide Web. In *Proceedings of the ACM/SPIE Conference on Multimedia Computing and Networking (MMCN)*, 2000.
- [3] J. Almeida, J. Krueger, D. Eager, and M. Vernon. Analysis of educational media server workloads. In *Proceedings of the 11th International Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV)*, 2001.
- [4] V. Almeida, A. Bestavros, M. Crovella, and A. de Oliveira. Characterizing reference locality in the WWW. In *Proceedings of the IEEE International Conference on Parallel and Distributed Information Systems (PDIS)*, 1996.
- [5] M. Arlitt and C. Williamson. Web server workload characterization: The search for invariants. In *Proceedings of the ACM SIGMETRICS*, 1996.
- [6] G. Banga and P. Druschel. Measuring the capacity of a Web server. In *Proceedings of the 1st USENIX Symposium on Internet Technologies and Systems (USITS)*, 1997.
- [7] P. Barford, A. Bestavros, A. Bradley, and M. Crovella. Changes in Web client access patterns: Characteristics and caching implications. *World Wide Web Journal*, 1998.
- [8] P. Barford and M. Crovella. Generating representative Web workloads for network and server performance evaluation. In *Proceedings of the ACM SIGMETRICS*, 1998.
- [9] L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker. Web caching and Zipf-like distributions: Evidence and implications. In *Proceedings of the IEEE INFOCOM Conference*, 1999.
- [10] K. chan Lan and J. Heidemann. Multi-scale validation of structural models of audio traffic. Technical Report ISI-TR-544, University of Southern California, 2001.

- [11] M. Chesire, A. Wolman, G. Voelker, and H. Levy. Measurement and analysis of a streaming-media workload. In *Proceedings of the 3rd USENIX Symposium on Internet Technologies and Systems (USITS)*, 2001.
- [12] M. Corporation. Windows media services 4.1. <http://www.microsoft.com/windows/windowsmedia/technologies/services.asp>.
- [13] M. Crovella and A. Bestavros. Self-similarity in World Wide Web traffic: evidence and possible causes. In *Proceedings of the ACM SIGMETRICS*, 1996.
- [14] C. Cunha, A. Bestavros, and M. Crovella. Characteristics of WWW client-based traces. Technical Report BU-CS-95-010, Boston University, 1995.
- [15] A. Downey. The structural cause of file size distributions. In *Proceedings of the International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems (MASCOTS'01)*, IEEE Computer Society, 2001.
- [16] S. Gribble and E. Brewer. System design issues for Internet middleware services: Deductions from a large client trace. In *Proceedings of the 1st USENIX Symposium on Internet Technologies and Systems (USITS)*, 1997.
- [17] D. Loguinov and H. Radha. Measurement study of low-bitrate internet video streaming. In *Proceedings of the ACM SIGCOMM Internet Measurement Workshop (IMW)*, 2001.
- [18] A. Mena and J. Heidemann. An empirical study of real audio traffic. In *Proceedings of the IEEE INFOCOM Conference*, 2000.
- [19] D. Menascé, V. Almeida, R. Riedi, F. Pelegrinelli, R. Fonseca, and W. M. Jr. In search of invariants for e-business workloads. In *Proceedings of the 2000 ACM Conference in E-commerce*, 2000.
- [20] M. Mitzenmacher. Dynamic models for file sizes and double pareto distributions. Preprint, 2002.
- [21] J. Padhye and J. Kurose. An empirical study of client interactions with a continuous-media courseware server. In *Proceedings of the 8th International Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV)*, 1998.
- [22] V. Padmanabhan and L. Qiu. The content and access dynamics of a busy Web site: Findings and implications. In *Proceedings of the ACM SIGCOMM*, 2000.
- [23] V. Paxson and S. Floyd. Wide-area traffic: the failure of poisson modeling. In *Proceedings of the ACM SIGCOMM*, 1994.

- [24] Y. Wang, M. Claypool, and Z. Zuo. An empirical study of realvideo performance across the internet. In *Proceedings of the ACM SIGCOMM Internet Measurement Workshop (IMW)*, 2001.
- [25] Zipf. *Human behavior and the principal of least-effort*. Addison-Wesley, 1949.