

Análise não supervisionada para inferência de qualidade de experiência de usuários residenciais

Gustavo H. A. Santos¹, Gabriel Mendonça¹,
Edmundo de Souza e Silva¹, Rosa M.M. Leão¹, Daniel S. Menasché¹

Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, RJ.

{gustavo, gabriel, edmundo, rosam, sadoc}@land.ufrj.br

Resumo. *A avaliação da qualidade de experiência dos usuários residenciais é de grande interesse para ISPs. No entanto, a obtenção da QoE percebida é custosa, dificultando a utilização de classificadores supervisionados. Este trabalho propõe um método baseado em aprendizado de máquina não supervisionado que detecta padrões estatísticos nas séries temporais a partir da detecção de pontos de mudança e da correlação espaço-temporal dos resultados de medições de QoS. Exemplificamos a aplicação do método em um conjunto de dados reais, mostrando que os resultados do modelo refletem uma métrica de QoE dos usuários obtida a partir de chamados técnicos realizados para o call center. Por fim, avaliamos a acurácia da execução online do método.*

Abstract. *The assessment of quality of experience for residential users is a major goal for ISPs. However, it is costly to obtain the QoE perceived by clients, making it difficult to use a supervised approach. We propose an unsupervised machine-learning based method to detect statistical patterns in time series using change point detection and spatial-time correlation. We apply our method to a real dataset, showing that our model results reflects a customer QoE metric obtained from complaints to the call center. Finally, we evaluate the accuracy of the method for online detection.*

1. Introdução

Não é recente o esforço de provedores de serviços da Internet para avaliar o impacto de métricas de desempenho da rede na qualidade *percebida* pelos usuários dos serviços. Com o crescimento da complexidade das redes residenciais e o aumento exponencial do uso de serviços de transmissão de vídeo e voz, é premente compreender as complexas interações entre o desempenho da rede e a qualidade de experiência dos usuários (QoE) [Bustamante et al. 2017]. Identificar as causas de uma QoE inadequada não é uma tarefa simples, uma vez que a QoE pode ser resultante de problemas em diferentes regiões, como na rede do ISP, na rede doméstica, nos servidores da aplicação, nos serviços de DNS etc.

Uma das maneiras mais comuns para inferir a qualidade de experiência dos usuários a partir de métricas de desempenho de rede se baseia em experimentos com voluntários em ambiente controlado [Charonyktakis et al. 2016, da Silva et al. 2008]. Entretanto, realizar esse tipo de procedimento é uma tarefa custosa, demorada e com baixo potencial de generalização, já que os resultados obtidos para um conjunto de aplicações ou ambientes podem não se aplicar a outros. Uma forma para mitigar alguns desses problemas é usar mecanismos de *crowdsourcing*. No entanto, problemas

como a falta de controle dos experimentos e a existência de usuários não confiáveis exigem o desenvolvimento de mecanismos sofisticados para garantir a integridade dos dados [Hoßfeld et al. 2014].

Uma possível maneira de se estimar a QoE dos clientes seria observar os registros de reclamações recebidas pelo ISP em seu *call center* [Pan et al. 2018]. Intuitivamente, chamados de usuários relatando insatisfação com os serviços de Internet teriam relação com a sua QoE no uso de aplicações. Esses eventos podem estar relacionados a variações estatísticas em métricas de desempenho de rede, como a taxa de perda de pacotes e a latência. Entretanto, correlacionar métricas de QoS com ligações recebidas em um *call center* não é uma tarefa simples. Por exemplo, uma ligação a um *call center* pode ser desestimulada pelo tempo perdido pelo usuário com o atendimento [Pan et al. 2018]. Além disso, um cliente só perceberá um problema de rede se estiver usando uma aplicação sensível ao problema no momento em que esse ocorre. Consequentemente, é de se esperar que seja reduzido o número de reclamações que efetivamente estejam relacionadas a problemas com a rede do provedor. Portanto, o uso de chamadas ao *call center* como rótulo para uma análise supervisionada não é trivial, podendo inclusive gerar resultados com viés (*bias*) ou *overfitting*.

Propomos uma abordagem baseada em aprendizado de máquina não supervisionado para extrair padrões de séries temporais reais obtidas de medições ativas realizadas em um provedor (ISP) de porte médio a partir da borda de rede (roteadores domésticos). Identificamos *pontos de mudanças estatísticas* nas séries temporais e fazemos uma correlação espaço-temporal dos intervalos entre esses pontos de cada série e a topologia da rede do ISP. Este nosso trabalho estende o apresentado em [Ximenes et al. 2018] de várias formas. Em primeiro lugar, o modelo escolhido para a detecção de pontos de mudança é mais facilmente parametrizável do que o utilizado em [Ximenes et al. 2018]. O novo modelo permite a interpretação dos resultados e é ainda mais robusto a *outliers*. Além disso, utilizamos um *dataset* com granularidade temporal bem mais fina do que o anteriormente considerado. Diferentemente do trabalho anterior, o *dataset* atual inclui períodos de indisponibilidade. Finalmente, avaliamos uma aplicação *online* do método e correlacionamos os resultados obtidos de forma não supervisionada com chamados de clientes ao *call center* do ISP.

O artigo exemplifica a utilidade da proposta usando um conjunto de dados reais e mostrando que é possível identificar e localizar regiões da rede do ISP que experimentam degradação de desempenho. Mostramos ainda que os resultados do modelo não supervisionado refletem a percepção de qualidade dos serviços de rede, correlacionando os chamados técnicos realizados para o *call center* com um indicador do modelo. Avaliamos também a acurácia da execução *online* do método.

