

Universidade Federal de Pernambuco
Centro de Informática
Pós-Graduação em Ciência da Computação

Estatística Aplicada
Prof^a Renata Souza



Modelagem do tráfego de um ambiente
virtual - Estudo de caso: Second Life

Eduardo Feitosa
Jeísa Oliveira
Rafael Antonello

Recife, abril de 2007

Sumário

1	Introdução	1
1.1	Motivação.....	1
1.2	Objetivo	1
2	Modelando Tráfego	2
2.1	Background.....	2
2.2	Tráfego de jogos online	3
2.3	Metodologia.....	3
2.3.1	Escolha dos dados	3
2.3.2	Escolha da distribuição analítica.....	4
2.3.3	Estimação dos parâmetros	4
2.3.4	Gráficos Q-Q Plot.....	4
2.3.5	Estimação da discrepância do modelo	4
2.3.6	Teste de Kolmogorov-Smirnov.....	5
3	Análise Exploratória	5
3.1	Trace Andando-1 Cliente	5
3.1.1	Métrica <i>Inter Arrival</i>	6
3.1.2	Métrica <i>Packet Size</i>	6
3.2	Trace Andando-1 Servidor.....	7
3.2.1	Métrica <i>Packet Size</i>	7
3.3	Traces Parado (Cliente e Servidor).....	8
4	Modelagem dos traces	8
4.1	Trace Parado 1 - Cliente	9
4.1.1	Métrica <i>Inter Arrival</i>	9
4.1.2	Métrica <i>Packet Size</i>	20
4.2	Trace Andando-1 Cliente	21
4.2.1	Métrica <i>Inter Arrival</i>	21
4.2.2	Métrica <i>Packet Size</i>	26
4.3	Trace Andando 1 - Servidor.....	27
4.3.1	Métrica <i>Inter Arrival</i>	27
4.3.2	Métrica <i>Packet Size</i>	27
4.4	Trace Parado-1 Servidor.....	32
4.4.1	Métrica <i>Inter Arrival</i>	32
4.4.2	Métrica <i>Packet Size</i>	32
5	Avaliação do Modelo	32
5.1	Traces Parado 2 3 4 5 - Cliente.....	32
5.1.1	Métrica <i>Inter Arrival</i>	32
5.2	Traces Andando 2 3 4 5 - Cliente.....	33
5.2.1	Métrica <i>Inter Arrival</i>	33
5.3	Traces Andando 2 3 4 5 - Servidor	34
5.3.1	Métrica <i>Packet Size</i>	34

5.4	Traces Parado 2 3 4 5 - Servidor	34
5.4.1	Métrica <i>Packet Size</i>	34
6	Conclusão	35
7	Referências	35

1 Introdução

O número de softwares que implementam e gerenciam mundos virtuais (VW do inglês *Virtual World*) têm atraído a atenção da mídia e das empresas nos últimos anos. A principal justificativa é que tais mundos virtuais apresentam diversas oportunidades aos usuários e visitantes como interação, níveis de socialização, além de espaço para divulgação, exposição e teste de produtos e conceitos.

Por definição, um mundo virtual é um ambiente de simulação ou simulado, gerado por computador e acessado por múltiplos usuários, representado por avatares, através de uma interface on-line, geralmente via Internet [1]. Mundos virtuais têm por objetivo fornecer a socialização entre usuários através de experiências e interações similares a do mundo real. Desta forma, eles acabam oferecendo novas formas de expressão para o comportamento humano, além de fornecer a “vantagem” da privacidade e o anonimato de seus usuários.

Devido ao interesse por mundos virtuais ainda ser recente, não existe uma classificação bem definida. Segundo Edward Castronova [2], mundos virtuais podem ser divididos em espaços de jogos e extensões do mundo real. Entretanto, a categorização mais aceita os classifica em jogos e mundos sociais. Os mundos virtuais sociais ou mundos sociais distinguem-se apresentando experiências mais realísticas. Os visitantes não devem necessariamente ganhar ou jogar um jogo, mas socializar-se com outros usuários. Enquanto os mundos virtuais de jogos focam um tema singular com convenções comuns tais como avatares, progressão via narrativa e uma série de eventos de competição, os mundos sociais tendem a ser muito menos estruturados, fornecendo um ajuste do tema de acordo com a realidade, ferramentas básicas para a construção do ambiente e a habilidade de hospedarem atividades e eventos. Em linhas gerais, os mundos sociais funcionam mais como comunidades e usam elementos de jogos.

Como exemplo de mundos virtuais pode-se destacar: The Sims Online [3] e Second Life [4]. Este último tem se tornado bastante popular devido sua capacidade de permitir que os próprios usuários (chamados de residentes) criem ambientes e objetos adicionais. De acordo com [5], o Second Life é um mundo virtual 3D criado e mantido por seus próprios residentes. Por permitir a interação entre múltiplos usuários, ele é considerado um MMORPG (*Massive Multiplayer Online Role Playing Game*). Entretanto, devido a capacidade de criar novos ambientes e objetos, e de cada residente ser dono de suas criações, ele também é classificado como MMOSG (*Massive Multiplayer Online Social Game*). Por isso, Second Life foi escolhido como foco de análise deste trabalho.

1.1 Motivação

Com o crescimento do número de usuários e do interesse de grandes empresas (IBM, Petrobras, MTV, entre outros) em fazer e divulgar negócios em mundos virtuais, os provedores de acesso à Internet (ISP, do inglês *Internet Service Provider*) têm percebido a necessidade de caracterizar os tráfegos gerados por tais aplicações.

O conhecimento e a compreensão do comportamento do tráfego de mundos virtuais tornou-se importante para atender a demanda por recursos (largura de banda, por exemplo), para manter a qualidade dos serviços e também como forma de gerar novos mecanismos de cobrança.

1.2 Objetivo

Este trabalho tem por objetivo caracterizar o comportamento do tráfego de mundos virtuais através do uso da modelagem estatística. Visando reduzir o escopo de trabalho, decidiu-se realizar um

estudo de caso sobre o software Second Life.

Especificamente, a modelagem do tráfego está dividida em duas métricas:

- *Inter Arrival*, que mede o tempo entre a chegada de pacotes em um determinado destino;
- *Packet Size*, que o mede o tamanho dos pacotes

Para melhorar o entendimento, o tráfego do Second Life será avaliado no sentido servidor-cliente, considerado tráfego de ENTRADA e no sentido cliente-servidor, considerado tráfego de SAÍDA.

O restante deste documento encontra-se organizado da seguinte forma. Na seção 2 encontra-se o *background* sobre a modelagem do tráfego e a descrição da metodologia proposta. A seção 3 apresenta a análise exploratória dos dados. A seção 4 ilustra o processo de modelagem, aplicando a metodologia proposta, e a seção 5 concentra o resultado da avaliação dos modelos. Finalmente, a seção 6 traz as conclusões sobre este estudo.

2 Modelando Tráfego

Como dito anteriormente, a modelagem do tráfego de redes é usado como base para o projeto de aplicações de rede e para planejamento de capacidade de sistemas de redes. Considerando o impacto que uma decisão errada teria nessa área, torna-se evidente a necessidade de modelos que ajudem nessa difícil tarefa de planejamento. Esta seção aborda os modelos comumente usados para tráfego de redes, em especial para tráfego de jogos na Internet. A metodologia utilizada para a modelagem realizada nesse trabalho é descrita em seguida.

2.1 Background

Há dois parâmetros principais gerados pelos modelos de tráfego de rede: distribuição do tamanho do pacote (*packet size*) e distribuição do tempo entre chegada (*inter arrival*) dos pacotes. Outros parâmetros como rotas e distribuição dos destinos são menos importantes. Simulações que usam traces gerados por modelos de tráfego de rede normalmente examinam um único nó na rede, como um roteador ou *switch*. Fatores que dependem especificamente de informações sobre a topologia da rede ou de roteamento são específicas para determinada topologia e simulações.

O problema da distribuição do tamanho do pacote é bem conhecida hoje. Modelos existentes de tamanho de pacotes têm provado ser simples e válidos. A maioria desses modelos não consideram a ordem de chegada dos pacotes. Já o problema da distribuição do tempo entre chegada dos pacotes é muito mais difícil, mas o conhecimento do tráfego de rede tem evoluído significativamente nos últimos anos, levando a uma série de evoluções nos modelos de tráfego.

Os modelos mais antigos eram baseados em distribuições de probabilidade simples, assumindo que isso se aplicava a agregações grandes. Por exemplo, distribuições de tráfego Poisson (com tempo entre chegada modelado por uma exponencial) eram frequentemente usadas nesses modelos [6]. Infelizmente, apesar de bem conhecidos, esses modelos nem sempre refletiam o sistema em análise. Tráfegos de redes reais apresentavam rajadas, as quais não apareciam nos modelos. Então novos modelos passaram a gerar tráfegos em rajada [7], [8].

Na década de 90, um grupo na Bellcore coletou uma grande quantidade de traces de tráfego de rede com alta qualidade e fina (ou baixa) granularidade. Análises desses dados mostraram a existência de auto-semelhança (*self-similarity*) [9] o que explicou a discrepância entre os modelos de tráfego e o tráfego de rede real. Modelos de tráfego com auto-semelhança são mais difíceis de ser desenvolvidos e analisados, e é difícil derivar parâmetros para esses modelos de tráfego real.

Contudo, muitos modelos têm evoluído para gerar tráfego com auto-semelhança [10] e têm sido criados sistemas para simplificar a análise de tráfegos existentes a fim de computar parâmetros realísticos para esses modelos [11]

2.2 Tráfego de jogos online

Na seção 1 foi exposto que o Second Life pode ser considerado um jogo, mais precisamente um *massively multiplayer online role playing game* (or MMORPG). Apesar de que o primeiro jogo multiusuário online, PLATO, ter sido desenvolvido em 1970, o primeiro MMORPG, chamado Neverwinter Nights, foi lançado em 1991 pela America Online. Outros jogos foram desenvolvidos em seguida, mas o tráfego de rede gerado pelos usuários destes jogos não era significativo e foi desprezado pelos ISPs até meados de 1999, quando estes jogos começaram a se popularizar e o número de usuários teve um crescimento de mais de 100% em menos de dois meses, atingindo a marca de 1 milhão no final do ano [12]. Desde então, o número de usuários vem crescendo exponencialmente e alguns ISPs passaram a ter como premissa o suporte a ambientes de jogo de alta qualidade para seus usuários. Isso levou todos os ISPs a examinar e avaliar componentes de rede novos, como modems, servidores e roteadores, para garantir o provimento de atrasos consideravelmente baixos [13].

Para desenvolver e avaliar hardwares para ambientes de jogo com baixa latência, são necessários modelos de tráfego de jogos online precisos. Alguns modelos têm sido propostos nos últimos anos: [13], [14], [15], [16], [17], [18], [19] e [20]. Cada um deles é específico para um jogo, e normalmente o jogo que possui o maior número de usuários no momento da publicação. Atualmente, o Second Life é um dos MMORPG com maior popularidade mundial, tendo atualmente mais de 5,4 milhões de usuários [21] e com perspectiva de atingir a marca de 25 milhões em 2008 [22]. Torna-se, portanto, evidente a necessidade de modelar o seu tráfego.

2.3 Metodologia

Em [13], é proposta uma metodologia para modelagem de tráfego de jogos online, a qual é utilizada também por [14]. Baseada nessa metodologia, este trabalho utiliza os passos descritos abaixo:

- Escolher os dados a serem modelados e uma distribuição analítica apropriada.
- Estimar os parâmetros da distribuição escolhida.
- Plotar o gráfico Q-Q Plot da distribuição com os valores estimados.
- Aplicar a métrica da discrepância para verificar a aderência da distribuição analítica.
- Aplicar o teste Kolmogorov-Smirnov.

A modelagem dos dados é realizada utilizando o software R [23], que é uma linguagem e um ambiente para gráficos e cálculos estatísticos. A seguir, serão descritos detalhes referentes aos passos da metodologia.

2.3.1 Escolha dos dados

O primeiro passo para a modelagem é escolher os dados que serão modelados. Para este trabalho, foram coletados 5 traces de 10 minutos, classificados de acordo com a mobilidade do usuário no jogo: andando e parado. Analisando visualmente os histogramas dos traces, observou-se que os mesmos aparentemente seguiam, para cada caso estudado, uma mesma distribuição. Por isso, decidiu-se fazer a modelagem com um dos traces e utilizar os quatro restantes para testar a modelagem. O trace 1 foi escolhido aleatoriamente para a modelagem.

2.3.2 Escolha da distribuição analítica

Para identificar a distribuição analítica que melhor se adequa aos dados escolhidos, o histograma de cada caso em estudo foi analisado e uma ou mais distribuições foram consideradas apropriadas para modelar os dados em questão. Em alguns casos, o histograma foi quebrado em duas ou mais partes, sendo cada parte modelada separadamente.

Os histogramas foram gerados usando a função `hist.scott()`, do pacote estatístico R, que plota o histograma com seleção automática da largura dos bins usando a fórmula de Scott:

$$h_n = 3.49sn^{-\frac{1}{3}}$$

onde n é o número de amostras e s é uma estimativa do desvio padrão.

2.3.3 Estimação dos parâmetros

Uma vez escolhida a distribuição, é necessário encontrar o valor dos parâmetros para os quais os dados melhor se ajustam à distribuição, através da estimação dos parâmetros. Optou-se pelo método da máxima verossimilhança para estimar os parâmetros, o qual é implementado pela função `fitdistr()` no R.

2.3.4 Gráficos Q-Q Plot

Para testar se os dados se adequam à distribuição escolhida com os valores estimados, um gráfico Q-Q Plot é gerado. Neste passo, a função `qqplot()` foi utilizada, tendo como entrada o conjunto de dados em estudo e um conjunto de dados randômicos gerado a partir da distribuição escolhida e dos parâmetros estimados.

2.3.5 Estimação da discrepância do modelo

Segundo [13], testes de aderência tradicionais como Qui-quadrado (X^2) e Kolmogorov-Smirnov normalmente falham com modelos analíticos para tráfego da Internet. Isto ocorre porque tais testes são viesados para conjuntos de dados grandes ou “bagunçados”. O mesmo ocorre com dados que exibem autocorrelação significativa [24]. Por isso, o autor sugere que seja estimada a discrepância entre os dados empíricos e o modelo analítico, ao invés de realizar um teste de aderência. Neste trabalho, serão usados os dois métodos para testar a adequação do modelo.

A métrica da discrepância, proposta por Pederson e Johnson em [25], mede a magnitude da divergência entre os dados e o modelo. Logo, quanto menor a discrepância, melhor o modelo. Basicamente, a ideia desse método é dividir as duas distribuições (empírica e analítica) em bins e fazer uma comparação bin a bin usando o teste X^2 .

De forma detalhada, uma distribuição contínua é discretizada em bins. Ou seja, dadas m observações de uma variável randômica Z , as observações podem ser particionadas em N bins. Seja \tilde{Z}_i o número de observações dentro do i -ésimo bin. O teste qui-quadrado é baseado na comparação do número de observações contidas no bin com o valor esperado do número de observações, de acordo com a distribuição analítica. O valor esperado do número de observações do bin i , representado por E_i , é dado por:

$$E_i = mP[z \in i] = m \int_i f_z(z) dz$$

O teste qui-quadrado é baseado em uma medida da diferença entre o valor esperado e observado do número de observações nos bins, dado por:

$$X^2 = \sum_{i=1}^N \frac{(\tilde{Z}_i - E_i)^2}{E_i}$$

A métrica da discrepância proposta por Pederson e Johnson é calculada pela fórmula abaixo:

$$\hat{\chi}^2 = \frac{X^2 - K - df}{m - 1}$$

onde:

- K é calculado pela fórmula:

$$K = \sum_{i=1}^N \frac{\tilde{Z}_i - E_i}{E_i}$$

- df representa o grau de liberdade, dado por:

$$df = N - 1 - \Theta\left(\hat{\theta}_z\right)$$

onde $\Theta\left(\hat{\theta}_z\right)$ é o número de parâmetros que foi usado para estimar a distribuição analítica.

Foi implementada uma função para calcular a métrica da discrepância, mostrada no Apêndice A.

2.3.6 Teste de Kolmogorov-Smirnov

Como explicado na seção anterior, o teste de Kolmogorov-Smirnov normalmente falha com modelos analíticos para tráfego da Internet. Os traces utilizados neste trabalho são de tráfego coletado na Internet, visto que foram gerados por um jogo online. Entretanto, decidiu-se realizar esse teste, implementado pela função *ks.test()*, para confirmar se o mesmo falharia, como afirmado em [13].

3 Análise Exploratória

Antes da atividade de modelagem, faz-se necessário um entendimento preliminar da natureza dos dados, suas distribuições de probabilidade, distribuições de probabilidade acumulada, e se os dados possuem ou não *outliers* (valores extremos). Esta seção descreve a análise exploratória dos dados realizada no conjunto de traces escolhidos para a modelagem do perfil de tráfego gerado pelo Second Life. As próximas sub-seções apresentam a análise exploratória dos traces escolhidos para modelagem de acordo com cada métrica avaliada.

3.1 Trace Andando-1 Cliente

Este trace representa o comportamento capturado quando um usuário encontra-se andando no mundo virtual e reflete o tráfego que é enviado do cliente para o servidor (tráfego de saída).

3.1.1 Métrica *Inter Arrival*

O primeiro passo da análise exploratória foi a análise visual do histograma. Como podemos observar na Figura 3.1(a), a distribuição do *Inter Arrival* forma uma distribuição bimodal, o que indica que a distribuição não será modelada por apenas uma distribuição. O histograma parece indicar, também, a presença de *outliers*. Na Figura 3.2(b) pode-se confirmar, através do Box-Plot, que realmente existem *outliers* nos dados observados. Uma vez que o *Inter Arrival* é influenciado pela carga de processamento da máquina, decidiu-se excluir esses *outliers*, visto que eles podem ter sido gerados pela influência de alguma tarefa com processamento intensivo por um processo concorrente à aplicação.

Na Figura 3.1(c) pode-se observar o histograma do *Inter Arrival* após a remoção dos *outliers*. Já na Figura 3.1 (d), examinando a ECDF do *Inter Arrival*, nota-se que 35% dos pacotes possuem *Inter Arrival* de até 0.02 ms, enquanto o restante (65%) dos pacotes possuem *Inter Arrival* maiores que 0.02 ms.

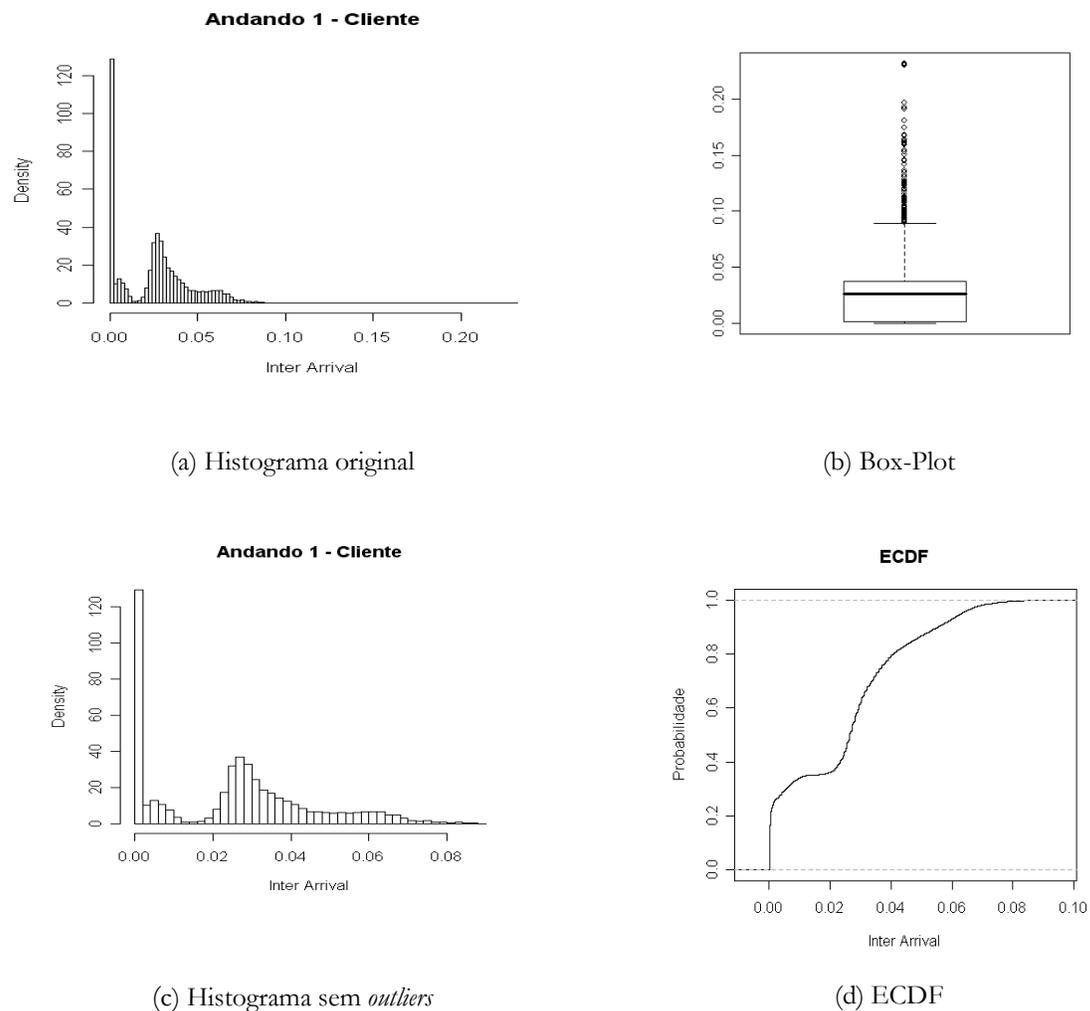


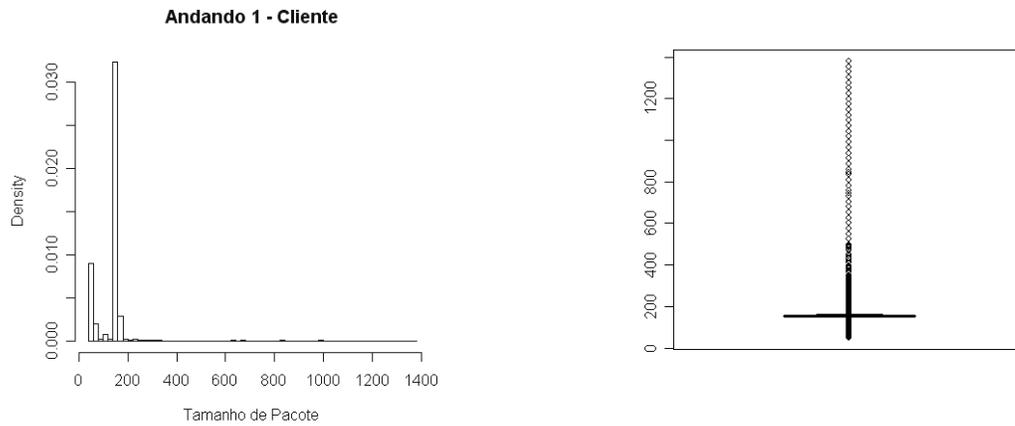
Figura 3.1 Histograma, Box-Plot, Histograma sem *outliers* e ECDF do trace Andando-1 Cliente - métrica *Inter Arrival*

3.1.2 Métrica *Packet Size*

A distribuição do tamanho de pacote gerada pelo cliente por ser observada na Figura 3.2(a). Olhando o histograma, é perceptível a presença de *outliers*, o que é confirmado pelo Box-Plot da

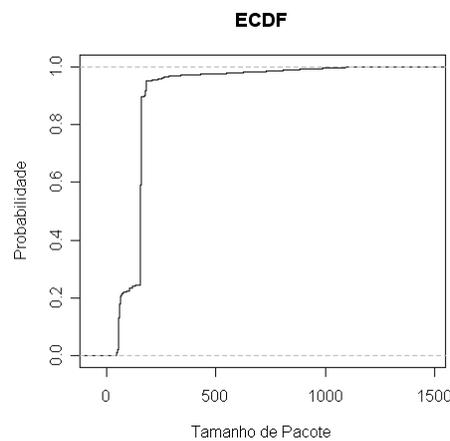
Figura 3.2(b). Mas diferente do *Inter Arrival*, aqui os *outliers* não serão removidos.

Analisando a ECDF presente na Figura 3.2(c) percebe-se que os pacotes gerados pelo cliente são em geral pequenos, visto que aproximadamente 95% dos pacotes têm tamanho inferior a 200 bytes.



(a) Histograma do tamanho de pacotes

(b) Box-Plot tamanho de pacotes



(c) ECDF Tamanho de pacote

Figura 3.2 Histograma, Box-plot, e ECDF do trace Andando-1 Cliente - métrica *Packet Size*

3.2 Trace Andando-1 Servidor

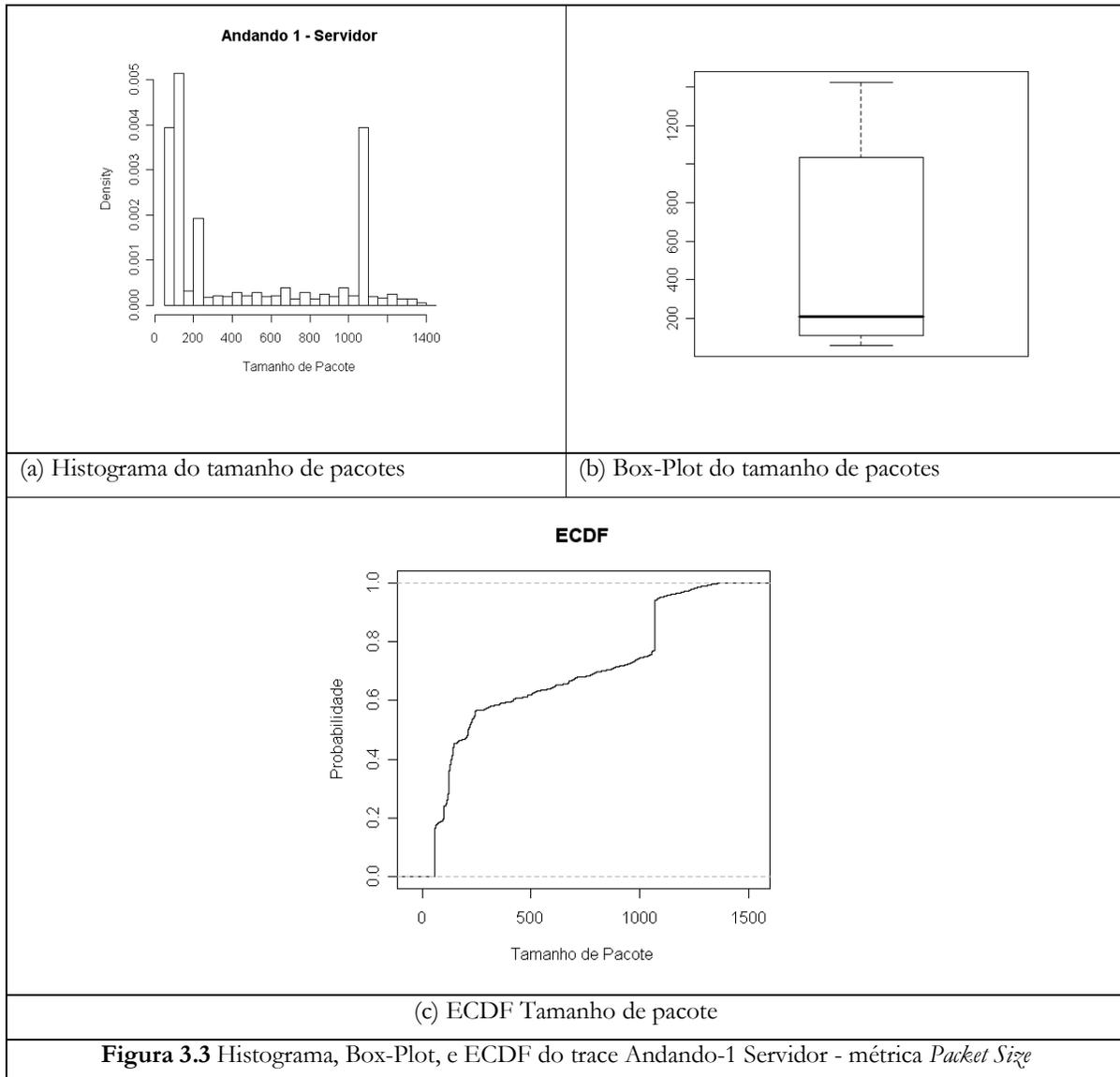
Este trace representa o comportamento capturado quando um usuário encontra-se andando no mundo virtual e reflete o tráfego que é enviado do servidor para o cliente (tráfego de entrada).

3.2.1 Métrica *Packet Size*

Abaixo, na Figura 3.3(a), encontra-se a distribuição de tamanho de pacotes gerado pelo servidor. Observando o histograma, pode-se perceber que se trata de uma distribuição bimodal e, conforme o Box-Plot da Figura 3.3(b), sem a presença de *outliers*.

Pela análise da Figura 3.4(c), nota-se que aproximadamente 55% dos pacotes têm até 250 bytes, enquanto o restante (45%) dos pacotes têm tamanho superior a 250 bytes. O confronto das Figuras 3.2(c) e 3.3(c) apoia a afirmação em [26] de que os pacotes do cliente para o servidor são em

geral muito menores que os pacotes do servidor para o cliente e possuem menor variação no seu tamanho.



3.3 Traces Parado (Cliente e Servidor)

Devido a restrições de espaço, e à similaridade entre as distribuições de *Inter Arrival* e *Packet Size* dos traces ANDANDO e PARADO, tais distribuições para a condição PARADO não serão comentadas. Contudo, os gráficos podem ser encontrados no Apêndice B.

4 Modelagem dos traces

Esta seção descreve a aplicação da metodologia apresentada na seção 2, no processo de modelagem dos traces de tráfego do jogo Second Life descritos na seção 3. As subseções seguintes detalham a modelagem dos traces de acordo com o sentido do tráfego. Primeiramente, serão modelados os traces no sentido Cliente-Servidor (chamados apenas de Cliente) nas situações PARADO e

ANDANDO, e métricas *Inter Arrival* e *Packet Size*. Em seguida, os traces no sentido Servidor-Cliente (chamados de Servidor) nas mesmas situações, com as mesmas métricas.

4.1 Trace Parado 1 - Cliente

4.1.1 Métrica *Inter Arrival*

A análise visual do histograma dos dados revela que: primeiro, não é possível presumir normalidade dos dados; segundo, percebe-se que a melhor opção para modelagem é dividir a distribuição em intervalos para permitir uma melhor captura das características. A Figura 4.1 mostra a divisão do histograma.

Optou-se por segmentar a distribuição em quatro (4) intervalos distintos. O primeiro apresenta valores de X menores ou iguais a 0.001, o que pela análise visual parece obedecer a um modelo determinístico. O segundo intervalo assume valores de X entre 0.001 e 0.02, pois, visualmente, lembra distribuições de calda longa como exponencial. O terceiro intervalo compreende a valores de X entre 0.02 e 0.08, e o quarto, valores acima de 0.08. Visualmente, ambos podem ser avaliados como normais.

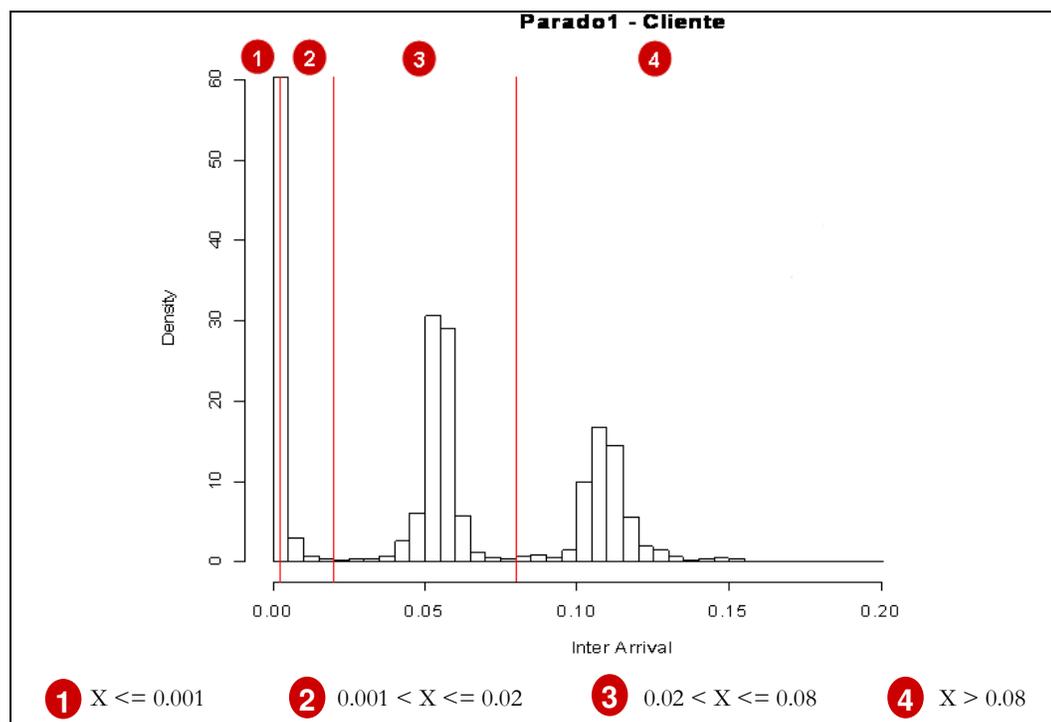


Figura 4.1 Histograma do trace Parado-1 Cliente - métrica *Inter Arrival*.

É importante ressaltar que a escolha de determinada distribuição para avaliar um intervalo não implica que tal distribuição deverá ser obrigatoriamente aplicada aos outros intervalos.

Intervalo $x \leq 0.001$

O intervalo para valores de X são menores que 0.001 aparece no histograma completo (Figura 4.1) como um pico. A métrica avaliada nessa seção é o *Inter Arrival*, ou seja, a diferença entre o tempo de chegada de dois pacotes consecutivos. Para um jogo online interativo, esse tempo deve ser o menor possível, para que qualquer movimento de um jogador ou mudança no ambiente causada por ele seja percebida praticamente em tempo real pelos outros jogadores. Portanto, é esperado (e desejável) que esse tempo esteja bem próximo de zero, o que realmente ocorre se não houver gargalos no caminho.

Isso justifica o pico para valores menores que ou iguais a 0.001.

De acordo com a metodologia adotada, é preciso escolher uma distribuição ou distribuições capazes de modelar o intervalo e, então, estimar seus parâmetros. A idéia inicial de modelar este intervalo como sendo determinístico foi descartada após a plotagem do histograma dos dados deste intervalo, conforme a Figura 4.2. Sendo assim, foram testadas as distriuições Weibull, Exponencial e Gama. Entretanto, a função de estimação dos parâmetros *fitdistr()* não convergiu para ambas as distribuições. Somente a distribuição Exponencial conseguiu ter seu parâmetro estimado, com valor 3617.217.

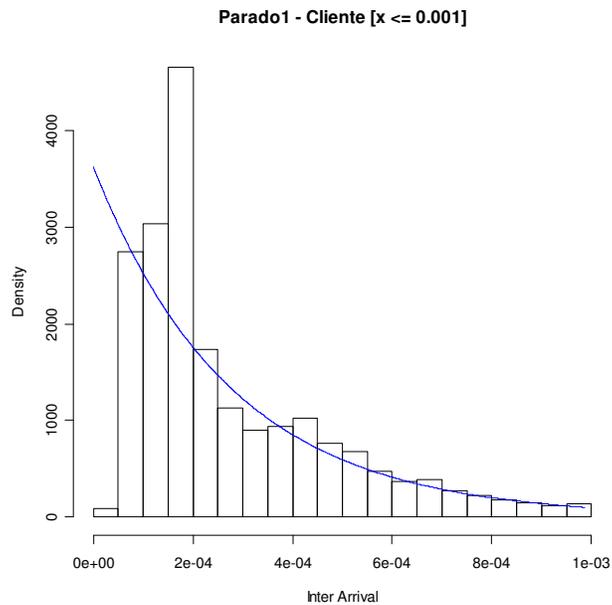
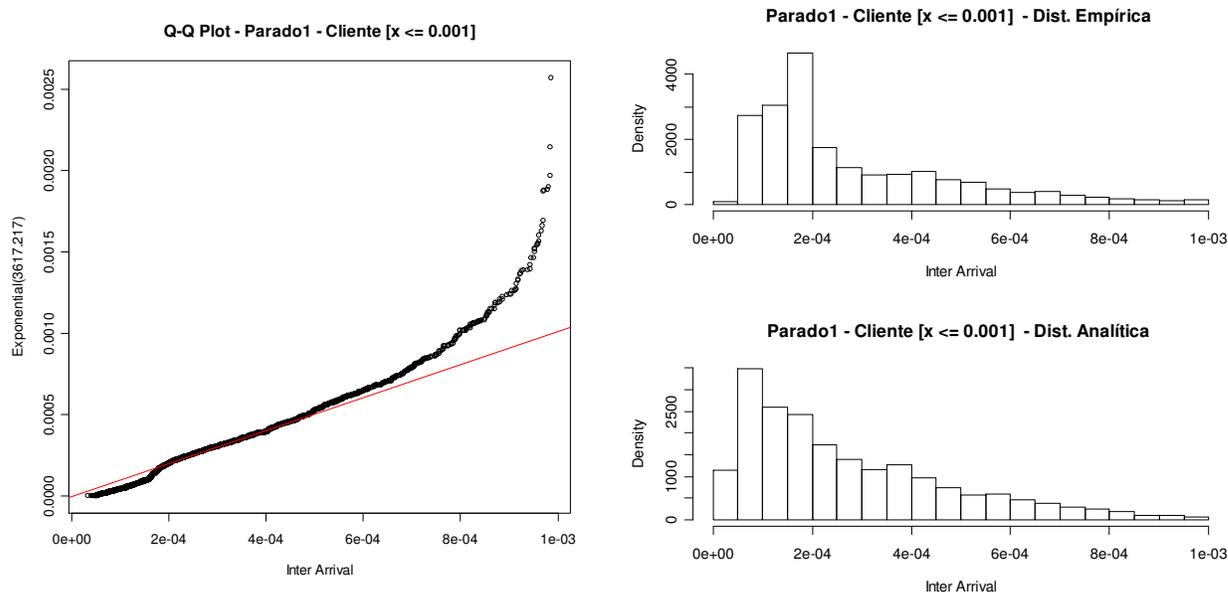


Figura 4.2 Histograma do trace Parado-1 Cliente, para $X < 0.001$ - métrica *Inter Arrival* e a função de densidade de probabilidade de uma Exponencial(3617.217)

Uma vez estimado o parâmetro, o passo seguinte é plotar o gráfico dos quantis para verificar a aderência da distribuição escolhida sobre os dados empíricos (etapa 3 da metodologia). A Figura 4.3(a) mostra o gráfico Q-Q Plot para a distribuição Exponencial.