O trabalho é organizado como se segue. A Seção 2 descreve o ambiente de medição utilizado e o conjunto de dados deste trabalho. O método proposto é detalhado na Seção 3, que inclui uma breve introdução sobre o problema de detecção de pontos de mudança estatística e o nosso modelo. Os resultados obtidos são mostrados na Seção 4. A Seção 5 discute trabalhos relacionados na literatura, enquanto a Seção 6 conclui o trabalho.

2. Dataset

Os dados deste trabalho resultaram de medições ativas realizadas por roteadores residenciais instalados em 2485 clientes de um ISP de médio porte e rodando o sistema operacional OpenWrt¹. As medições foram realizadas por 47 dias, de 08/08/2018 a 23/09/2018. Conforme argumentado em [Sundaresan et al. 2011], coletar dados de desempenho de rede a partir de roteadores residenciais fornece uma visão mais precisa da qualidade do serviço prestada ao usuário, em contraste com a coleta realizada no interior da rede do ISP. A Figura 1 mostra a arquitetura de nosso ambiente de medição.

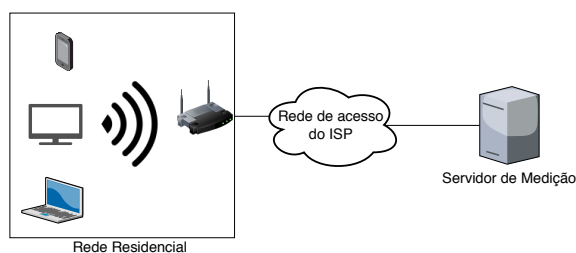


Figura 1. Infraestrutura de medição

A cada minuto as seguintes métricas são coletadas: o número de *bytes* (*upload* e *download*), a latência bidirecional e a taxa de perda de pacotes bidirecional de cada usuário. A latência e a taxa de perda são obtidas a partir do envio de uma rajada de 100 pacotes ICMP a intervalo de 10 milissegundos entre pacotes e destinados a um servidor localizado na rede do ISP. As rotas utilizadas durante as medições incluem apenas equipamentos dentro da rede do ISP. A partir desses dados também é possível inferir períodos de indisponibilidade de um usuário. O tráfego (*upload* e *download*) é medido de maneira passiva a partir do contador de *bytes* associado a cada interface do roteador.

Trabalhos da literatura mostram que a perda de pacotes tem maior impacto na QoE do que a latência [da Silva et al. 2008, Charonyktakis et al. 2016]. Uma análise preliminar do nosso conjunto de dados, omitida por falta de espaço, confirma essa hipótese. Desta forma, utilizamos somente a taxa de perda de pacotes como métrica básica de desempenho para os estudos que se seguem. Entretanto, a latência média ou outras métricas podem ser consideradas conjuntamente, sem maiores modificações na nossa metodologia.

A Figura 2 mostra a distribuição do percentual de perda de pacotes considerando todas as medições realizadas. A figura indica que 93% das medições não apresentaram perdas. Além disso, dentre as medições com alguma perda de pacotes, a maioria (55.6%) tinha apenas 1 pacote perdido em 100.

Os dados mostram também que a taxa de perda ao longo do tempo possui um padrão diário sendo menor no período da madrugada e no começo da manhã, e aumentando no período entre 19 e 22 horas, quando o tráfego médio é maior.

Na Figura 3 exemplificamos a variação da taxa de perda com o tempo de um único usuário dos 2485 participantes das medições. No início do período considerado, o cliente apresentou uma porcentagem de perda predominantemente menor ou igual a 1%. Um pico pode ser observado no dia 26/08, em que 5% das medições do usuário apresentaram

¹www.openwrt.org

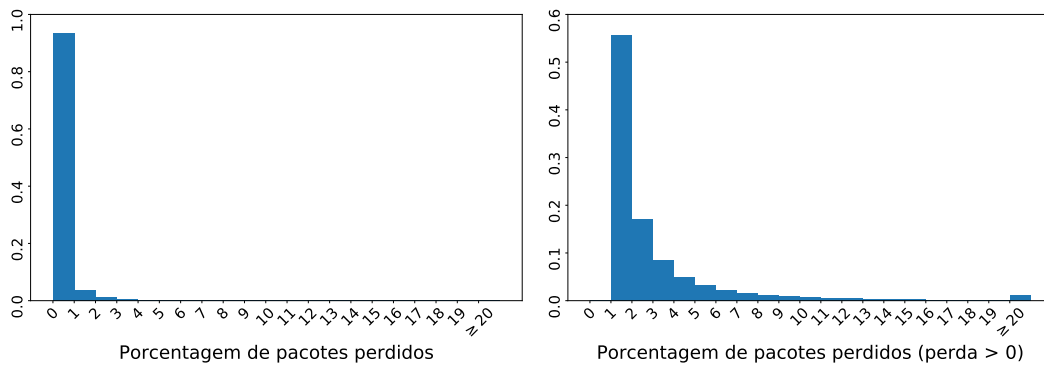


Figura 2. Distribuição de perda

taxa maior ou igual a 3%. Percebe-se também uma mudança no comportamento a partir do dia 06/09, quando ao menos 95% das medições não apresentaram perda.

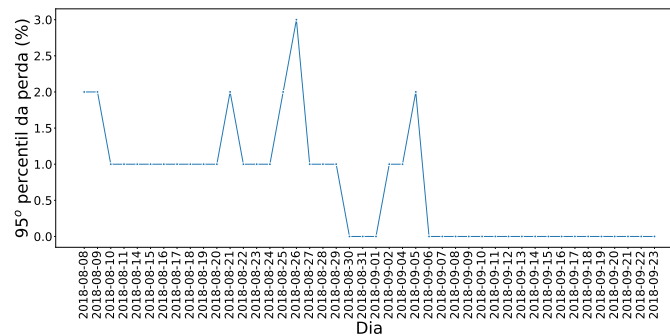


Figura 3. Variação temporal da perda de um usuário

No que se segue descrevemos o pré-processamento realizado nas séries temporais para facilitar a análise. Períodos de indisponibilidade também são considerados como amostras válidas das séries.