(a) Q-Q Plot

(b) Histograma das distribuições empírica versus analítica

Figura 4.3 Q-Q Plot e Histograma do trace Parado1 - Cliente, para $X \leq 0.001$ - métrica *Inter Arrival* versus dados randomicamente gerados seguindo uma $\text{Exponential}(3617.217)$

O passo seguinte da metodologia é, através do parâmetro estimado, realizar o teste de discrepância. O resultado da discrepância para distribuição Exponencial tem valor igual a 0.174, o que indica um bom ajuste da distribuição escolhida aos dados. A Figura 4.3(b) ilustra a comparação bin-a-bin entre as distribuições empírica e analítica, realizada no cálculo da discrepância.

E para concluir a seqüência da metodologia, foi realizado o teste de aderência Kolmogorov-Smirnov. Como esperado, de acordo com as afirmações realizadas por [13] (ver Seção 2.3.5), o teste falhou.

Intervalo $0.001 < x \leq 0.02$

O histograma da segunda partição, composta pelos valores de X que estão entre 0.001 e 0.02, encontra-se ilustrado na Figura 4.4.

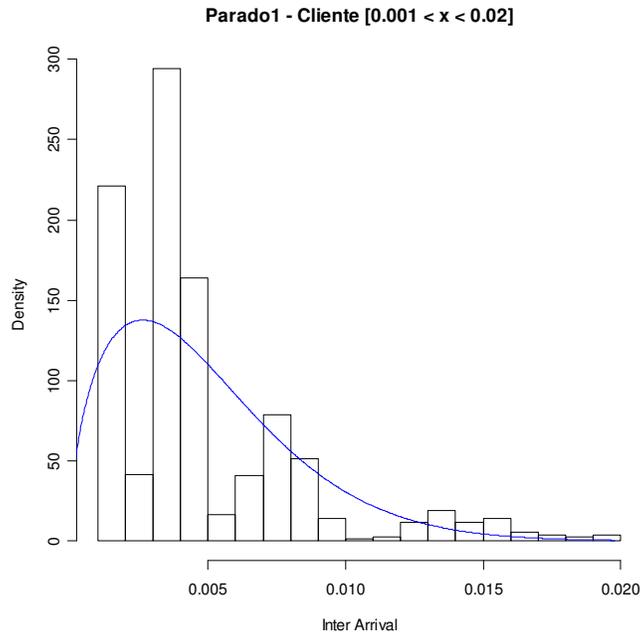


Figura 4.4 Histograma do trace Parado-1 Cliente, para $0.001 \leq X < 0.02$ - métrica *Inter Arrival* e a função de densidade de probabilidade de uma Weibull(1.505, 0.005)

Para o intervalo entre 0.001 e 0.02 foram testadas as distriuições Weibull, Exponencial e Gama. A Tabela 4.1 contém os estimadores de cada distribuição obtidos através da função *fitdistr()* do R.

Tabela 4.1 Estimadores das distribuições do trace Parado1-Cliente para o intervalo entre 0.001 a 0.02 – métrica *Inter Arrival*.

Distribuição	Shape	Scale	Rate
Weibull	1.505	0.005	-
Exponencial	-	-	208.078
Gama	1.576	0.003	-

Com os parâmetros estimados, gera-se o gráfico dos quantis para verificar a aderência das distribuições escolhidas sobre os dados empíricos. A Figura 4.5 mostra o gráfico Q-Q Plot para todas as distribuições avaliadas. Fica claro que nenhuma das distribuições mostra-se ajustada aos dados, entretanto, pode-se preceber que a distribuição Weibull é a que melhor se adequa.

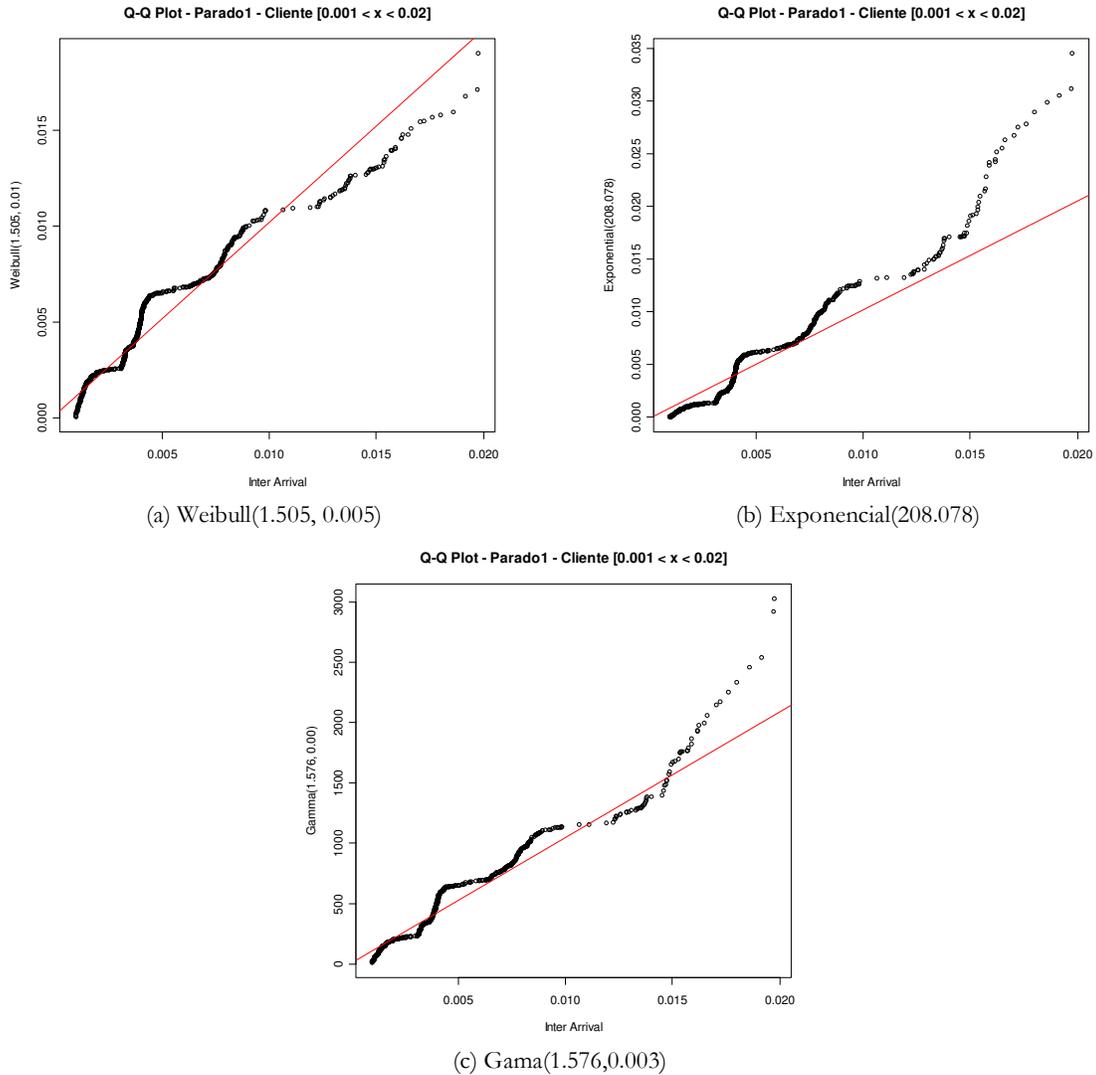
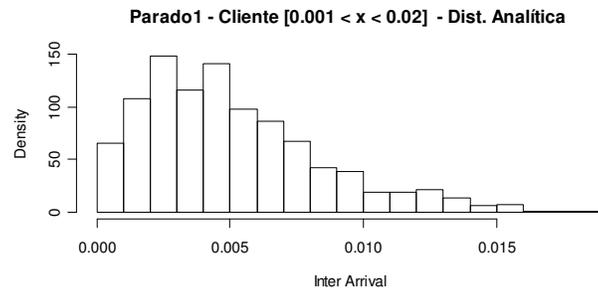
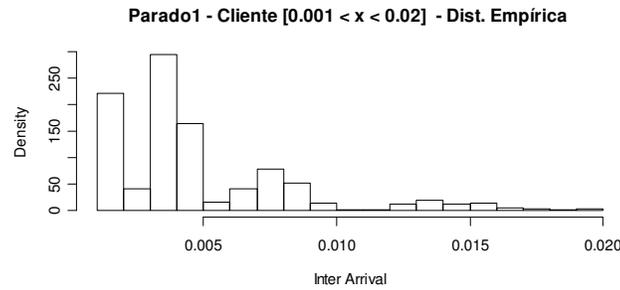
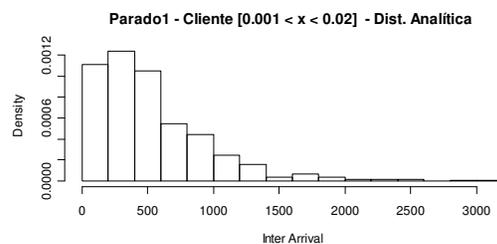
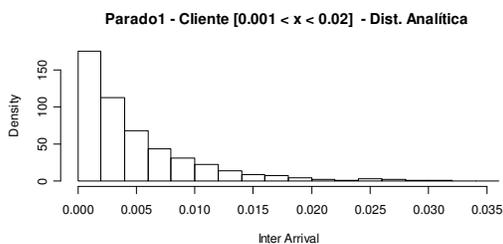
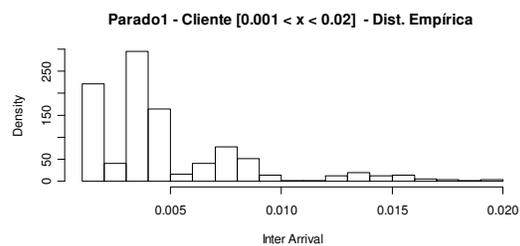
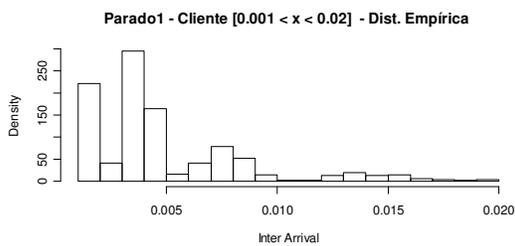


Figura 4.5 Q-Q-Plot das distribuições analíticas do trace Parado-1 Cliente, para $0.001 \leq X < 0.02$ - métrica *Inter Arrival*.

Em seguida, também com os parâmetros estimados, é realizado o teste de discrepância. O resultado da discrepância, aponta a distribuição Weibull como a de melhor resultado, tendo registrado 0.787. As distribuições Exponencial e Gama apresentaram discrepância de 1.094 e 1.512, respectivamente. A Figura 4.6 ilustra a comparação bin-a-bin entre as distribuições empírica e analítica, realizada no cálculo da discrepância, para as distribuições Weibull (a), Exponencial (b) e Gama (C).



(a) Weibull(1.505, 0.005)



(b) Exponencial(208.078)

(c) Gama(1.576,0.003)

Figura 4.6 Histograma das distribuições empírica versus analítica do Parado-1 Cliente, para $0.001 \leq X < 0.02$ - métrica *Inter Arrival*.

No último passo da metodologia, foi realizado o teste de aderência Kolmogorov-Smirnov. Como esperado, de acordo com as afirmações realizadas por [13] (ver Seção 2.3.5), o teste falhou para todas as distribuições.

Intervalo $0.02 \leq x \leq 0.08$

O histograma da erceira partição, composta pelos valores de X que estão entre 0.02 e 0.08, encontra-se ilustrado na Figura 4.7.

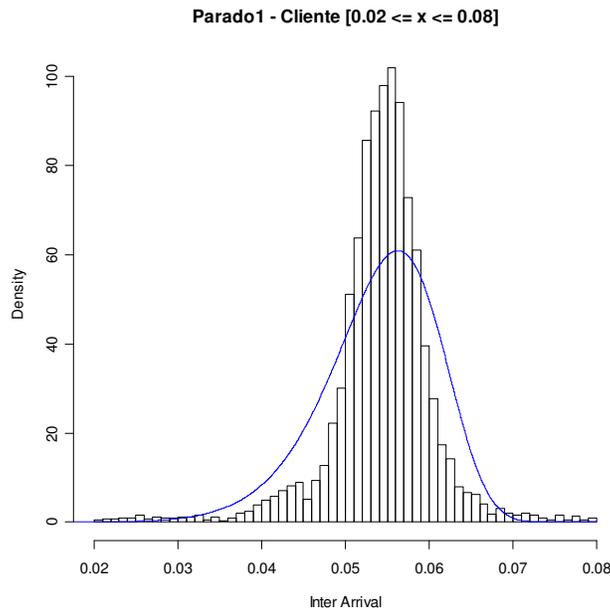


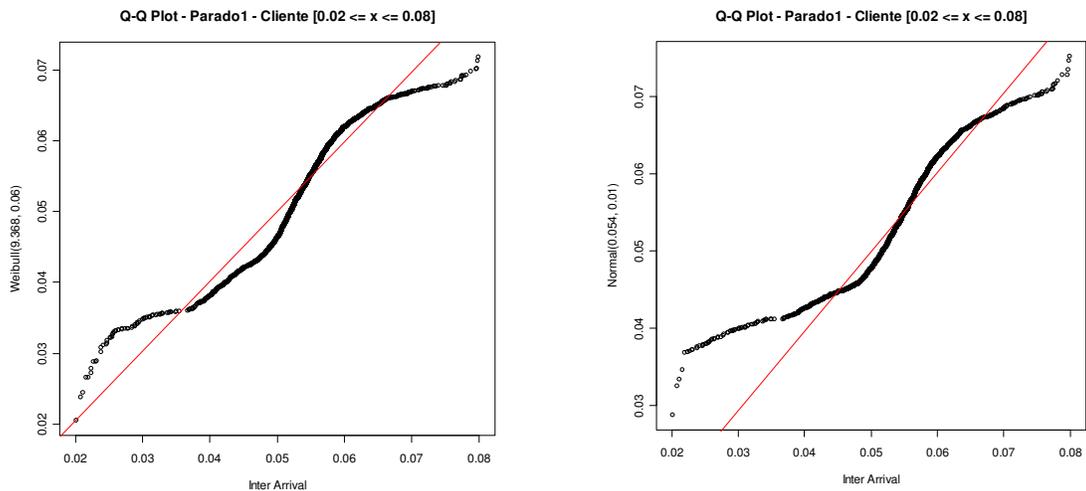
Figura 4.7 Histograma do trace Parado-1 Cliente, para $0.02 \leq X \leq 0.08$ - métrica *Inter Arrival* e a função de densidade de probabilidade de uma Weibull(9.368, 0.057)

Para o intervalo entre 0.02 e 0.08 foram testadas as distriuições Weibull, Normal e Gama. A Tabela 4.2 apresenta os estimadores de cada distribuição.

Tabela 4.2 Estimadores das distribuições do trace Parado1-Cliente para o intervalo entre 0.02 a 0.08 – métrica *Inter Arrival*.

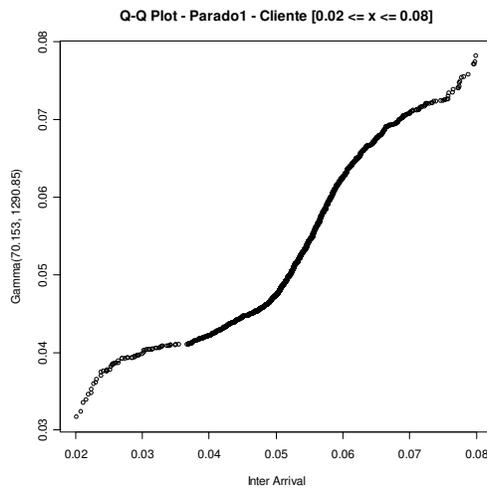
Distribuição	Shape	Scale	Mean	Std. Dev.
Weibull	9.368	0.057	-	-
Normal	-	-	0.054	0.006
Gama	70.153	1290.849	-	-

Com os parâmetros estimados, foram plotados os gráficos dos quantis de cada distribuição para verificar a aderência, conforme a Figura 4.8.



(a) Weibull(9.368, 0.057)

(b) Normal(0.054, 0.006)

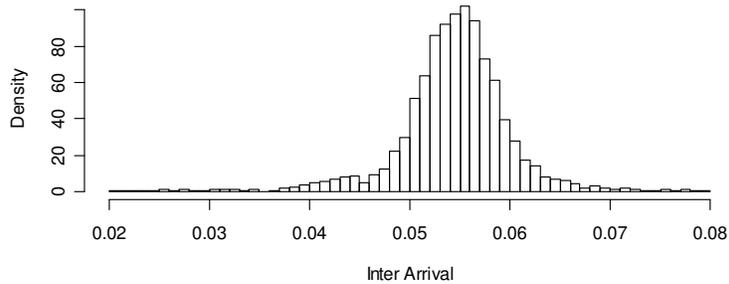


(c) Gama(70.153, 1290.849)

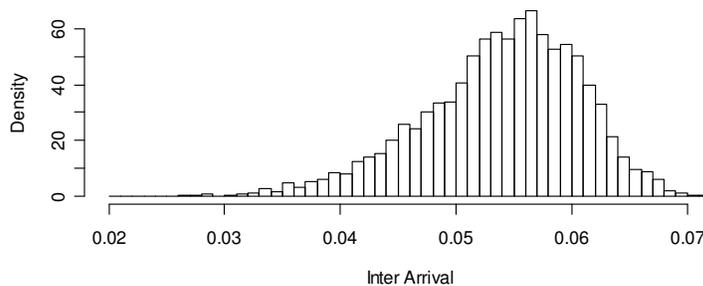
Figura 4.8 Q-Q-Plot das distribuições analíticas do trace Parado-1 Cliente, para $0.02 \leq X < 0.08$ - métrica *Inter Arrival*

Por fim, o teste de discrepância aponta a distribuição Weibull com o melhor resultado, 0.257. As distribuições Normal e Gama apresentaram 2.878 e 6.313, respectivamente. A Figura 4.9 representa a comparação bin-a-bin entre as distribuições empírica e analítica, realizada no cálculo da discrepância, para as distribuições Weibull (a), Normal (b) e Gama (C).

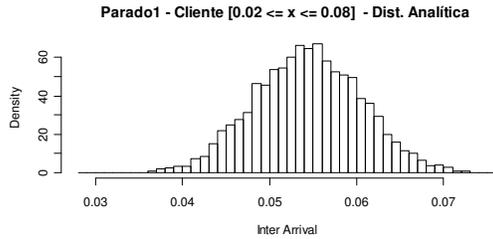
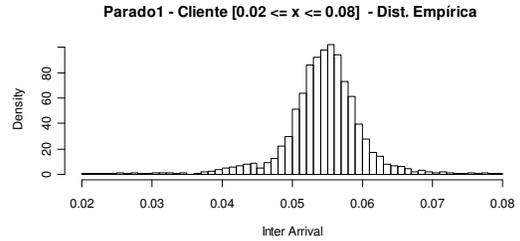
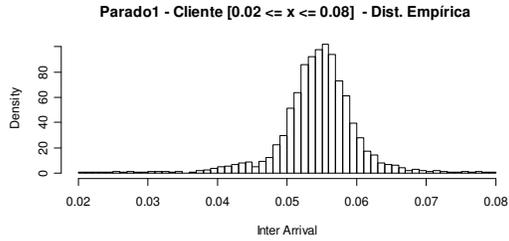
Parado1 - Cliente [0.02 <= x <= 0.08] - Dist. Empírica



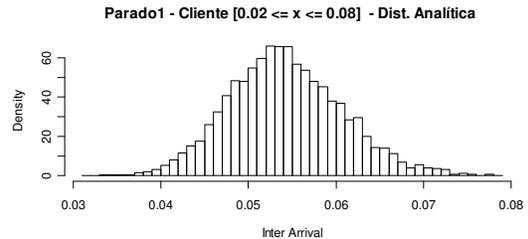
Parado1 - Cliente [0.02 <= x <= 0.08] - Dist. Analítica



(a) Weibull(9.368, 0.057)



(b) Normal(0.054, 0.006)



(c) Gama(70.153, 1290.849)

Figura 4.9 Histograma das distribuições empírica versus analítica do Parado-1 Cliente, para $0.02 \leq X < 0.08$ - métrica *Inter Arrival*.