As séries temporais são construídas a partir das medidas realizadas a cada minuto e portanto o tempo é discretizado em *bins* de 1 minuto. Neste trabalho, dividimos as medidas coletadas de cada cliente em períodos de 24 horas. Desta forma, cada série temporal terá potencialmente $n = 1440$ amostras (ou *bins* de 1 minuto). Para cada *bin*, é verificada a quantidade de tráfego na rede residencial do cliente gerado ou recebido pelos dispositivos de sua rede (informalmente chamado de tráfego “do cliente”). O tráfego do cliente, quando gerado concomitantemente a uma medição ativa, interfere com o tráfego de medição ativa no roteador. Portanto, pode afetar os resultados das medições [Sundaresan et al. 2011]. Sendo assim, é importante filtrar os resultados coletados em um *bin* sempre que o tráfego do cliente for maior do que um limiar θ . O valor de θ é escolhido proporcionalmente à vazão nominal dos clientes. Os *bins* filtrados devido ao tráfego do cliente são marginalizados durante o período de treinamento conforme descrito em [Yu and Kobayashi 2003].

Identificamos também os *bins* com ausência de medições. Uma medição não realizada pode ser consequência da inexistência de rota ao servidor de medições do ISP, o

que pode ser relevante para a análise (há outras causas para ausência de medições como, por exemplo, roteador desligado, mas essas causas podem ser facilmente identificadas). Esses *bins* são rotulados com um “símbolo” (ϕ), para indicar indisponibilidade.

Para garantir um número mínimo de amostras com uma medição de perda de pacotes, descartamos séries temporais com um número de medições inferior a um limiar κ , dentre as $n = 1440$ potenciais amostras diárias. Neste trabalho, escolhemos $\theta = 2.5$ Mbps, por ser esse valor de tráfego suficientemente baixo para não afetar uma medida. Escolhemos $\kappa = 1000$ amostras, igual a aproximadamente 70% do total de amostras diárias em potencial. Usando esses valores, nosso *dataset* resultou em um total de 76741 séries temporais. Deste total, separamos 51041 séries para treinamento, correspondentes ao intervalo de 08/08/2018 a 10/09/2018, e 25700 para testes, correspondentes ao intervalo de 11/09/2018 a 23/09/2018.

3. Metodologia

Nesta seção apresentamos a metodologia proposta. Ela é baseada no treinamento (não supervisionado) de uma HMM para inferir a qualidade da rede de acesso de um cliente em um determinado momento. Aplica-se o modelo para cada cliente, sendo em seguida realizada uma correlação espaço-temporal visando identificar regiões na rede com potenciais problemas.

3.1. Detecção de Pontos de Mudança

Seja uma série temporal de observações $\mathbf{x} = x_1, \dots, x_t, \dots, x_n$, onde x_t representa o valor da série no tempo t e n representa o número de amostras. Técnicas de detecção de pontos de mudanças estatísticas têm como objetivo identificar o número de pontos de mudança k e o conjunto de pontos $\tau = \tau_1, \dots, \tau_k$ tal que o intervalo (τ_i, τ_{i+1}) possua características estatísticas distintas dos intervalos imediatamente anterior a τ_i e posterior a τ_{i+1} . Diversas abordagens para a detecção desses pontos podem ser encontradas na literatura [Aminikhanghahi and Cook 2017]. Modelos baseados em cadeias de Markov ocultas (*hidden Markov models* - HMM) [Kehagias 2004, Luong et al. 2013, Montañez et al. 2015] têm sido amplamente usados para esse fim, sendo esse o método escolhido neste trabalho.

HMMs são processos estocásticos que modelam a distribuição de uma série temporal associando a cada amostra x_t , denominada símbolo observável, uma variável latente s_t denominada estado oculto. Este modelo supõe que a transição entre dois estados ocultos possui a propriedade de Markov. Uma das principais vantagens da utilização de HMMs é a possibilidade de interpretação do estado oculto associado a cada amostra e, em consequência, a atribuição de um significado para diferentes segmentos da série temporal. Uma visão detalhada sobre HMMs pode ser encontrada em [Rabiner 1989, de Souza e Silva et al. 2011].

Neste trabalho, consideramos como símbolos observáveis as medidas de taxa de perda de cada cliente. A partir dos símbolos observáveis e da definição do número de estados ocultos, estima-se a distribuição da taxa de perda associada a cada estado. Com base nas distribuições obtidas durante o período de treinamento, pode-se tentar extrair alguma interpretação física para os estados ocultos, como veremos nas próximas seções.

Cada estado de uma HMM corresponde potencialmente a uma distribuição distinta para os valores observáveis (taxa de perda). Portanto, uma mudança de estado na cadeia oculta é naturalmente mapeada a um ponto de mudança de distribuição dos símbolos observáveis. Os estados associados a uma sequência de observações podem ser inferidos a partir do algoritmo de Viterbi, que obtém a sequência mais provável de estados ocultos $\mathbf{S} = s_1, s_2, \dots, s_n$ a partir da sequência de símbolos observáveis $\mathbf{O} = O_1, O_2, \dots, O_n$. Desta forma, uma transição de um estado s_i para $s_{i+1} \neq s_i$ representa um ponto de mudança de distribuição. Devido às vantagens citadas, utilizamos HMMs para a identificação de pontos de mudança.

Em [Montañez et al. 2015], os autores argumentam que a utilização de HMMs pode gerar um número elevado de transições que não refletem os pontos de mudança reais. Para mitigar o problema, os autores propõem a adição de um fator de regularização durante o aprendizado dos parâmetros da cadeia. Entretanto, a modificação proposta pelo método torna o processo de treinamento mais complexo.

Para evitar aumentar a complexidade do modelo, neste trabalho propomos um filtro recursivo simples para lidar com o problema. Seja $\mathbf{S} = s_1, s_2, \dots, s_n$ a sequência de estados ocultos obtidos após a aplicação do algoritmo de Viterbi. A partir de \mathbf{S} obtemos a sequência filtrada $\mathbf{S}' = s'_1, s'_2, \dots, s'_n$ como se segue. Seja $\mathbf{I}_{s_t}(s_i)$ a função indicadora definida como: $\mathbf{I}_{s_t}(s_i) = 1$ se $s_t = s_i$ e zero caso contrário. Considerando uma janela de amostras de tamanho δ , obtemos a sequência \mathbf{S}' como se segue. Inicializa-se com $s'(1) = s(1)$ e para $t > 1$:

$$s'_t = \begin{cases} s_t & \text{se } \frac{1}{\delta} \sum_{i=t-\delta+1}^t \mathbf{I}_{s_t}(s_i) = 1, t \geq \delta \\ s'_{t-1} & \text{caso contrário} \end{cases}$$

Nos nossos experimentos, a sequência filtrada \mathbf{S}' conseguiu reduzir satisfatoriamente o número de *outliers*, isto é, estados em \mathbf{S} com pequeno tempo de permanência por visita.

No que se segue, nosso objetivo é diferenciar séries temporais que possam indicar potenciais problemas na rede. Para atingir esse objetivo, será necessária a interpretação dos estados ocultos da HMM treinada com dados reais. Essa interpretação depende da distribuição dos símbolos e da qualidade de serviço aceitável em relação à métrica taxa de perdas. Informalmente, rotulamos como estados de qualidade “ruim” aqueles cuja distribuição de símbolos seja tal que a probabilidade de gerar taxas de perdas “altas” não seja desprezível. Semelhantemente, em estados com qualidade “boa” a probabilidade de uma taxa de perda alta é muito baixa, enquanto estados com qualidade “média” são os “intermediários”. Usaremos também um outro rótulo (“inacessível”) para indicar indisponibilidade da rede. Na Seção 4, onde mostraremos o modelo treinado, os rótulos ficarão evidentes.

Baseado nos rótulos dos estados, no que se segue atribuiremos rótulos para cada série temporal (já filtrada) tomando como base o tempo total de permanência nos estados, durante n minutos. Intuitivamente, séries temporais com rótulo “vermelho” possuem longo tempo em estados “ruins”. Caso contrário, a série é dita “azul”. É importante ressaltar que a atribuição de rótulos aos estados e conseqüentemente às séries depende do que se considera como qualidade ruim ou boa a partir da distribuição da taxa de perda de pacotes. Por outro lado, a metodologia independe dessa atribuição.

Seja \mathcal{R} o conjunto de estados do modelo considerados “ruins”. Definimos o rótulo $L(S')$ como uma função indicadora igual a 1, se o tempo total de permanência no conjunto \mathcal{R} durante n é maior que um limiar γ , e igual a 0 caso contrário. O parâmetro γ depende dos requisitos de qualidade de serviço almejados pelo ISP. Nos nossos estudos, o valor de γ/n foi atribuído empiricamente igual a 1,5% (0,015), relativamente baixo. Entretanto, como veremos na Seção 4, é possível correlacionar os resultados do modelo com chamados técnicos dos clientes para o ISP.

3.2. Correlação espaço-temporal

A partir dos rótulos $L(S')$ definimos um algoritmo de correlação espaço-temporal com o objetivo de agrupar regiões da rede que possuam padrões estatísticos similares. O algoritmo de correlação espaço-temporal assume que as rotas entre os pontos de medição e o servidor são estáticas durante a janela de tempo considerada. Como consequência, a topologia da rede pode ser representada por uma árvore em que cada nó equivale a um ou mais equipamentos de rede. Supomos ainda que a degradação de performance de um nó E afeta o desempenho de todos os clientes que fazem parte da subárvore de raiz E .

A abordagem proposta é adaptada do algoritmo de detecção de causa raiz de [Ximenes et al. 2018]. As principais diferenças são: (a) a associação de rótulos $L(S')$ para cada série temporal em S' , possibilitando a interpretação dos padrões estatísticos observados; (b) o método de [Ximenes et al. 2018] exige que os as métricas de QoS (e.g. perda de pacotes) sejam coletadas a partir dos nós folhas da árvore resultante da topologia de rede. Essa restrição é relaxada neste trabalho, de forma que as métricas de desempenho podem ser coletadas de qualquer nó da topologia.

Para a aplicação da correlação, associa-se a cada ponto de medição um “voto” de acordo com a função $L(S')$ definida na subseção anterior. Para cada nó E da árvore de topologia calcula-se uma função $\Gamma(E)$, obtida a partir dos votos relacionados a todos os nós pertencentes à subárvore de raiz E . Em outras palavras, o cálculo de $\Gamma(E)$ é dado através de um mecanismo de votação realizado a partir dos rótulos $L(S')$ atribuídos às séries obtidas pelos pontos de medição.

Diferentes métodos de votação podem ser utilizados para definir $\Gamma(E)$, como votação com peso ou votação baseada em limiar. Uma formalização de diferentes mecanismos de votação pode ser encontrada em [Parhami 1994]. A utilização de modelos probabilísticos, como no nosso caso, permite a definição de pesos para os votos de acordo com a *likelihood* associada a cada série temporal. Por questão de simplicidade e falta de espaço, consideramos para este trabalho o mecanismo de votação majoritária.

A Figura 4 exemplifica a execução do método de correlação espaço-temporal utilizando votação majoritária. Para cada nó E da árvore, o processo de votação considera os votos dos medidores conectados à subárvore de raiz E . Ao fim da votação, obtém-se grupos distintos de nós cujas medições compartilham padrões estatísticos similares.