Para concluir a seqüência da metodologia, foi realizado o teste de aderência Kolmogorov-Smirnov. Como esperado, de acordo com as afirmações realizadas por [13] (ver Seção 2.3.5), o teste falhou para todas as distribuições.

Intervalo $x > 0.08$

O histograma da terceira partição, composta pelos valores de X maiores que 0.08, encontra-se ilustrado na Figura 4.10.

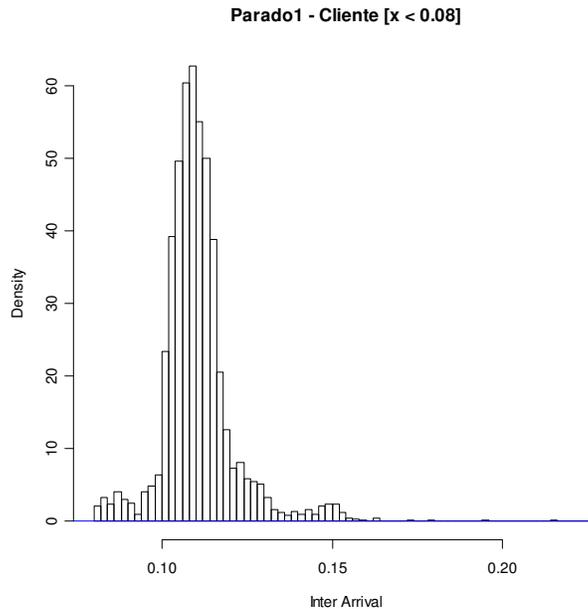


Figura 4.10 Histograma do trace Parado-1 Cliente, para $X > 0.08$ - métrica *Inter Arrival* e a função de densidade de probabilidade de uma Gama(21.060,0.005)

Para valores de x maior que 0.08 foram testadas as distriuições Weibull, Normal e Gama. A Tabela 4.3 apresenta os estimadores de cada distribuição.

Tabela 4.3 Estimadores das distribuições do trace Parado1-Cliente para valores de $X > 0.08$ – métrica *Inter Arrival*.

Distribuição	Shape	Scale	Mean	Std. Dev.
Weibull	7.220	0.116	-	-
Normal	-	-	0.110	0.011
Gama	21.060	0.005	-	-

Com os parâmetros estimados, foram plotados os gráficos dos quantis de cada distribuição para verificar a aderência, conforme a Figura 4.11.

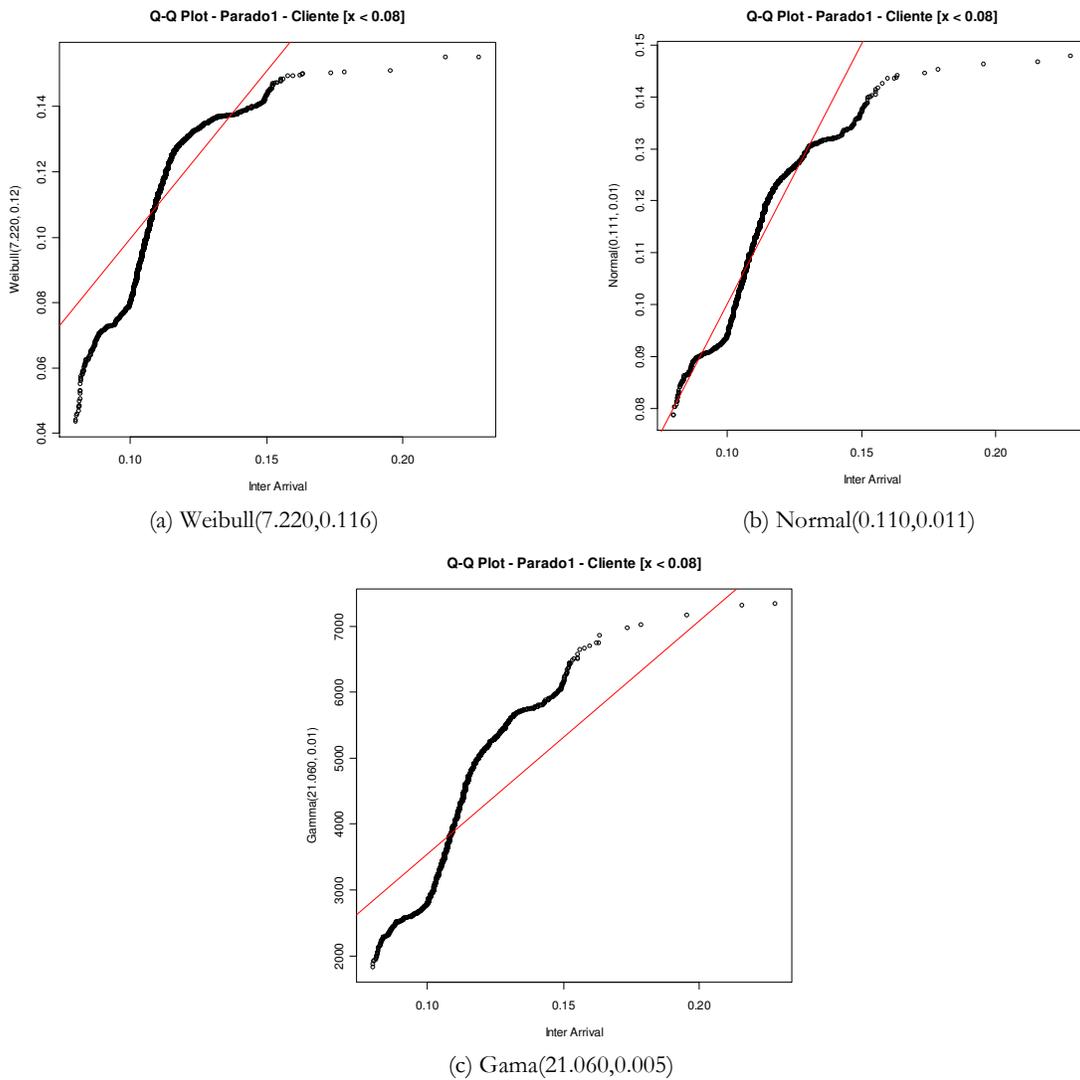


Figura 4.11 Q-Q-Plot das distribuições analíticas do trace Parado-1 Cliente, para $X > 0.08$ - métrica *Inter Arrival*.

O teste de discrepância apresentou um resultado diferente dos outros intervalos. A

Como esperado, de acordo com as afirmações realizadas por [13] (ver Seção 2.3.5), o teste falhou para todas as distribuições.

A Tabela 4.4 apresenta os resultados finais da modelagem do trace Parado-1 sentido Cliente. Em outras palavras, apenas as distribuições de menor discrepância para cada intervalo são mostradas na tabela.

Tabela 4.4 Modelagem para o trace Parado 1 - Cliente.

Intervalo	Distribuição	Shape	Scale	Rate	Discrepância
$x < 0.001$	Exponencial	-	-	3617.217	0.174
$0.001 < x \leq 0.02$	Weibull	1.505	0.005	-	0.787
$0.02 < x \leq 0.08$	Weibull	9.368	0.057	-	0.257
$X > 0.08$	Gama	21.060	0.005	-	1.657

4.1.2 Métrica *Packet Size*

O histograma dos dados para a métrica *Packet Size* encontra-se ilustrado na Figura 4.13(a). Observando o histograma, percebe-se facilmente que alguns tamanhos de pacotes são bastante freqüentes, o que resulta nos picos sinalizados com setas na Figura 4.13(a). Tais picos levam à conclusão de que o modelo determinístico é o mais adequado para modelar os dados deste trace.

Para descobrir quais são esses valores de maior freqüência, foi gerado um gráfico que plota os dados do trace, i.e., plota a seqüência dos tamanhos dos pacotes coletados no trace. Neste gráfico, ilustrado na Figura 4.13(b), é possível perceber que existem vários tamanhos diferentes para os pacotes coletados no trace analisado, contudo, alguns apresentam uma freqüência muito superior à maioria, que correspondem às linhas que aparecem no gráfico. Os 4 (quatro) tamanhos que apresentam uma freqüência significativa resultam nas linhas escuras da Figura 4.13(b) e nos picos do histograma na Figura 4.13(a)

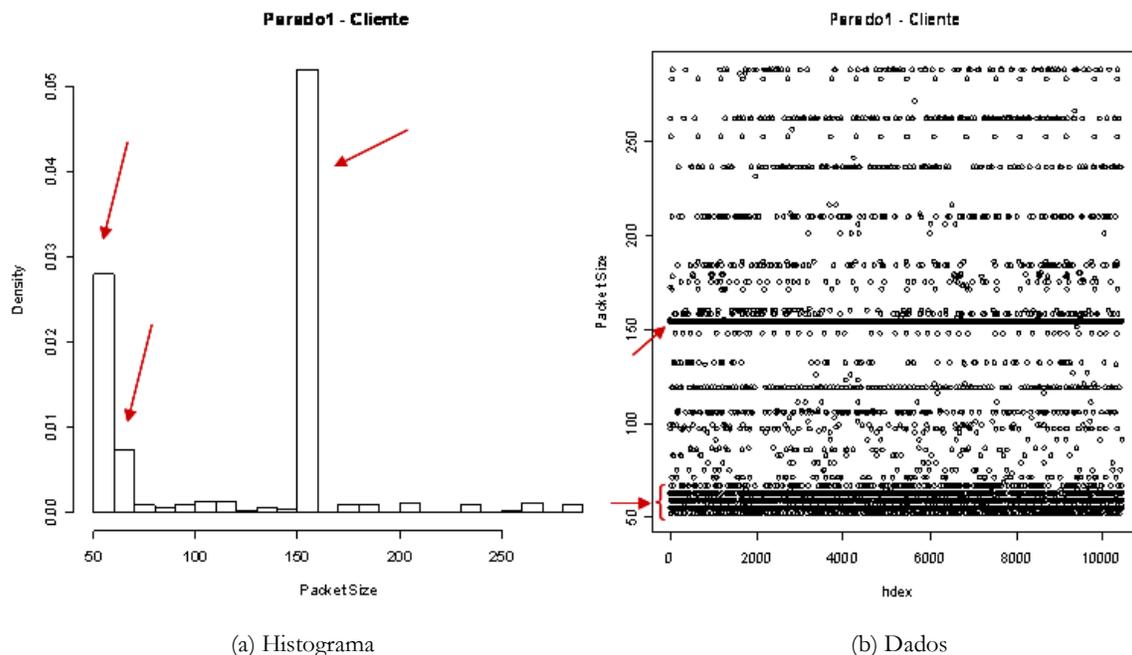


Figura 4.13 Histograma e dados do trace Parado-1 Cliente - métrica *Packet Size*.

A análise numérica dos dados revela que: 14.9% dos pacotes avaliados neste trace possuem tamanho igual a 55 bytes; 10.4% possuem tamanho igual a 59 bytes; 5.1% possuem tamanho igual a 63 bytes e 50.4% possuem tamanho igual a 154 bytes. Juntos, estes 4 tamanhos de pacotes representam 80.8% do trace, enquanto os demais tamanhos, que representam 19.2%, não atingem a frequência de 2% cada.

4.2 Trace Andando-1 Cliente

4.2.1 Métrica *Inter Arrival*

Analisando visualmente o histograma dos dados do trace, é possível observar que estes não seguem uma única distribuição. Em outras palavras, é necessário dividir os dados em intervalos para melhor modelá-los. Para determinar os limites dos intervalos, foi realizada uma análise detalhada dos dados, através de histogramas, como descrito a seguir.

Examinando o histograma completo, observou-se os pontos em que sua forma mudava e, a partir desses pontos estimados, foram plotados novos histogramas, a fim de determinar com maior precisão tais pontos. Dessa forma, os dados foram divididos em três intervalos, segmentando o histograma como mostra a Figura 4.14.

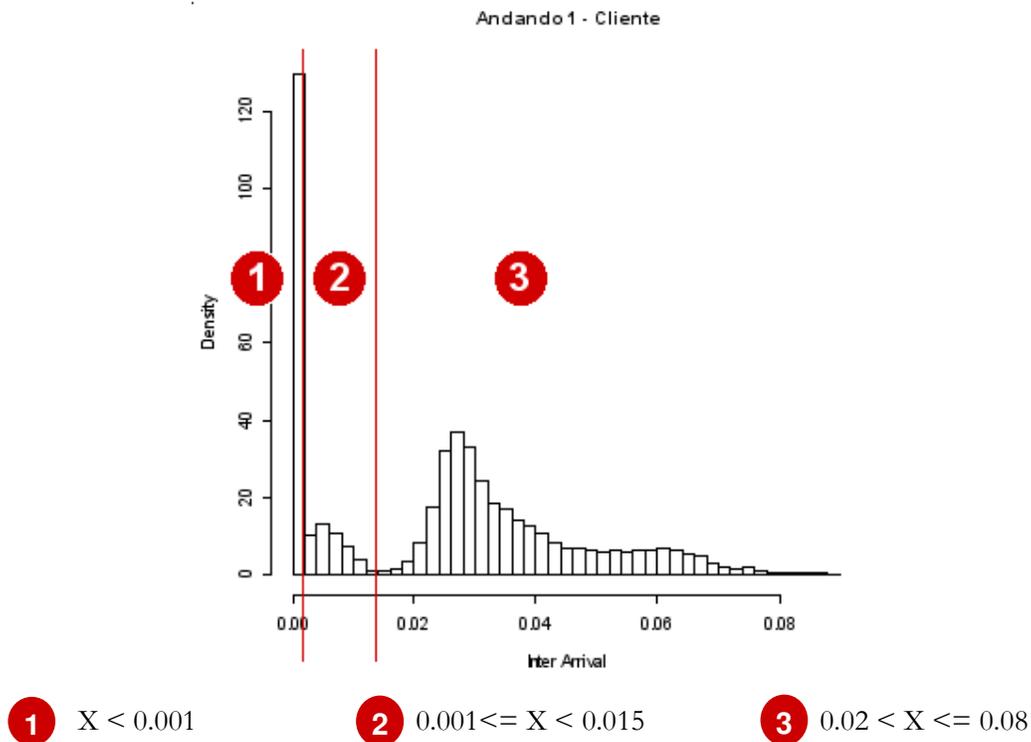


Figura 4.14 Histograma do trace Andando-1 Cliente - métrica *Inter Arrival*.

Intervalo $X < 0.001$

A primeira partição, isto é, o intervalo em que os valores de X são menores que 0.001, aparece no histograma completo como um pico. Como explicado na seção 4.1.1, a métrica avaliada nessa seção é o *Inter Arrival*, ou seja, a diferença entre o tempo de chegada de dois pacotes consecutivos. Para

um jogo online interativo, esse tempo deve ser o menor possível, para que qualquer movimento de um jogador ou mudança no ambiente causada por ele seja percebida praticamente em tempo real pelos outros jogadores. Portanto, é esperado (e desejável) que esse tempo esteja muito próximo de zero, o que realmente ocorre se não houver atrasos no caminho. Isso justifica o pico para valores menores que 0.001.

Analisando separadamente este trecho, obtém-se o histograma ilustrado na Figura 4.15. Através da análise visual, as distribuições Weibull, Gamma e Exponencial foram escolhidas como candidatas para modelar os dados. Contudo, no passo 2 da metodologia (estimação dos parâmetros), as distribuições Weibull e Gamma foram descartadas, pois a função *fitdistr()* não convergiu para ambas as distribuições. Assim, apenas a distribuição Exponencial foi utilizada nesta modelagem.

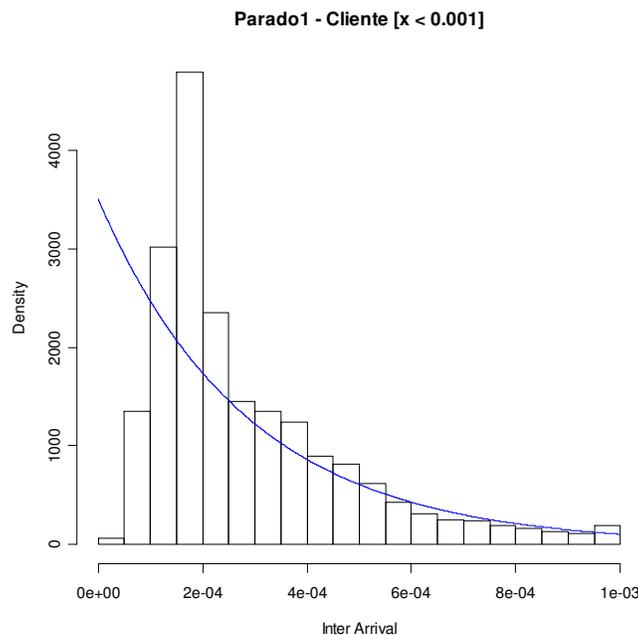
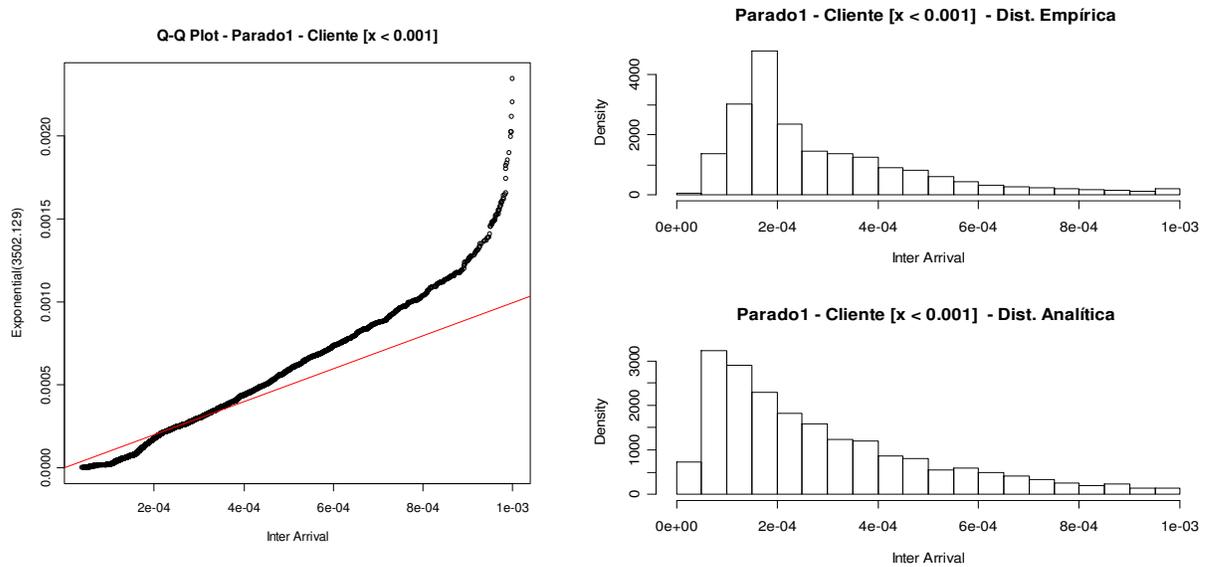


Figura 4.15 Histograma do trace Andando-1 Cliente, para $X < 0.001$ - métrica *Inter Arrival* e a função de densidade de probabilidade de uma Exponencial(3502.129)

A estimação do parâmetro para a Exponencial resultou no valor 3502.129. A Figura 4.15 apresenta (através da linha azul) a função de densidade de probabilidade da distribuição Exponencial com o parâmetro estimado sobre o histograma dos dados.