4. Resultados

Nesta seção, apresentamos os resultados da aplicação do método proposto. Discutimos o modelo de detecção de pontos de mudança obtido na Seção 4.1. A Seção 4.2 exemplifica a aplicação da correlação espaço-temporal com o objetivo de localizar regiões da rede que experimentam degradação em seu desempenho. Mostramos a relação dos resultados

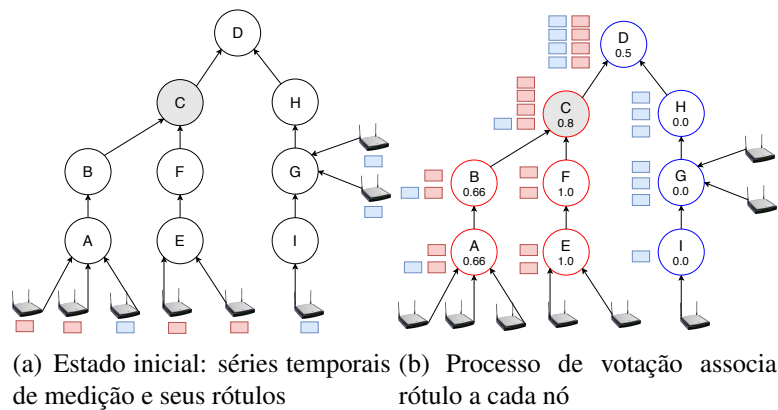


Figura 4. Aplicação do método de votação majoritária

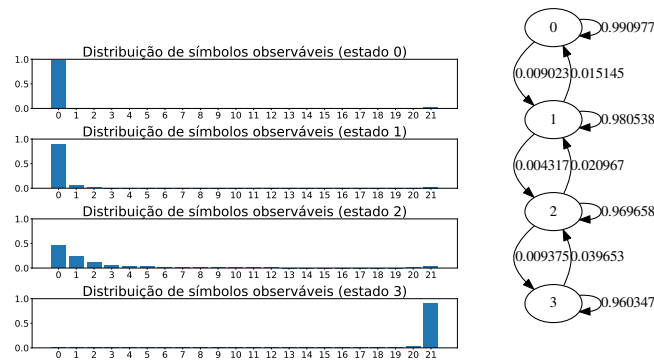


Figura 5. Modelo de Markov oculto obtido após o treinamento

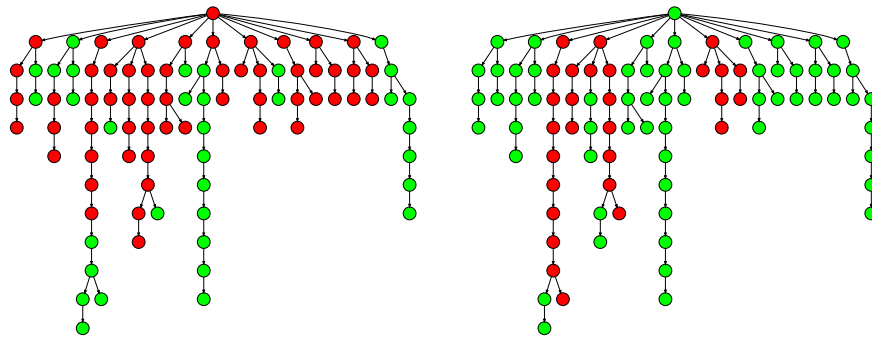
do nosso modelo com a QoE dos usuários na Seção 4.3. Por fim, a Seção 4.4 discute a aplicação *online* do método.

4.1. Detecção de pontos de mudança

O número de estados, símbolos e estrutura da HMM utilizada resultou de uma escolha criteriosa entre diferentes modelos, comparação da *likelihood* etc., mas omitimos detalhes por restrição de espaço. Consideramos uma cadeia de Markov oculta com 4 estados que gera símbolos de acordo com uma distribuição categórica. As transições podem ocorrer apenas entre estados ocultos adjacentes como mostra a Figura 5. Codificamos as medições de perda utilizando 22 símbolos.

A Figura 5 mostra o modelo obtido após o processo de treinamento. Percebe-se que o estado 2 apresenta a distribuição com a maior média de perda, enquanto os dois primeiros estados geram perda 0 com alta probabilidade. O estado 3, por outro lado, possui uma alta probabilidade de gerar símbolos relativos a indisponibilidade (símbolo 21). O estado 2 pode ser interpretado como o período em que a rede apresenta problemas, enquanto o estado 3 representa períodos em que a rede se torna indisponível.

Para a obtenção da sequência de estados S' , aplicamos o algoritmo de Viterbi e realizamos a etapa de pós-processamento considerando $\delta = 5$ e $\gamma = 20$. Com o objetivo de detectar pontos de mudança relacionados a períodos com maior probabilidade de taxas



(a) Correlação espaço temporal (08/08) (b) Correlação espaço temporal (23/09)

Figura 6. Correlação espaço-temporal aplicada nos dias 08/08/2018 (esq.) e 23/09/2018 (dir.). Nós em vermelho experimentam degradação de performance.

mais altas de perda, ou de indisponibilidade (períodos “ruins”), definimos $\mathcal{R} = \{2, 3\}$.

4.2. Correlação espaço-temporal

Para exemplificar a correlação espaço-temporal, consideramos um subconjunto de 42 clientes conectados a uma mesma parte da topologia da rede. Esse subconjunto foi escolhido por conta do número relevante de medidores instalados na rede do ISP.