No próximo passo, o gráfico Q-Q Plot é gerado, como ilustrado na Figura 4.16(a). Percebe-se pelo Q-Q Plot que a distribuição se encaixa perfeitamente no intervalo entre 0.0002 e 0.0004 ms. A partir desse ponto, a aderência vai diminuindo. Contudo, a métrica da discrepância, calculada no passo seguinte, afirma que o modelo pode ser considerado muito bom, visto que seu valor é 0.242. A comparação bin-a-bin entre as distribuições empírica e analítica, realizada no cálculo da discrepância, pode ser visualizada na Figura 4.16(b).



(a) Q-Q Plot (b) Histograma das distribuições empírica versus analítica

Figura 4.16 Q-Q Plot e Histograma do trace Andando-1 Cliente, para $X < 0.001$ - métrica *Inter Arrival* versus dados randomicamente gerados seguindo uma *Exponencial(3502.129)*

Para concluir a sequência da metodologia, foi realizado o teste de aderência Kolmogorov-Smirnov. Como esperado, de acordo com as afirmações realizadas por [13] (ver Seção 2.3.5), o teste falhou.

Intervalo $0.001 \leq X < 0.015$

O histograma da segunda partição, composta pelos valores de X que estão entre 0.001 e 0.015, encontra-se ilustrado na Figura 4.17.

Através da análise visual do histograma abaixo, as distribuições Weibull, Gamma e Exponencial foram escolhidas como candidatas para modelar os dados. Todos os passos da modelagem foram realizados para cada uma dessas distribuições, contudo apenas a modelagem que apresentou a menor discrepância, a Exponencial, será apresentada nesta seção. Os gráficos referentes às demais estão no Apêndice C.

A estimação do parâmetro para a Exponencial resultou no valor 205.841. A Figura 4.17 apresenta (através da linha azul) a função de densidade de probabilidade da distribuição Exponencial com o parâmetro estimado sobre o histograma dos dados.

Através da métrica da discrepância, calculada no passo seguinte, pode-se afirmar que o modelo é muito bom, pois o valor da discrepância é 0.376. A Figura 4.18(b) ilustra a comparação bin-a-bin entre as distribuições empírica e analítica, realizado no cálculo da discrepância.

No último passo da metodologia, foi realizado o teste de aderência Kolmogorov-Smirnov. Como esperado, de acordo com as afirmações realizadas por [13] (ver Seção 2.3.5), o teste falhou.

Intervalo $X < 0.001$

O histograma da última partição deste trace, isto é, do intervalo em que os valores de X são maiores que 0.015, encontra-se ilustrado na Figura 4.19. Através da análise visual do histograma, as distribuições Weibull, Gamma e Exponencial foram escolhidas como candidatas para modelar os dados. Todos os passos da modelagem foram realizados para cada uma dessas distribuições, porém apenas a modelagem que apresentou a menor discrepância, a Exponencial, será apresentada nesta seção. Os gráficos referentes às demais estão no Apêndice C.

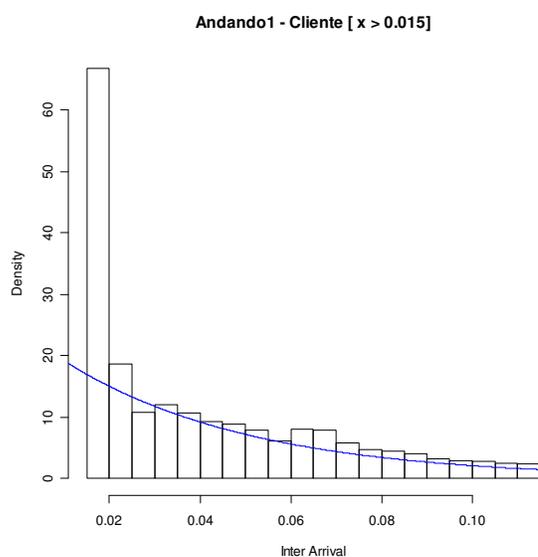


Figura 4.19 Histograma do trace Andando-1 Cliente, para $X > 0.015$ - métrica *Inter Arrival* e a função de densidade de probabilidade de uma Exponencial(24.505)

A estimação do parâmetro para a Exponencial, realizada no segundo passo da metodologia, resultou no valor 24.505. A função de densidade de probabilidade da distribuição Exponencial com o parâmetro estimado encontra-se plotada sobre o histograma dos dados na Figura 4.19.

O gráfico dos quantis foi plotado no próximo passo. A Figura 4.20(a) ilustra o Q-Q Plot da distribuição dos dados do trace no intervalo estudado *versus* a Exponencial(24.505). Percebe-se pelo Q-Q Plot que a distribuição se encaixa perfeitamente no intervalo entre 0.02 e 0.055 ms. A partir desse ponto, a aderência vai diminuindo. Contudo, a métrica da discrepância, calculada no passo seguinte, afirma que o modelo pode ser considerado muito bom, visto que seu valor é 0.358. A comparação bin-a-bin entre as distribuições empírica e analítica, realizada no cálculo da discrepância, pode ser visualizada na Figura 4.20(b).

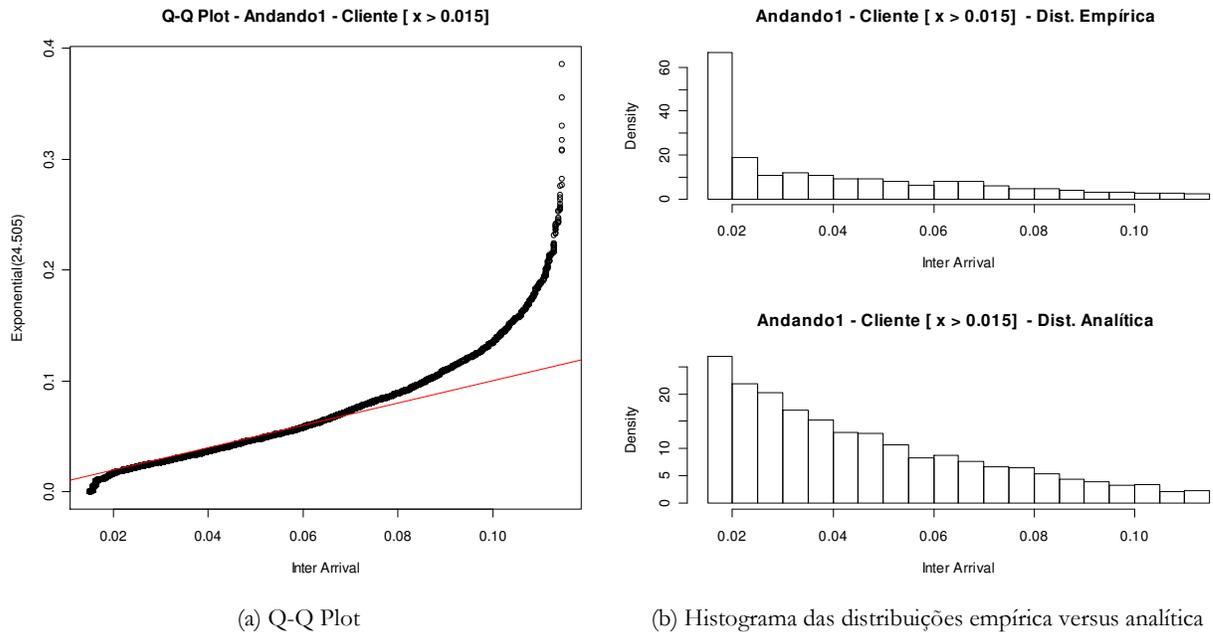


Figura 4.20 Q-Q Plot e Histograma do trace Andando-1 Cliente, para $X \leq 0.015$ - métrica *Inter Arrival* versus dados aleatoriamente gerados seguindo uma Exponencial(24.505)

Para concluir a seqüência da metodologia, foi realizado o teste de aderência Kolmogorov-Smirnov. Como esperado, de acordo com as afirmações realizadas por [13] (ver Seção 2.3.5), o teste falhou.

A Tabela 4.5 apresenta os resultados finais da modelagem do trace Andando-1 sentido Cliente. Em outras palavras, apenas as distribuições de menor discrepância para cada intervalo são mostradas na tabela.

Tabela 4.5 Modelagem para o trace ANDANDO 1 Cliente.

Intervalo	Distribuição	Rate	Discrepância
$X < 0.001$	Exponencial	3502.129	0.242
$0.001 < X \leq 0.015$	Exponencial	205.841	0.376
$X > 0.015$	Exponencial	24.505	0.358

4.2.2 Métrica *Packet Size*

O histograma dos dados para a métrica *Packet Size* encontra-se ilustrado na Figura 4.21(a). Observando o histograma, percebe-se facilmente que alguns tamanhos de pacotes são bastante freqüentes, o que resulta nos picos sinalizados com setas na Figura 4.21(a). Tais picos levam à conclusão de que o modelo determinístico é o mais adequado para modelar os dados deste trace.

Para descobrir quais são esses valores de maior freqüência, foi gerado um gráfico que plota os dados do trace, i.e., plota a seqüência dos tamanhos dos pacotes coletados no trace. Neste gráfico, ilustrado na Figura 4.21(b), é possível perceber que existem apenas 7 tamanhos diferentes para os pacotes coletados no trace analisado, que correspondem às linhas que aparecem no gráfico. E desses 7 (sete) valores, apenas 4 (quatro) apresentam uma freqüência significativa, resultando nas linhas escuras da Figura 4.21(b) e nos picos do histograma na Figura 4.21(a).

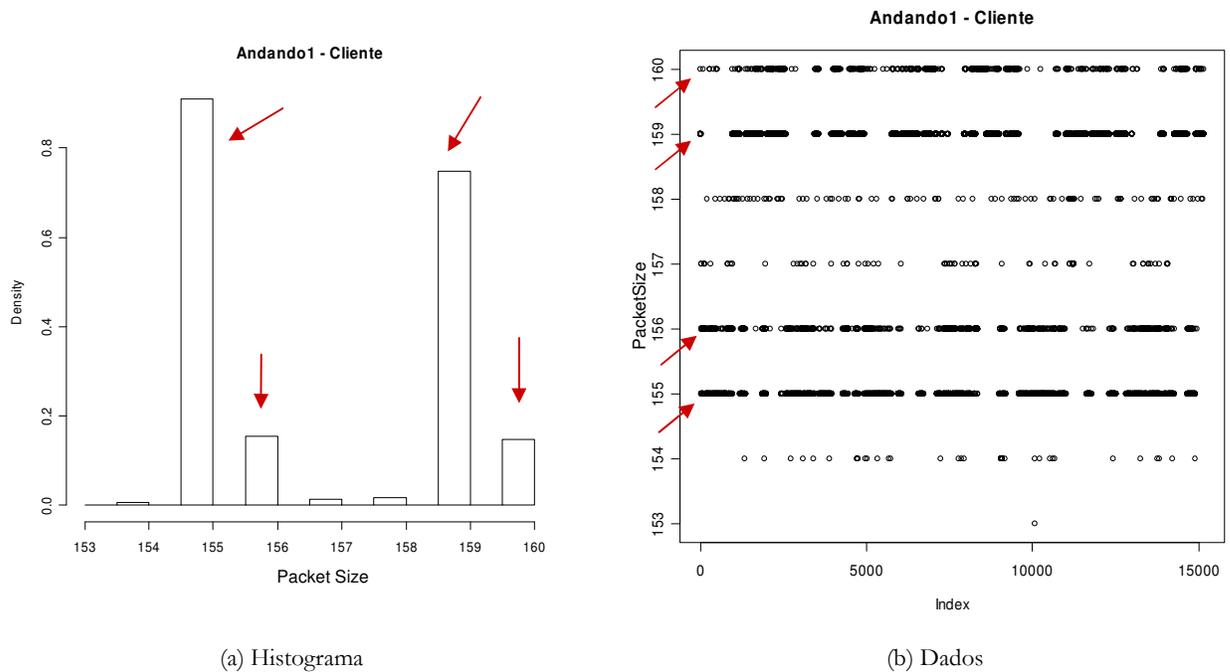


Figura 4.21 Histograma e dados do trace Andando-1 Cliente - métrica *Packet Size*

A análise numérica dos dados revela que: 45.5% dos pacotes deste trace possuem tamanho igual a 155 bytes; 7.7% possuem tamanho igual a 156 bytes; 37.4% possuem tamanho igual a 159 bytes e 7.3% possuem tamanho igual a 160 bytes. Esses 4 valores incluem 97.9% dos pacotes, o que justifica desconsiderar os outros tamanhos na modelagem.

4.3 Trace Andando 1 - Servidor

4.3.1 Métrica *Inter Arrival*

Como descrito na seção 4.2.1, esta métrica representa a diferença entre o tempo de chegada de dois pacotes consecutivos. Uma vez que os traces foram capturados no Cliente, não foi possível coletar dados para modelar esta métrica no Servidor.

4.3.2 Métrica *Packet Size*

A análise visual do histograma dos dados revela que não é possível presumir normalidade dos dados e que a melhor opção para modelagem parece ser dividir a distribuição em intervalos para permitir uma melhor captura das suas características. Optou-se por segmentar a distribuição em três (3) intervalos distintos. O primeiro apresenta valores de X menores a 500, o segundo com valores de X entre 500 e 1067 e o terceiro intervalo compreende valores acima de 1067. A Figura 4.22 apresenta o histograma dos dados e a divisão dos segmentos.

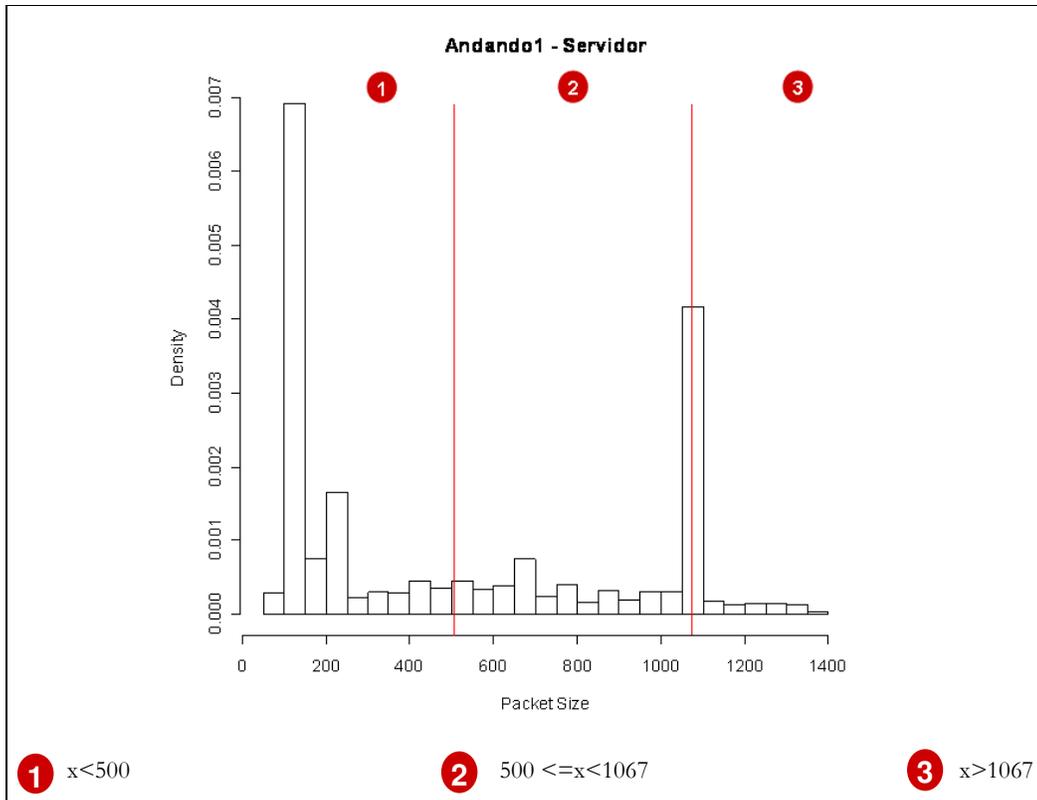


Figura 4.22 Histograma do trace Andando-1 Servidor - métrica *Packet Size*.

Intervalo $x < 500$

O histograma da primeira partição, composta pelos valores de X que são menores que 500, encontra-se ilustrado na Figura 4.23.

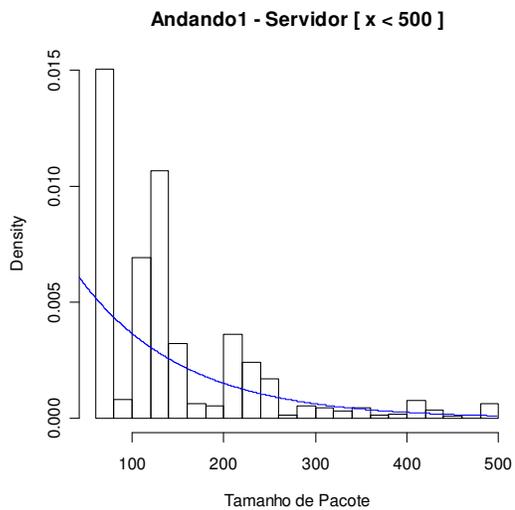


Figura 4.23 Histograma do trace Andando-1 Servidor, para $X < 500$ - métrica *Packet Size* e a função de densidade de probabilidade de uma Exponencial(0.009)

Através da análise visual do histograma acima, as distribuições Exponencial e Gama foram escolhidas como candidatas para modelar os dados. Todos os passos da modelagem foram realizados para cada uma dessas distribuições, contudo apenas a modelagem que apresentou a menor discrepância, a Exponencial, será apresentada nesta seção. Os gráficos referentes à distribuição Gamma estão no Apêndice D.

A estimação do parâmetro para a Exponencial resultou no valor 0.009. A Figura 4.23 plota a função de densidade de probabilidade da distribuição Exponencial com o parâmetro estimado sobre o histograma dos dados.

De posse dos parâmetros estimados, o gráfico dos quantis deve ser plotado para verificar a aderência das distribuição escolhidas aos dados empíricos. Percebe-se pelo Q-Q Plot, ilustrado na Figura 4.24(a), que a distribuição se encaixa perfeitamente em apenas poucos intervalos. Contudo, a métrica da discrepância, calculada no passo seguinte, afirma que o modelo pode ser considerado bom, visto que seu valor é 0.540. A comparação bin-a-bin entre as distribuições empírica e analítica, realizada no cálculo da discrepância, pode ser visualizada na Figura 4.23(b).

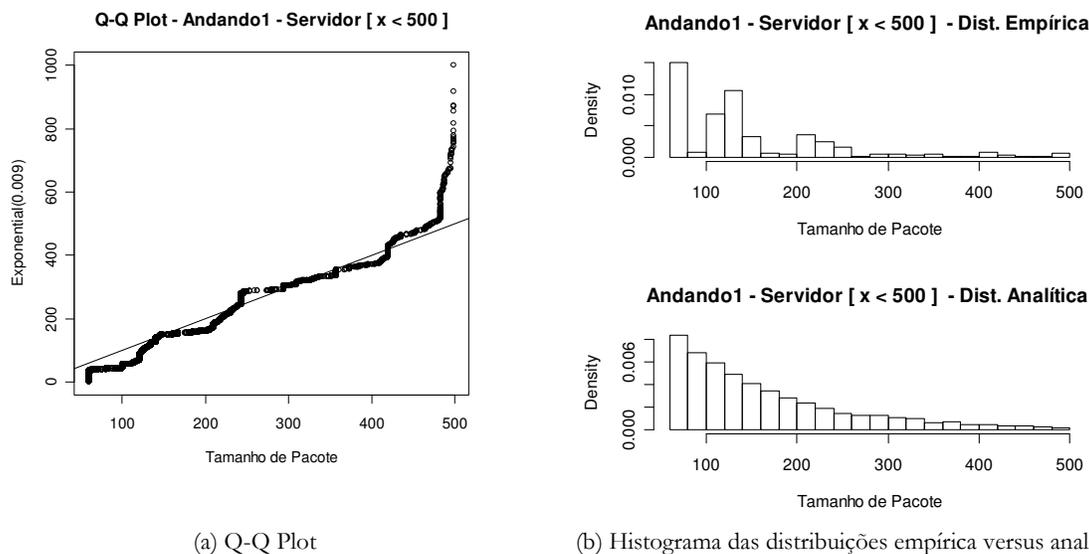


Figura 4.24 Q-Q Plot e Histograma do trace Andando-1 Servidor, para $X < 500$ - métrica *Packet Size* versus dados randomicamente gerados seguindo uma Exponencial(0.009)

Para concluir a sequência da metodologia, foi realizado o teste de aderência Kolmogorov-Smirnov. Como esperado, de acordo com as afirmações realizadas por [13] (ver Seção 2.3.5), o teste falhou.

Intervalo $500 \leq x < 1067$

O histograma da segunda partição, composta pelos valores de X que se encontram entre 500 e 1067, encontra-se ilustrado na Figura 4.25.

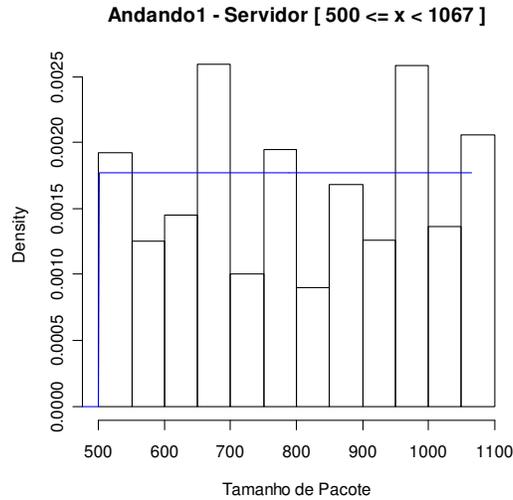


Figura 4.25 Histograma do trace Andando-1 Servidor, para $500 < X \leq 1067$ - métrica *Packet Size* e a função de densidade de probabilidade de uma Uniforme(500, 1067)

Através da análise visual do histograma acima, percebeu-se que a melhor opção de distribuição para modelar esse intervalo seria uma Uniforme. E, dado que os resultados para essa distribuição foram satisfatórios, não foram testadas outras distribuições.

A Figura 4.25 plota a função de densidade de probabilidade da distribuição Uniforme(500, 1067) sobre o histograma dos dados.

O gráfico Q-Q Plot, ilustrado na Figura 4.26(a), indica que a distribuição se adequa em aproximadamente metade dos intervalos e a métrica da discrepância com valor de 0.238 aponta que o modelo pode ser considerado muito bom. A comparação bin-a-bin entre as distribuições empírica e analítica, realizada no cálculo da discrepância, pode ser visualizada na Figura 4.26(b).

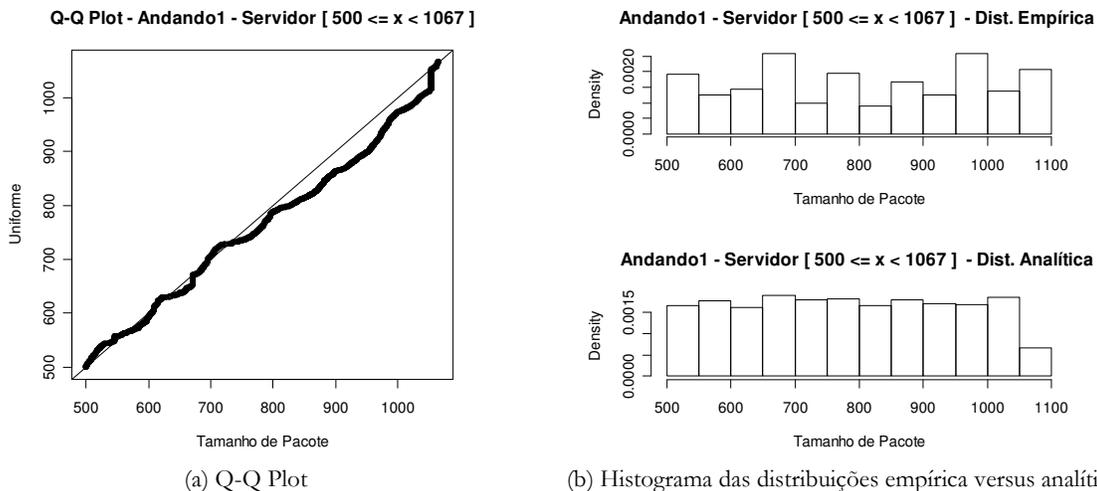


Figura 4.26 Q-Q Plot e Histograma do trace Andando-1 Servidor, para $500 \leq X < 1067$ - métrica *Packet Size* versus dados aleatoriamente gerados seguindo uma Uniforme(500,1100)

Intervalo $x \geq 1067$

O histograma da última partição deste trace, isto é, do intervalo em que os valores de X são maiores que 1067, encontra-se ilustrado na Figura 4.27. Através da análise visual do histograma, as distribuições Gamma e Extreme foram escolhidas como candidatas para modelar os dados. Todos os passos da modelagem foram realizados para cada uma dessas distribuições, porém apenas a modelagem que apresentou a menor discrepância, a Extreme, será apresentada nesta seção. Os gráficos referentes à distribuição Gama podem ser visualizados no Apêndice D.

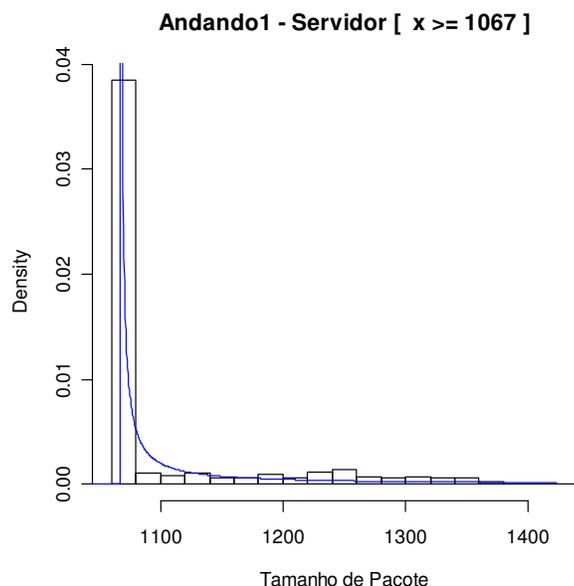


Figura 4.27 Histograma do trace Andando-1 Servidor, para $X > 1067$ - métrica *PacketSize* e a função de densidade de probabilidade de uma Extreme(1075.734, 47.598, 5.449)

A estimação do parâmetro para a Extreme, realizada no segundo passo da metodologia, resultou nos valores 1075.734, 47.598 e 5.449, para *shape*, *scale* e *location*, respectivamente. A função de densidade de probabilidade da distribuição Extreme com os parâmetros estimado encontra-se plotada sobre o histograma dos dados na Figura 4.27.

O gráfico dos quantis foi plotado no próximo passo. A Figura 4.28(a) ilustra o Q-Q Plot da distribuição dos dados do trace no intervalo estudado versus a Extreme(1075.734, 47.598, 5.449). O gráfico Q-Q Plot ficou pouco legível, provavelmente devido ao último ponto que se encontra na escala de 10^{19} , enquanto os demais parecem estar próximos de 0. Contudo, a métrica da discrepância, calculada no passo seguinte, afirma que o modelo pode ser considerado excelente, visto que seu valor é 0.186. A comparação bin-a-bin entre as distribuições empírica e analítica, realizada no cálculo da discrepância, pode ser visualizada na Figura 4.28(b).

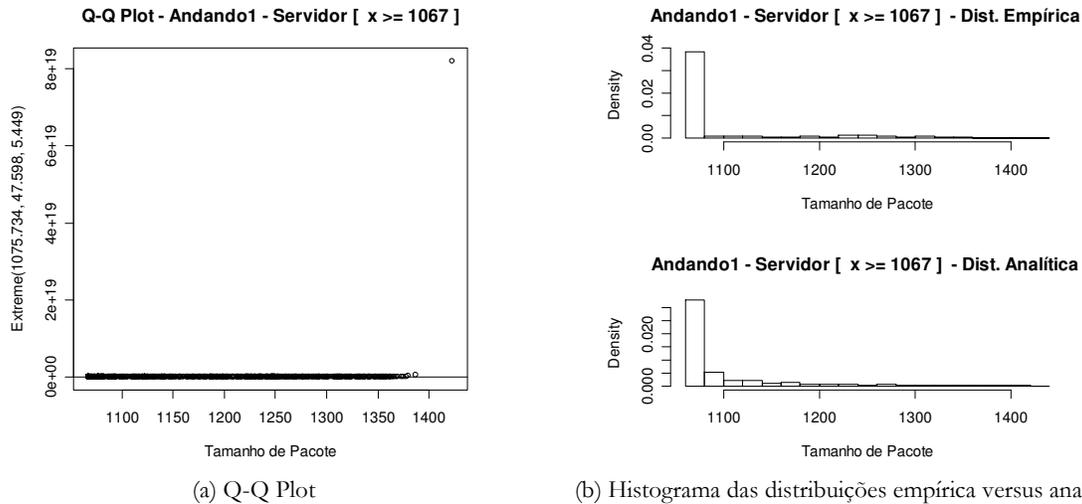


Figura 4.28 Q-Q Plot e Q-Q Plot e Histograma do trace Andando-1 Servidor, para $X \geq 1067$ - métrica *Packet Size* versus dados aleatoriamente gerados seguindo uma distribuição Extreme(1075.734, 47.598,5.449)

No último passo da metodologia, foi realizado o teste de aderência Kolmogorov-Smirnov. Como esperado, de acordo com as afirmações realizadas por [13] (ver Seção 2.3.5), o teste falhou.

4.4 Trace Parado-1 Servidor

4.4.1 Métrica *Inter Arrival*

Assim como para o trace Andando-1 Servidor, esta métrica não foi modelada, pois representa a diferença entre o tempo de chegada de dois pacotes consecutivos. Uma vez que os traces foram capturados no Cliente, não foi possível coletar dados para modelar esta métrica no Servidor.

4.4.2 Métrica *Packet Size*

Através de análise visual do histograma deste trace, observou-se que há uma grande semelhança com o histograma do trace Andando-1 Servidor. Assim, decidiu-se testar a modelagem realizada para Andando-1 Servidor nos traces 2, 3, 4 e 5 de Parado Servidor. A avaliação gerou excelentes resultados, que serão apresentados na seção 5.2, e, portanto, não foi gerado um modelo a partir deste trace.

5 Avaliação do Modelo

Esta seção visa avaliar a modelagem proposta na seção 4. Como descrito na seção 2.3.1, o trace 1 foi utilizado para modelar o tráfego e os quatro traces restantes foram utilizados para testar e validar a modelagem. As subseções seguintes apresentam os resultados obtidos na avaliação.

5.1 Traces Parado 2 3 4 5 - Cliente

5.1.1 Métrica *Inter Arrival*

Similar ao descrito na seção 4.1, estes traces representam o comportamento capturado quando um

usuário encontra-se parado no mundo virtual e reflete o tráfego que é enviado do cliente para o servidor (tráfego de saída) e, da mesma forma que o trace Parado 1, também foram divididos em quatro partições ($x < 0.001$; $0.001 \leq x < 0.02$; $0.02 \leq x < 0.8$; $x \geq 0.8$).

A Tabela 5.1 apresenta o resultado do teste de discrepância para todos os traces em cada partição. Observando os valores na tabela, pode-se notar que as partições II e III tiveram um bom resultado, com valores próximos a 1. Enquanto as partições I e IV tiveram um desempenho ruim, com todos os valores acima de 2 e chegando a atingir 14.139 para o trace Parado-2 na partição IV. Também pode-se perceber que o trace Parado-4 é atípico, apresentando valores altos para discrepância em todas as partições.

Tabela 5.1 Discrepância – Avaliação para modelagem do trace Parado-1 Cliente – métrica *Inter Arrival*

Trace	Partição			
	I ($X < 0.001$) Exponencial(3617.217, 0.174)	II ($0.001 \leq X < 0.02$) Weibull(1.505, 0.005)	III ($0.02 \leq X < 0.08$) Weibull(9.368, 0.057)	IV ($X \geq 0.8$) Gama(21.060, 0.005)
Parado 2 Cliente	2.519	0.492	1.400	14.139
Parado 3 Cliente	2.419	0.719	0.259	2.200
Parado 4 Cliente	4.819	1.475	3.049	11.030
Parado 5 Cliente	2.316	0.398	1.148	7.881

Os resultados da avaliação da modelagem para a situação PARADO no sentido Cliente mostram que a modelagem não foi muito bem sucedida para as partições I e IV. O motivo da alta discrepância para tais partições pode ser devida ao particionamento errado da distribuição ou a ter assumido de forma equivocada que todos os traces seguissem as mesmas distribuições. É necessário fazer uma análise mais cuidadosa para remodelar essas partições.

5.2 Traces Andando 2 3 4 5 - Cliente

5.2.1 Métrica *Inter Arrival*

Similar ao descrito na seção 4.2, estes traces representam o comportamento capturado quando um usuário encontra-se andando no mundo virtual e reflete o tráfego que é enviado do cliente para o servidor (tráfego de saída) e, da mesma forma que o trace Parado 1, também foram divididos em três partições ($x < 0.001$; $0.001 \leq x < 0.015$; $x \geq 0.015$).

A Tabela 5.2 apresenta o resultado do teste de discrepância para todos os traces em cada partição. Observando os valores na tabela, pode-se notar que a modelagem da partição não foi satisfatória, pois obteve valores extremamente altos para discrepância. Em contrapartida, a segunda partição obteve ótimos resultados, com todos os valores abaixo de 1. A partição III teve resultado razoável, com valores próximos de 3, sendo recomendável uma nova modelagem.

Tabela 5.2 Discrepância – Avaliação para modelagem do trace Andando-1 Cliente – métrica *Inter Arrival*

Trace	Partição		
	I ($X < 0.001$) Exponencial(3502.129, 0.242)	II ($0.001 \leq X < 0.015$) Exponencial(205.841, 0.376)	III ($X > 0.015$) Exponencial(24.505, 0.376)
Parado 2 Cliente	21.466	0.419	3.543
Parado 3 Cliente	23.831	0.889	3.691
Parado 4 Cliente	18.825	0.488	2.854
Parado 5 Cliente	28.086	0.544	3.289

Os resultados da avaliação da modelagem para a situação ANDADO no sentido Cliente mostram que a modelagem foi excelente para a partição II, mas as outras partições devem ser remodeladas. e IV.

5.3 Traces Andando 2 3 4 5 - Servidor

5.3.1 Métrica *Packet Size*

Similar ao descrito na seção 4.3, estes traces representam o comportamento capturado quando um usuário encontra-se em movimento no mundo virtual e reflete o tráfego que é enviado do servidor para o cliente (tráfego de entrada) e, da mesma forma que o trace Andando 1, também foram divididos em três partições ($X < 500$; $500 \leq X < 1067$; $X \geq 1067$).

A Tabela 5.3 apresenta o resultado do teste de discrepância para todos os traces em cada partição. Observando os valores na tabela, pode-se notar que a modelagem de todas as partições foi muito boa, obtendo discrepância com valores menores que 1 para quase todos os casos.

Tabela 5.3 Discrepância – Avaliação para modelagem do trace Andando-1 Servidor – métrica *Packet Size*

Trace	Partição		
	I ($X < 500$) Exponencial(0.009)	II ($500 \leq X < 1067$) Uniforme(500,1067)	III ($X \geq 1067$) Exponencial(0.009)
Andando 2 Servidor	1.025	0.168	0.533
Andando 3 Servidor	0.975	0.186	0.636
Andando 4 Servidor	0.435	0.301	0.577
Andando 5 Servidor	0.928	0.219	1.063

Os resultados da avaliação da modelagem para a situação ANDANDO no sentido Servidor mostram que a modelagem foi bastante satisfatória e pode ser considerada uma boa modelagem.

5.4 Traces Parado 2 3 4 5 - Servidor

5.4.1 Métrica *Packet Size*

Similar ao descrito na seção 4.4, estes traces representam o comportamento capturado quando um usuário encontra-se em movimento no mundo virtual e reflete o tráfego que é enviado do servidor para o cliente (tráfego de entrada) e, da mesma forma que o trace Parado 1, também foram divididos em três partições ($X < 500$; $500 \leq X < 1067$; $X \geq 1067$).

A Tabela 5.4 apresenta o resultado do teste de discrepância para todos os traces em cada partição. Os valores na tabela mostram que essa foi a melhor modelagem de todas, pois os valores obtidos para a discrepância foram excelentes em todos os casos, sendo o maior valor obtido igual a 0.640.

Tabela 5.4 Discrepância – Avaliação para modelagem do trace Parado-1 Servidor – métrica *Packet Size*

Trace	Partição		
	I ($X < 500$) Exponencial(0.009)	II ($500 \leq X < 1067$) Uniforme(500,1067)	III ($X \geq 1067$) Exponencial(0.009)
Parado 2 Servidor	0.617	0.136	0.192
Parado 3 Servidor	0.627	0.109	0.147
Parado 4 Servidor	0.640	0.163	0.529
Parado 5 Servidor	0.614	0.154	0.197

Os resultados da avaliação da modelagem para a situação PARADO no sentido Servidor mostram que a modelagem foi bastante satisfatória e pode ser considerada uma ótima modelagem.

6 Conclusão

Neste estudo foi proposto um modelo para as distribuições de tamanho de pacote e *inter arrival* para tráfego de entrada e saída do Second Life. Também foi apresentada uma avaliação do modelo através do cálculo da discrepância entre as distribuições geradas pelo modelo e algumas distribuições obtidas empiricamente dos traces.

Para a tarefa de modelagem, foi utilizada a técnica de “*split distribution*” para a obtenção de um modelo mais fiel aos dados empíricos. Foram conduzidos testes com várias distribuições (Gamma, Exponential, Normal, Extreme, Uniforme e Weibull), sempre procurando o menor valor possível para a discrepância.

A avaliação mostrou que os modelos conseguiram se adequar à maioria dos traces coletados (mesmo os que não foram modelados), o que indica que os modelos conseguem descrever com fidelidade o comportamento da aplicação em várias situações diferentes. Contudo, alguns modelos não obtiveram resultado satisfatório, sendo necessária a remodelagem para esses casos.

Como trabalhos futuros, destacam-se:

- A remodelagem das partições que obtiveram resultados insatisfatórios;
- O estudo mais detalhado da distribuição Extreme, que apresentou bons resultados em trechos da modelagem;
- A procura por novas ou não usuais distribuições, como por exemplo shifted Gama, que assim como a distribuição Extreme usa o parâmetro de localização;
- A implementação do modelo em um simulador de rede amplamente utilizado, como o NS2[27];
- A geração de tráfego em um simulador de rede e avaliação da aderência dos dados simulados com os empíricos.

7 Referências

- [1] Book, B., “What is a Virtual World”, Virtual World Review, 2006, <http://www.virtualworldreview.com>
- [2] Castronova, Edward. “The Right to Play”, 2003, <http://www.nyls.edu/pdfs/castronova.pdf>
- [3] Electronic Arts, “The Sims Online”, 2007, <http://www.ea.com/official/thesims/thesimsonline/us/nai/index.jsp>
- [4] Rymaszewski, M. et al., “Second Life: The Official Guide”, Sybex, 2006
- [5] Second Life, Inc., “Second Life”, 2007, <http://www.secondlife.com>
- [6] Marathe, M., and Hawe, W. “Predicted Capacity of Ethernet in a University Environment,” Proceedings of Southcon 1982, March 1982, pp 1-10.