As Figuras 6(a) e 6(b) mostram o resultado da correlação espaço-temporal para dois dias distintos (08/08/2018 e 23/09/2018). Adotamos um mecanismo de votação majoritária para definir os rótulos $\Gamma(E)$. Percebe-se que a região não apresentava um bom desempenho de rede no dia 08/08: a maior parte da topologia identificava problemas, incluindo a raiz da árvore. Durante o período considerado, diversas modificações foram realizadas pelo ISP com o objetivo de sanar os problemas da região. A partir da Figura 6(b), é possível notar que as mudanças realizadas surtiram efeito, já que boa parte da região apresentou melhoras. No entanto, um conjunto relevante de nós ainda apresentava degradação em sua performance, o que sugere que a atuação do ISP nessa região ainda é necessária.

Em seguida, avaliamos a evolução temporal da qualidade do serviço prestado nessa região ao longo do período de 08/08/2018 a 23/09/2018. Mostramos na Figura 7 a fração de séries temporais em estados do conjunto \mathcal{R} (vermelho) em cada minuto. Nota-se uma clara melhora no desempenho dos clientes associados a essa topologia, com uma fração cada vez menor de séries temporais visitando estados em \mathcal{R} . Um exemplo dessa mudança é apresentado na Figura 8, que mostra as séries temporais dos dias 08/08 e 23/09 de um cliente conectado a um nó que apresentou melhora. Percebe-se que esse cliente observava um período significativo com taxas de perdas relevantes antes da intervenção do ISP (pontos em vermelho). Ao final do período considerado, o cliente apresenta uma grande melhora em seu desempenho, tendo apenas eventos isolados de perda.

4.3. Avaliação de QoE

A partir da parceria com o ISP, obtivemos acesso a informações sobre os chamados realizados por clientes dessa região. A existência de um chamado permite identificar momentos em que usuários possuem sua qualidade de experiência impactada. Consideramos para

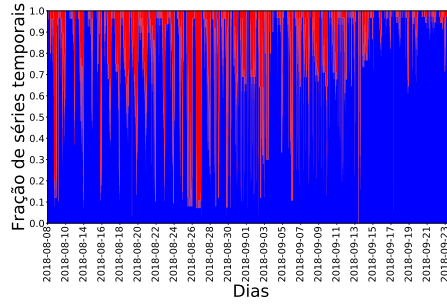
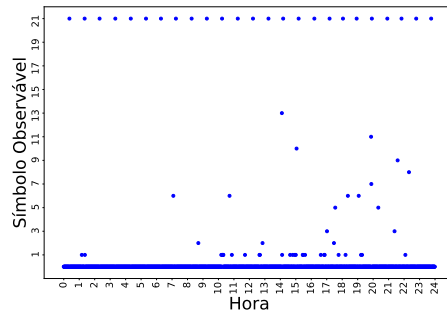
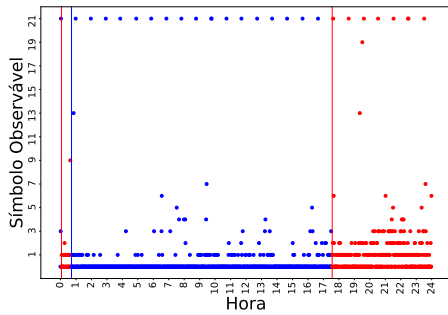


Figura 7. Fração de amostras das séries temporais associadas a estados em \mathcal{R} . Estados em \mathcal{R} são representados pela cor vermelha.



(a) Série temporal antes da intervenção (08/08) (b) Série temporal após intervenção (23/09)

Figura 8. Mudança de comportamento da rede de usuário após intervenção na rede do ISP

esta análise que um nó E da topologia apresenta uma QoE insatisfatória num determinado dia se ao menos um cliente conectado a E (com ou sem roteador de medição) abriu um chamado relatando problemas técnicos na rede. O número de chamados ao ISP é pequeno relativamente ao número total de usuários. Desta forma, consideramos para a avaliação dos resultados a relação entre a abertura de ao menos um chamado técnico relacionado ao nó E e o rótulo $\Gamma(E)$ associado ao nó, utilizando como métrica o *recall*.

Relembramos que o valor do rótulo de um nó E ($\Gamma(E)$), é calculado a partir de um processo de votação que se inicia a partir dos nós com medidores usando $L(S')$, e caminha em direção ao nó da árvore. A Tabela 1 mostra os resultados obtidos. Nota-se que os resultados do modelo refletem a qualidade de experiência observada pelos usuários da seguinte forma: quase 89% dos nós com chamados foram identificados a partir do modelo quando consideramos o subconjunto de nós com ou sem medidores conectados. Resultados ainda melhores são obtidos quando consideramos apenas nós com medidores diretamente conectados, onde 95% dos nós com chamados foram identificados. Desta forma, a aplicação do método proposto permite ao ISP agir de maneira pró-ativa, identificando regiões com degradação de desempenho nas quais intervenções melhorariam a qualidade de experiência dos usuários e diminuiriam a quantidade de chamados.

4.4. Aplicação *online* do algoritmo

A análise realizada anteriormente considera que cada série temporal x possui um conjunto de amostras de um dia ($n = 1440$). Para a utilização do método de ma-

Conjunto de nós	Recall	Número de nós com chamado
Nós com medidores diretamente conectados	0.951	61
Nós com rótulo $\Gamma(E)$ associado	0.886	123

Tabela 1. Relação entre chamados e rótulos $\Gamma(E)$ associados a nós da rede

neira *online*, consideramos que a sequência de estados ocultos S^* é obtida através da aplicação do algoritmo de Viterbi utilizando uma janela deslizante centrada de tamanho $2\alpha + 1$. Cada estado oculto s_t é estimado a partir de uma janela de amostras $W = x_{t-\alpha}, x_{t-\alpha+1}, \dots, x_t, \dots, x_{t+\alpha-1}, x_{t+\alpha}$. Em consequência da utilização de uma janela centrada de amostras de tamanho $2\alpha + 1$, a identificação do estado oculto s_t é feita com um atraso de α amostras.