- [7] Jain, R., Routhier, S.A. "Packet Trains - Measurements and a New Model for Computer Network Traffic," IEEE Journal on Selected Areas in Communications, Vol. 4, No. 6, September 1986, pp. 986-995.
- [8] H. Heffes and D. M. Lucantoni "A Markov Modulated Characterization of Packetized Voice and Data Traffic and Related Statistical Multiplexer Performance," IEEE Journal on Selected Areas in Communications, Vol. 4, No. 6, September 1986, pp. 856-868.
- [9] Willinger, W. "The Discovery of Self-Similar Traffic," In Performance Evaluation: Origins and Directions G. Haring, C. Lindemann, and M. Reiser, Eds. Lecture Notes In Computer Science, vol. 1769. Springer-Verlag, London, 513-527.
- [10] Ashok Erramilli, Matthew Roughan, Darryl Veitch, Walter Willinger. 2002. "Self-Similar Traffic and Network Dynamics," In Proc. of the IEEE., Vol 90, no. 5.
- [11] Abry, P., and Veitch, D. "Wavelet analysis of long-range-dependent traffic," IEEE Transactions on Information Theory 44, 1 (1998), 215.
- [12] "Total MMOG Active Subscriptions", <http://www.mmogchart.com/Chart4.html>
- [13] M. S. Borella, "Source Models of Network Game Traffic".Computer Communications 23(4), pp. 403-410, Feb. 2000.
- [14] Lakkakorpi J., Heiner A., Ruutu J. Measurement and Characterization of Internet Gaming Traffic. Research Seminar on Networking, Helsinki University of Technology, Networking Laboratory, February, 2002.
- [15] T. Lang, P. Branch, G. Armitage. A Synthetic Traffic Model for Quake 3. ACM SIGCHI ACE2004 conference, Singapore, June 2004.
- [16] J. Faerber. Network game traffic modelling. ACM NetGames Workshop 2002, Germany, April 2002.
- [17] W. Feng, F. Chang, W. Feng, J. Walpole. Provisioning Online Games: A Traffic Analysis of a Busy Counter-Strike Server. SIGCOMM Internet Measurement Workshop, November 2002.
- [18] T. Lang, G. Armitage, P. Branch, H-Y. Choo. A Synthetic Traffic Model for Half Life. Australian Telecommunications Networks & Applications Conference (ATNAC) 2003, Melbourne, Australia, December 2003.
- [19] T. Lang, G. Armitage, "A ns-2 model for the System Link game Halo", Australian Telecommunications Networks & Applications Conference (ATNAC) 2003, Melbourne, Australia, December 2003.
- [20] S. Zander, G. Armitage. A Traffic Model for the Xbox Game Halo 2. 15th ACM International Workshop on Network and Operating System Support for Digital Audio and Video (NOSSDAV 2005), Washington (USA), June 2005.

- [21] “Second Life – What is Second Life”, <http://secondlife.com/whatis/>
- [22] “Second Life Research – Second Life Resident Statistics”, <http://secondliferesearch.blogspot.com/2007/03/second-life-residents-statistics.html>
- [23] “The R Project for Statistical Computing”, <http://www.r-project.org/>
- [24] L.J. Glesser, D.S. Moore, "The effect of dependence on chi-squared and. empiric distribution tests of fit". The Annals of Statistics, 11:1100–1108, 1983.
- [25] Carl E. Minton, "Modeling and Estimation Techniques for Wide-Area Network Traffic with Atypical Components", Master's Dissertation, Virginia Polytechnic Institute and State University, April 2002.
- [26] Branch, P., Armitage, G. Extrapolating server to client IP traffic from empirical measurements of first person shooter games. In Proceedings of 5th ACM SIGCOMM Workshop on Network and System Support For Games, October, 2006.
- [27] “The Network Simulator - ns2”, <http://www.isi.edu/nsnam/ns/>

Apêndice A

Scripts do R - Discrepância

```
# Estimates lambda2 discrepancy
# [x2, lambda2, dev] = lambdasquare(empirical, analytical, param, bins, min, max, xlab, mainlab)
# param = parameters of the analytical distribution
# bins = custom bins vector or bin width
# min,max:
# Use 0,0 to not make any truncations
# Use 1,1 to truncate analytical to match min and max from empirical
# Use different values to truncate both distros at your pleasure.
```

```

# xlab = xlab of the histograms
# mainlab = mainlab of the histograms

lambdasquare <- function(empirical, analytical, param, bins, mmin, mmax, xlab, mainlab) {
  if (length(bins) == 1) {
    w <- bins
  }

  # Truncations
  if (mmin == mmax) {
    if (mmin != 0) {
      # Truncate analytical to match min and max from empirical
      analytical <- analytical[(analytical >= min(empirical)) & (analytical <= max(empirical))]
      if (length(bins) == 1) {
        bins <- seq(from=(min(empirical)+w/2), to=max(empirical), by=w)
      }
    } else {
      if (length(bins) == 1) {
        bins <- seq(from=(min(min(empirical),min(analytical))+w/2),
to=max(max(empirical),max(analytical)), by=w)
      }
    }
  } else {
    # Truncate both to mmin and mmax
    analytical <- analytical[(analytical >= mmin) & (analytical <= mmax)];
    empirical <- empirical[(empirical >= mmin) & (empirical <= mmax)];
    if (length(bins) == 1) {
      bins <- seq(from=(min(min(empirical),min(analytical))+w/2), to=max(max(empirical),max(analytical)),
by=w);
    }
  }

  # Just for ease
  n <- length(empirical);

  par(mfrow=c(2,1),bg="white")

  # Compute histograms
  # y = absolute frequency of the empirical distro
  y <- hist(empirical, nclass=length(bins), xlab=xlab, main=paste(mainlab," - Dist. Empírica",seq=""), freq=FALSE)
  y <- y$counts
  # z = absolute frequency of the analytical distro
  z <- hist(analytical, nclass=length(bins), xlab=xlab, main=paste(mainlab," - Dist. Analítica",seq=""), freq=FALSE)
  z <- z$counts

  # Check if analytical has zero frequency bins
  if (min(abs(z), na.rm=TRUE) == 0) {
    bins <- bins[z != 0]
    y <- y[z != 0]
    z <- z[z != 0]
  }

  # Check if empirical has zero frequency bins
  if (min(abs(y), na.rm=TRUE) == 0) {
    print("Warning: There are bins with zero frequency in empirical.");
    print("The computation of the formula is correct, but the chosen bin width probably not.");
  }

  # Compute chi-square discrepancy
  # Expected freq = number of samples of empirical * relative frequency of i_th bin of analytical
  npi <- n * (z/length(analytical));
  x2 <- sum(((y-npi)^2) / npi);

```

```

# Compute lambda-square discrepancy
D <- y - npi;
E <- npi;
k <- sum(D / E);
df <- length(bins) - 1 - length(param);
lambda2 <- (x2 - k - df)/(n-1);

# Compute variance associated with this estimate of lambda2
T <- sum( (D^3 - 2*(D*E) + 5/2*(D^2) + 3/2*(D+E)) / (E^2) );
dev <- sqrt(2*df + 4*n*lambda2 + 4*n*lambda2^2 + 4*T) / n;

list(x2=x2, lambda2=lambda2, dev=dev)

}

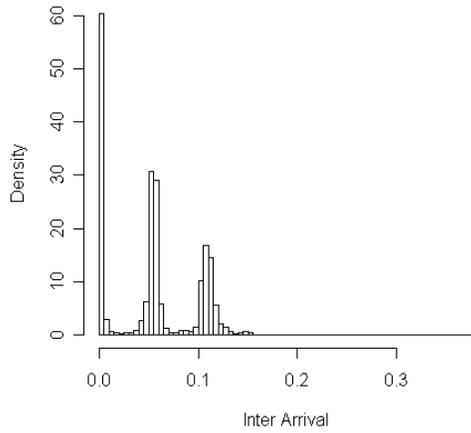
##testando a funcao
#x <- andando$entrada$pktSize
#fitdistr(x,"normal")
#lambdasquare(x,rnorm(length(x),mean=66.63,sd=11.38),0,10,0,0)

```

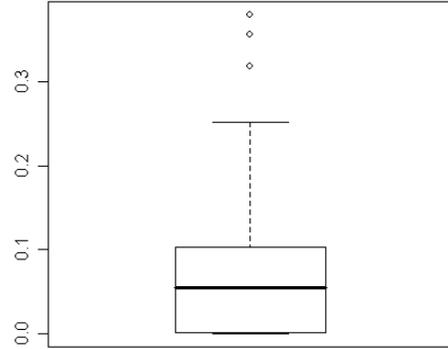
Apêndice B

Análise Exploratória – Traces Parado 1 (Cliente e Servidor)

Parado 1 - Cliente

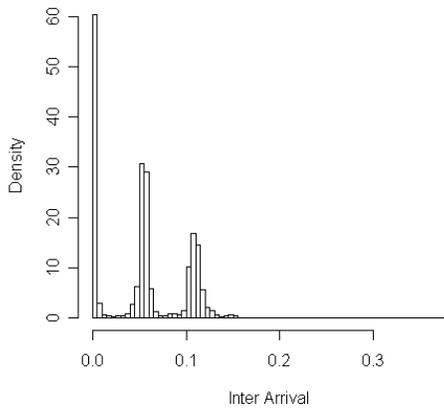


(a) Histograma original



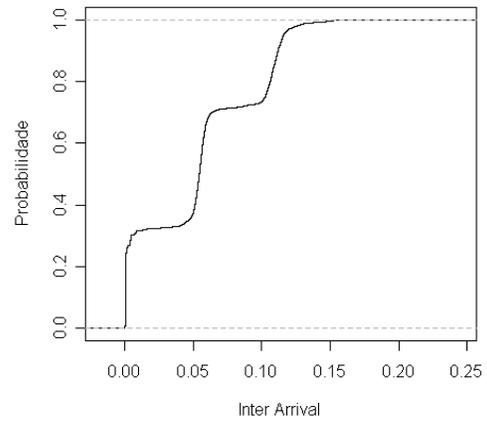
(b) Box-Plot

Parado 1 - Cliente



(c) Histograma sem *outliers*

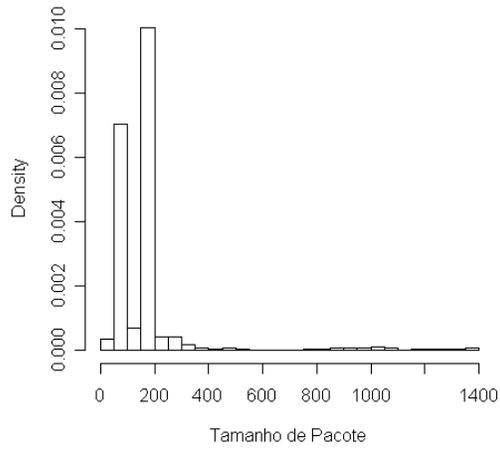
ECDF



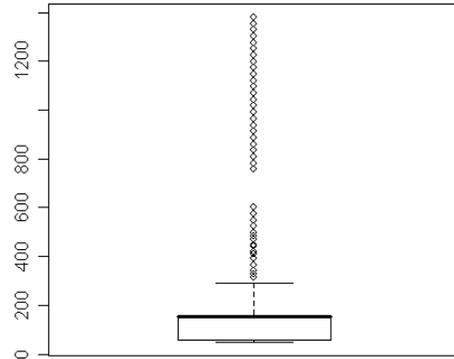
(d) ECDF

Figura B.1 Histograma, Box-Plot, Histograma sem outliers e ECDF do trace Parado-1 Cliente - métrica *Inter Arrival*

Parado 1 - Cliente

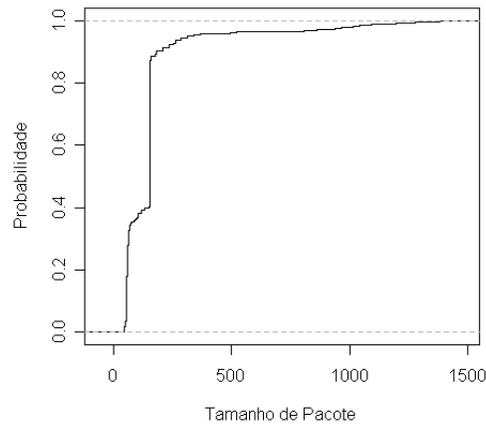


(a) Histograma do tamanho de pacotes



(b) Box-Plot tamanho de pacotes

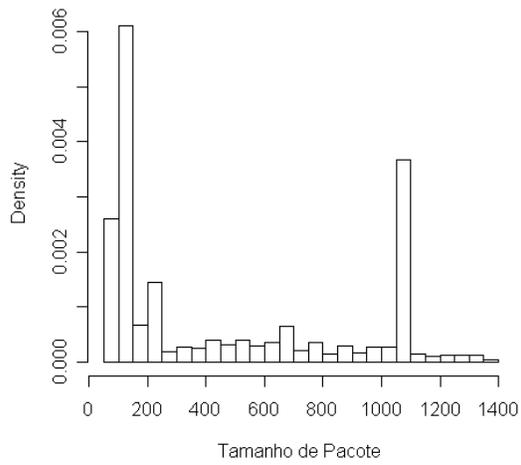
ECDF



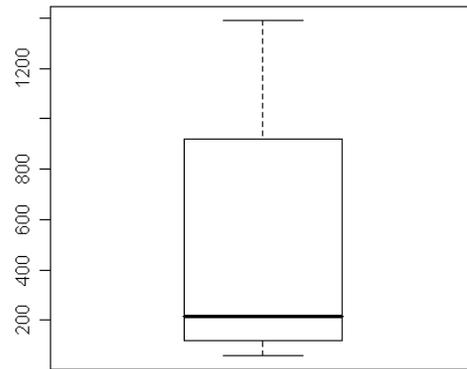
(c) ECDF Tamanho de pacote

Figura B.2 Histograma, Box-Plot, e ECDF do trace Parado-1 Cliente - métrica *Packet Size*

Parado 1 - Servidor

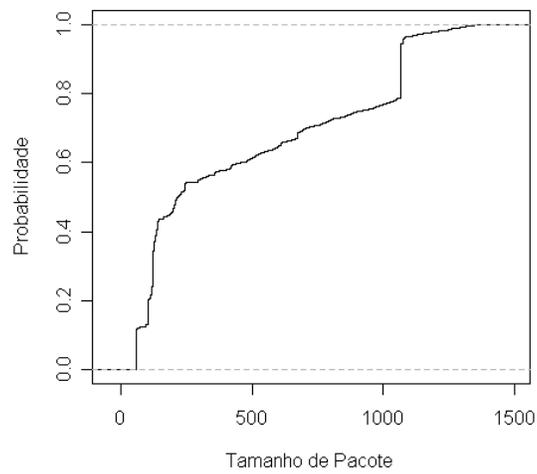


(a) Histograma do tamanho de pacotes



(b) Box-Plot tamanho de pacotes

ECDF



(c) ECDF Tamanho de pacote

Figura B.3 Histograma, Box-Plot, e ECDF do trace Parado-1 Servidor - métrica *Packet Size*

Apêndice C

Modelagem do Trace Andando 1 – Cliente - Métrica *Inter Arrival*

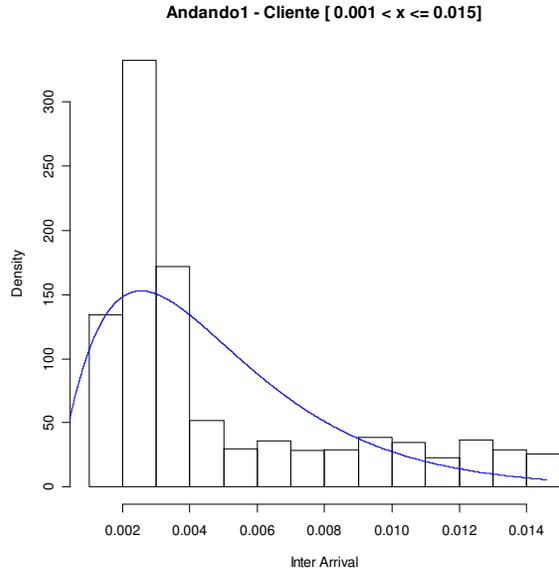


Figura C.1 Histograma do trace Andando-1 Cliente, para $0.001 < X \leq 0.015$ - métrica *Inter Arrival* e a função de densidade de probabilidade de uma $\text{Gamma}(2.097, 431.43)$

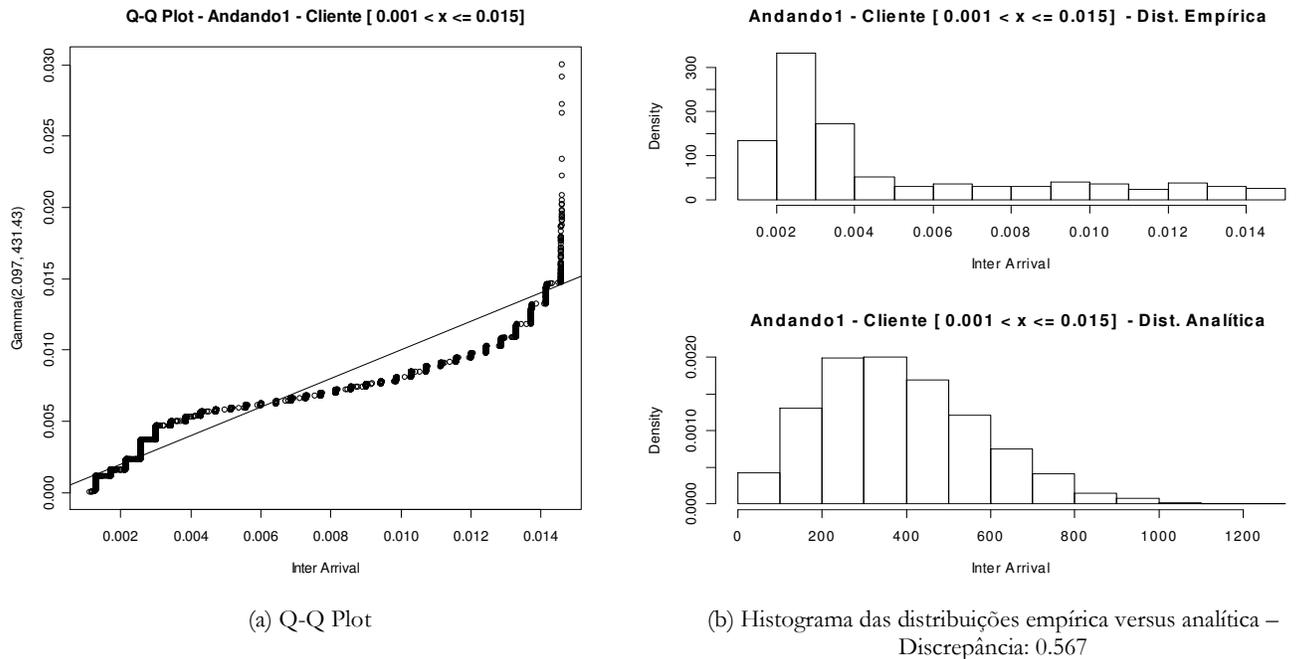


Figura C.2 Q-Q Plot e Histograma do trace Andando-1 Cliente, para $0.001 < X \leq 0.015$ - métrica *Inter Arrival* – $\text{Gamma}(2.097, 431.43)$

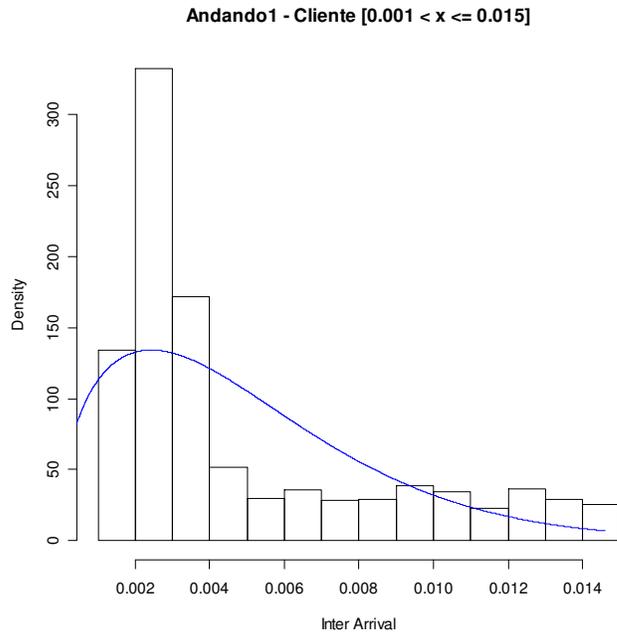
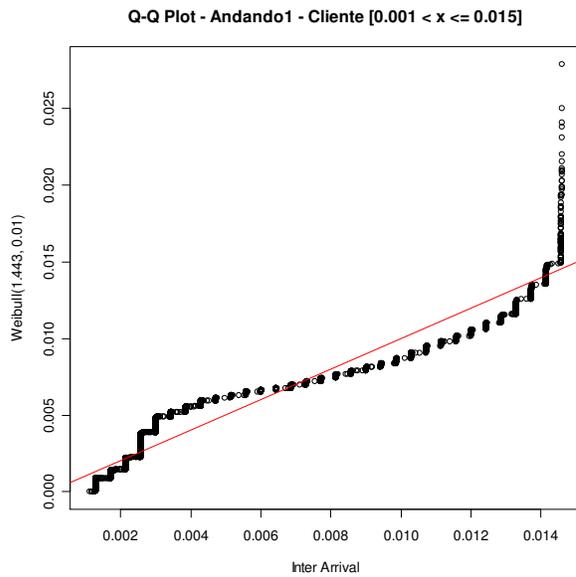
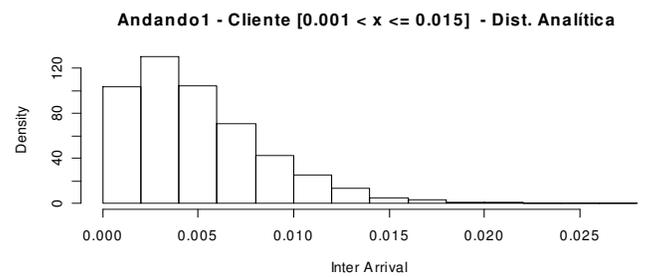
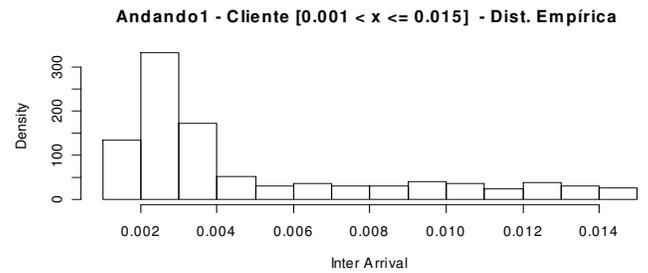


Figura C.3 Histograma do trace Andando-1 Cliente, para $0.001 < X \leq 0.015$ - métrica *Inter Arrival* e a função de densidade de probabilidade de uma Weibull(1.443,0.005)



(a) Q-Q Plot



(b) Histograma das distribuições empírica versus analítica – Discrepância: 7.109

Figura C.4 Q-Q Plot e Histograma do trace Andando-1 Cliente, para $0.001 < X \leq 0.015$ - métrica *Inter Arrival* – Weibull(1.443,0.005)

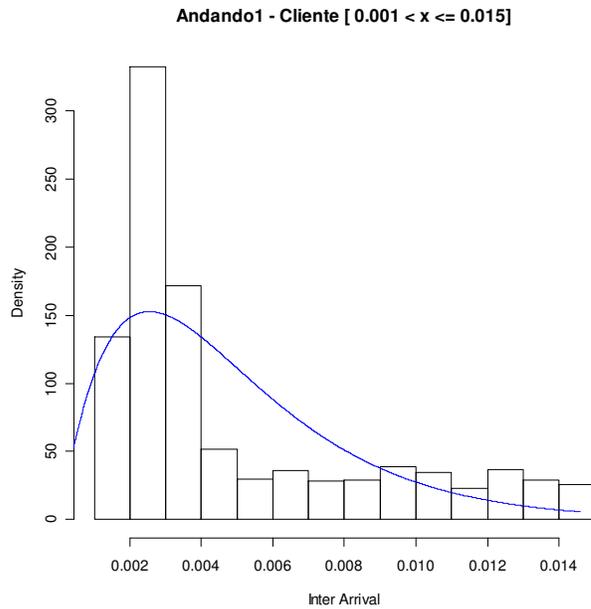
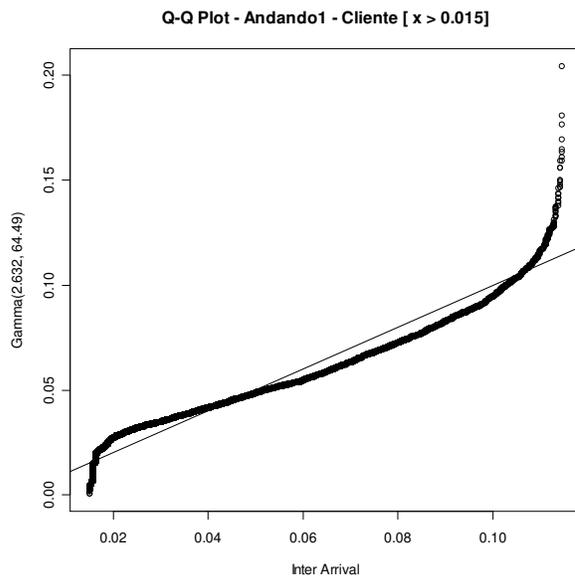
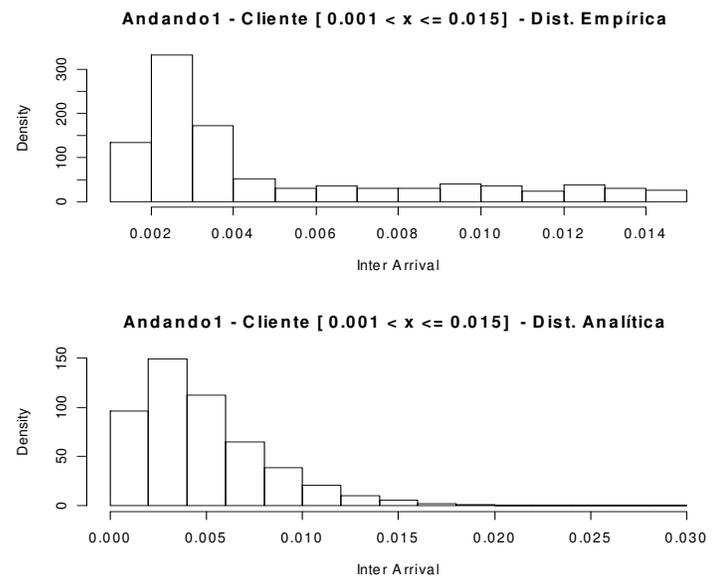


Figura C.5 Histograma do trace Andando-1 Cliente, para $X > 0.015$ - métrica *Inter Arrival* e a função de densidade de probabilidade de uma $\text{Gamma}(2.631, 64.490)$



(a) Q-Q Plot



(b) Histograma das distribuições empírica versus analítica – Discrepância: 0.699

Figura C.6 Q-Q Plot e Histograma do trace Andando-1 Cliente, para $X > 0.015$ - métrica *Inter Arrival* – $\text{Gamma}(2.631, 64.490)$

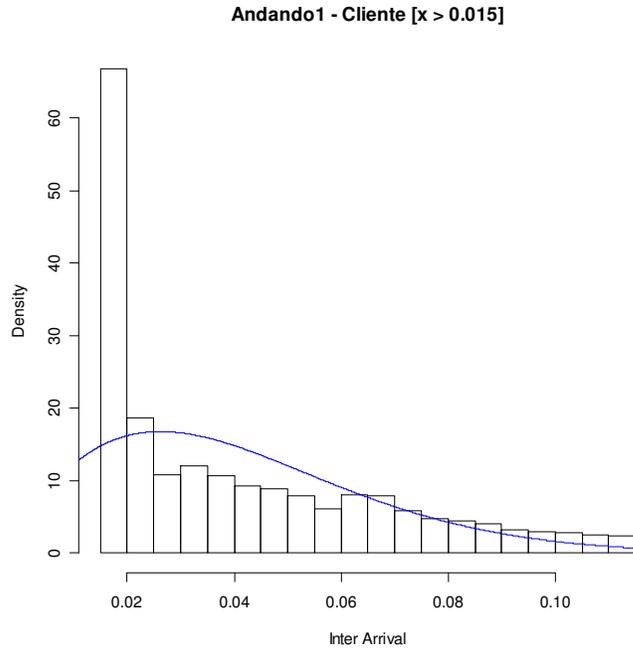
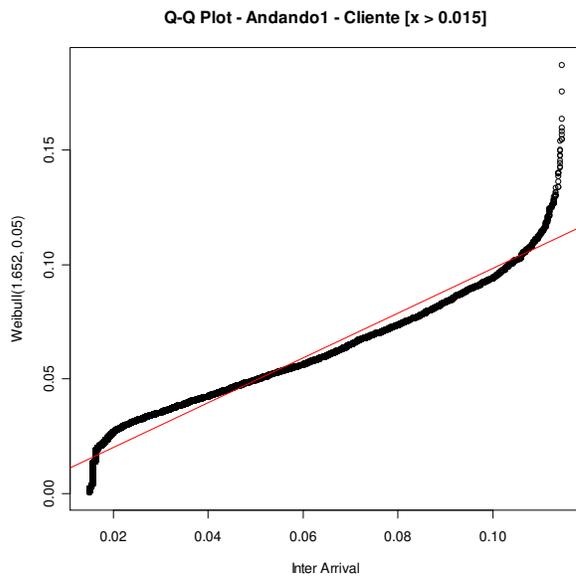
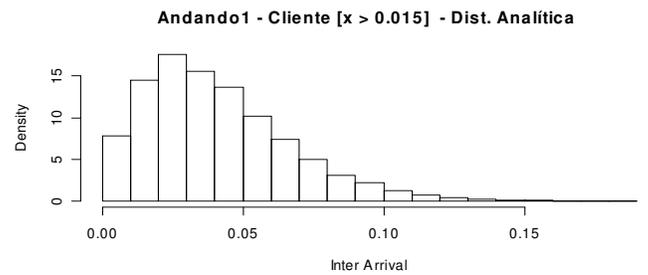
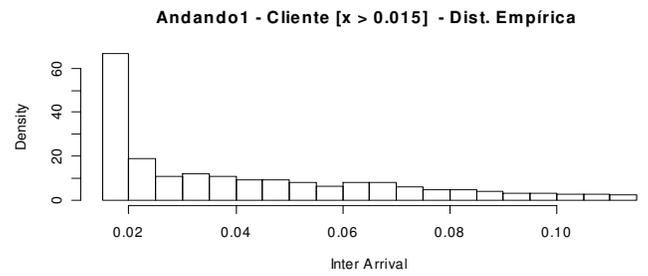


Figura C.7 Histograma do trace Andando-1 Cliente, para $X > 0.015$ - métrica *Inter Arrival* e a função de densidade de probabilidade de uma Weibull(1.652,0.046)



(a) Q-Q Plot



(b) Histograma das distribuições empírica versus analítica – Discrepância: 7.407

Figura C.8 Q-Q Plot e Histograma do trace Andando-1 Cliente, para $X > 0.015$ - métrica *Inter Arrival* – Weibull(1.652,0.046)

Apêndice D

Modelagem do Trace Andando 1 – Servidor - Métrica *Packet Size*

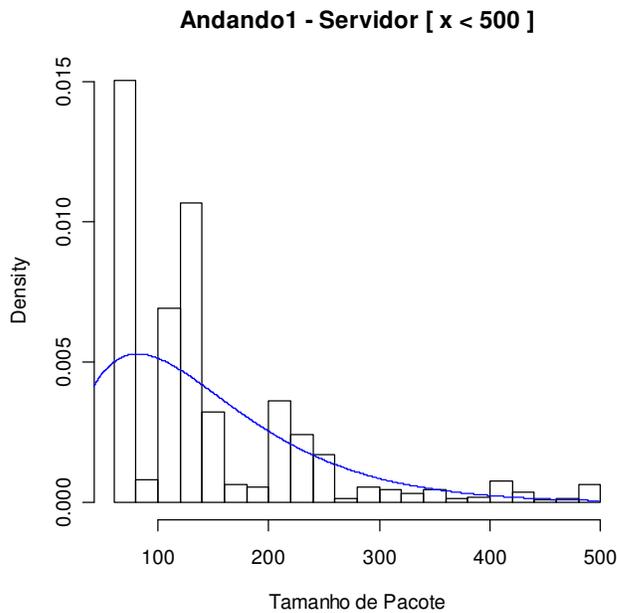
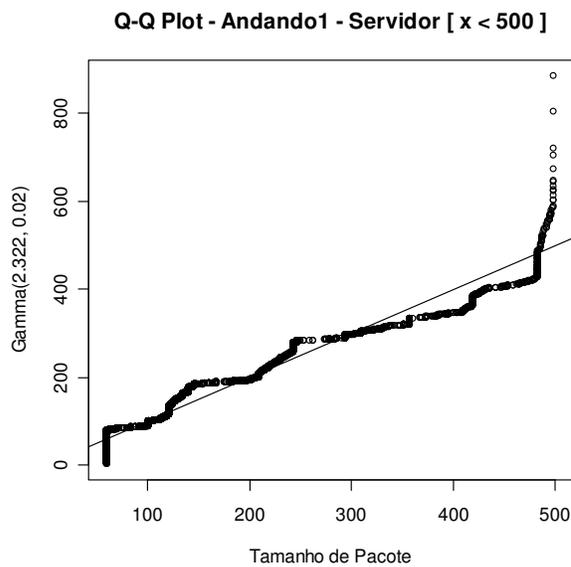
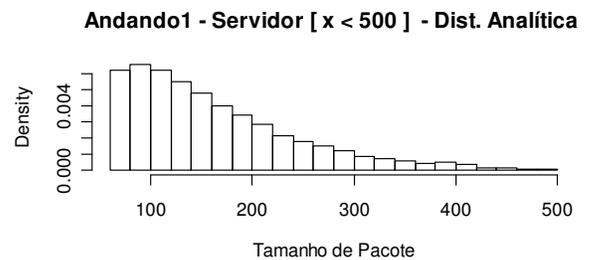
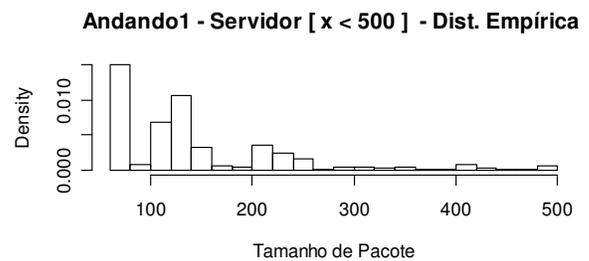


Figura D.1 Histograma do trace Andando-1 Servidor, para $X < 500$ - métrica *Packet Size* e a função de densidade de probabilidade de uma $\text{Gamma}(2.322, 0.02)$



(a) Q-Q Plot



(b) Histograma das distribuições empírica versus analítica – Discrepância: 0.723

Figura D.2 Q-Q Plot e Histograma do trace Andando-1 para $X < 500$ - métrica *Packet Size* - $\text{Gamma}(2.322, 0.02)$

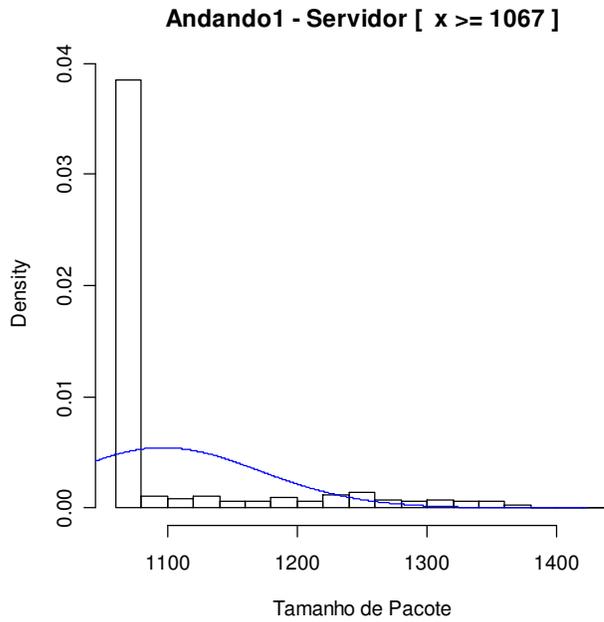
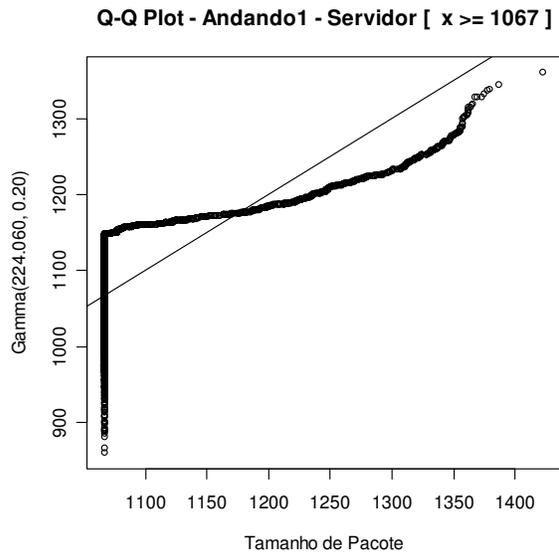
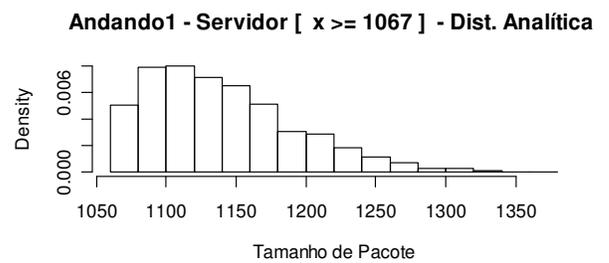
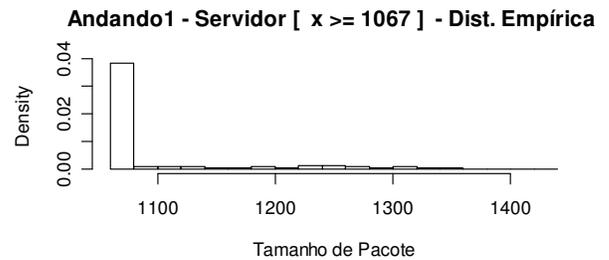


Figura D.3 Histograma do trace Andando-1 Servidor, para $X \geq 1067$ - métrica *Packet Size* e a função de densidade de probabilidade de uma Gamma(22.406, 0.02)



(a) Q-Q Plot



(b) Histograma das distribuições empírica versus analítica – Discrepância: 5.898

Figura D.4 Q-Q Plot e Histograma do trace Andando-1 para $X \geq 1067$ - métrica *Packet Size* - Gamma(22.406, 0.02)