Para avaliar o desempenho do algoritmo *online*, comparamos a sequência de estados ocultos S obtida a partir da série temporal de um dia e a sequência de estados ocultos S^* obtida a partir da aplicação *online* do método. Considera-se que S corresponde ao *ground truth* e que o objetivo do método *online* é a obtenção de uma sequência S^* muito similar à sequência S . Desta forma, utilizamos como métrica de comparação a fração de estados ocultos s_t iguais nas sequências S e S^* .

A Figura 9 mostra a acurácia obtida por diferentes tamanhos de janela. A partir deste gráfico percebe-se que a utilização de uma janela de 121 amostras possui uma acurácia de aproximadamente 0,96 no conjunto de testes, enquanto a utilização de uma janela de 61 amostras possui uma acurácia de aproximadamente 0,87. Em consequência, concluímos que janelas de amostras com menos de um dia podem ser utilizadas uma vez que as sequências de estados (obtidas a partir do algoritmo de Viterbi) são praticamente idênticas. O tamanho da janela deve ser ajustado de acordo com o *tradeoff* entre o tempo α necessário para a detecção do estado oculto e a acurácia desejada.

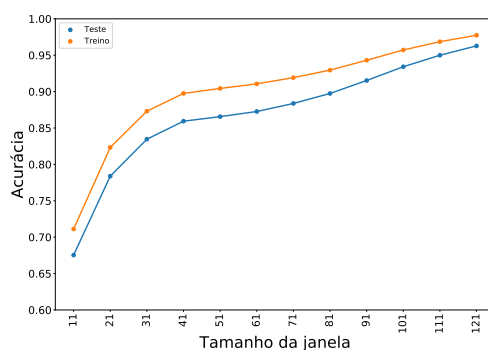


Figura 9. Taxa de acerto para aplicação de Viterbi *online*

5. Trabalhos Relacionados

A avaliação da qualidade de experiência é vastamente explorada na literatura [Hora et al. 2016][Charonyktakis et al. 2016][da Silva et al. 2008]. Muitos trabalhos, como [Charonyktakis et al. 2016][da Silva et al. 2008], são baseados em medições da QoE realizadas a partir de testes em laboratório, que são custosos e de difícil execução. Abordagens baseadas em *crowdsourcing* são apresentadas como alternativas à utilização

de testes em laboratório visando a redução de custos. Entretanto, desafios como a existência de usuários não confiáveis e a falta de controle dos experimentos exigem a aplicação de mecanismos sofisticados para garantir que os resultados sejam válidos [Hofsfeld et al. 2014]. [Pan et al. 2018] propõe a utilização de dados de satisfação dos usuários obtidos a partir de chamadas ao *call center*. No entanto, a abordagem utilizada é supervisionada e aplicada a redes celulares.

Outros trabalhos da literatura utilizam resultados de medições para detecção de problemas. Muitos não levam em consideração a topologia da rede, se aplicando apenas a redes celulares [Padmanabha Iyer et al. 2018, Deb et al. 2017, Pan et al. 2018]. As ferramentas propostas em [Adams et al. 2016, Peng et al. 2017] utilizam técnicas de *probing* dinâmico entre múltiplos pontos da rede para detectar falhas e desconexões em *data centers* com base no conhecimento da topologia da rede. Por outro lado, focamos em nosso trabalho na detecção de problemas em um ISP residencial. Já em [Yan et al. 2012], os autores propõem o sistema Argus, que atua na identificação e localização de problemas a partir da agregação espacial e temporal de métricas de QoS fim-a-fim em uma rede fixa. Enquanto nosso trabalho se utiliza de técnicas de detecção de pontos de mudança, [Yan et al. 2012] aplicam técnicas de detecção de anomalia baseadas no algoritmo de Holt-Winters.

Este trabalho é uma extensão de [Ximenes et al. 2018]. Ambos se baseiam na detecção de pontos de mudança e realizam correlação espaço-temporal para encontrar regiões da rede com performance similar. [Ximenes et al. 2018] aplica um algoritmo de detecção baseado em otimização proposto por [Maidstone et al. 2017], que é sensível a *outliers* e possui difícil parametrização. Nosso trabalho, por outro lado, realiza detecção de pontos de mudança utilizando cadeias de Markov ocultas, que são: (1) mais robustas a *outliers*, (2) mais facilmente parametrizáveis e (3) mais fáceis de se interpretar. Além disso, como mostramos na Seção 4.4, nosso modelo pode ser aplicado na detecção *online* de problemas, o que traz grandes benefícios para o ISP ao permitir a mitigação mais rápida de falhas na rede. Por fim, ressaltamos que o presente trabalho conta com uma base de dados mais completa contendo séries de perda de pacotes mais refinadas², informações de indisponibilidade de rede e registro de chamados técnicos realizados pelos usuários da rede, que nos permitem relacionar a degradação no serviço prestado pelo ISP com a queda na qualidade percebida pelo cliente.

6. Conclusão e trabalhos futuros

Neste trabalho propomos uma metodologia não supervisionada para avaliar a QoE de usuários residenciais a partir de dados reais obtidos de um ISP de porte médio. As principais contribuições do trabalho são: (a) um modelo baseado em cadeias de Markov ocultas (HMM) usado para a detecção de pontos de mudanças estatísticas de séries temporais. O modelo foi parametrizado a partir de séries reais de taxa de perda de pacotes, coletadas de medições realizadas a partir de roteadores residenciais de quase 2500 clientes do ISP por quase 2 meses; (b) nosso modelo permite a identificação de mudanças estatísticas em séries temporais e, em particular, períodos de degradação da QoS dos usuários participantes das medições. O modelo é extensível a outras métricas de QoS como, por exemplo,

²Nosso intervalo de amostragem é de 1 minuto, enquanto que em [Ximenes et al. 2018] os intervalos são de no mínimo 30 minutos

a latência; (c) usamos a topologia da rede do ISP para identificar, através do modelo não supervisionado e da correlação espaço-temporal dos resultados, regiões com QoS não satisfatória; (d) correlacionamos os resultados da metodologia com os chamados (técnicos) de clientes realizados para a operadora. Tais chamados constituem uma importante medida da qualidade de experiência dos clientes. Finalmente, mostramos que o método pode ser aplicado de forma *online*. Como trabalhos futuros incluiremos outras métricas de QoS nos modelos e ainda ampliaremos os períodos de coleta de dados. Planejamos a implementação do método *online* em um número significativo de clientes.

Agradecimento: Este trabalho é parcialmente suportado por projeto de cooperação MCTIC-RNP/NSF, MCTIC/FAPESP, e projetos do CNPq e FAPERJ, além de bolsas CAPES.

Referências

- Adams, A., Lapukhov, P., and Zeng, J. H. (2016). NetNO-RAD: Troubleshooting networks via end-to-end probing. <https://code.facebook.com/posts/1534350660228025/netnorad-troubleshooting-networks-via-end-to-end-probing/>.
- Aminikhanghahi, S. and Cook, D. J. (2017). A survey of methods for time series change point detection. *Knowledge and information systems*, 51(2):339–367.
- Bustamante, F., Clark, D., and Feamster, N. (2017). Workshop on Tracking Quality of Experience in the Internet: Summary and Outcomes. *SIGCOMM Comput. Commun. Rev.*, 47(1):55–60.
- Charonyktakis, P., Plakia, M., Tsamardinos, I., and Papadopouli, M. (2016). On user-centric modular QoE prediction for VoIP based on machine-learning algorithms. *IEEE Transactions on Mobile Computing*, 15:1443–1456.
- da Silva, A. P. C., Varela, M., de Souza e Silva, E., Leão, R. M. M., and Rubino, G. (2008). Quality assessment of interactive voice applications. *Computer Networks*, 52(6):1179–1192.
- de Souza e Silva, E., Leão, R. M. M., and Muntz., R. R. (2011). Performance evaluation with hidden markov models. In *Performance Evaluation of Computer and Communication Systems. Milestones and Future Challenges*, pages 112–128.
- Deb, S., Ge, Z., Isukapalli, S., Puthenpura, S., Venkataraman, S., Yan, H., and Yates, J. (2017). Aesop: Automatic policy learning for predicting and mitigating network service impairments. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1783–1792.
- Hoßfeld, T., Keimel, C., Hirth, M., Gardlo, B., Habigt, J., Diepold, K., and Tran-Gia, P. (2014). Best practices for QoE crowdtesting: QoE assessment with crowdsourcing. *IEEE Transactions on Multimedia*, 16(2):541–558.
- Hora, D. N. d., Teixeira, R., van Doorselaer, K., and van Oost, K. (2016). Predicting the effect of home Wi-Fi quality on web QoE. In *Proceedings of the 2016 Workshop on QoE-based Analysis and Management of Data Communication Networks*, Internet-QoE '16, pages 13–18.

- Kehagias, A. (2004). A hidden markov model segmentation procedure for hydrological and environmental time series. *Stochastic Environmental Research and Risk Assessment*, 18(2):117–130.
- Luong, T. M., Rozenholc, Y., and Nuel, G. (2013). Fast estimation of posterior probabilities in change-point analysis through a constrained hidden markov model. *Computational Statistics and Data Analysis*, 68:129 – 140.
- Maidstone, R., Hocking, T., Rigai, G., and Fearnhead, P. (2017). On optimal multiple changepoint algorithms for large data. *Statistics and Computing*, 27(2):519–533.
- Montañez, G. D., Amizadeh, S., and Laptev, N. (2015). Inertial hidden markov models: Modeling change in multivariate time series. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, AAAI'15, pages 1819–1825. AAAI Press.
- Padmanabha Iyer, A., Erran Li, L., Chowdhury, M., and Stoica, I. (2018). Mitigating the latency-accuracy trade-off in mobile data analytics systems. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*, pages 513–528. ACM.
- Pan, H., Zhou, S., Jia, Y., Niu, Z., Zheng, M., and Geng, L. (2018). Data-driven user complaint prediction for mobile access networks. *Journal of Communications and Information Networks*, 3(3):9–19.
- Parhami, B. (1994). Voting algorithms. *IEEE transactions on reliability*, 43(4):617–629.
- Peng, Y., Yang, J., Wu, C., Guo, C., Hu, C., and Li, Z. (2017). deTector: a Topology-aware Monitoring System for Data Center Networks. In *2017 USENIX Annual Technical Conference (USENIX ATC 17)*, pages 55–68.
- Rabiner, L. R. (1989). A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE*.
- Sundaresan, S., de Donato, W., N.Feamster, Teixeira, R., Crawford, S., and Pescapè, A. (2011). Broadband internet performance: A view from the gateway. In *Proceedings of the ACM SIGCOMM 2011*.
- Ximenes, D., Mendonça, G., Santos, G. H. A., de Souza e Silva, E., Leão, R. M., and Menasché, D. S. (2018). O problema de detecção e localização de eventos em séries temporais aplicado a redes de computadores. *Workshop em Desempenho de Sistemas Computacionais e de Comunicação (WPerformance CSBC)*, 17(1/2018).
- Yan, H., Flavel, A., Ge, Z., Gerber, A., Massey, D., Papadopoulos, C., Shah, H., and Yates, J. (2012). Argus: End-to-end service anomaly detection and localization from an isp's point of view. In *INFOCOM, 2012 Proceedings IEEE*, pages 2756–2760.
- Yu, S.-Z. and Kobayashi, H. (2003). A hidden semi-markov model with missing data and multiple observation sequences for mobility tracking. *Signal Processing*, 83(2):235–250.