k-DynMix: Um Mecanismo de Proteção Dinâmica de Privacidade em Mix-Zones

Ekler Paulino de Mattos[1,2], Augusto C. S. A. Domingues[1], Fabrício A. Silva[3] Heitor S. R. Filho[1], and Antonio A. F. Loureiro[1] *

[1]Departamento de Ciência da Computação - Universidade Federal de Minas Gerais
Belo Horizonte, MG

[2]Universidade Federal de Mato Grosso do Sul - Campus de Coxim
Coxim, MS

[3]Universidade Federal de Viçosa – Campus Florestal
Florestal, MG

{ekler.mattos, augusto.souza, ramosh, loureiro}@dcc.ufmg.br,

fabricio.asilva@ufv.br

Abstract. *Mix-zones are an anonymization-based privacy protection mechanism used in various contexts against tracking attacks. However, mix-zones depend on factors that affect their performance, e.g., defining fair privacy levels. This work proposes k-DynMix, a dynamic mix-zone that adjusts the level of privacy over time in online mode and linear complexity, according to the flow of vehicles, to achieve greater anonymization. In the experiments, we analyzed real and synthetic datasets comparing k-DynMix with two prediction engines to estimate privacy over time and with classic mix-zones using mix-zone coverage and Anonymization Quality metrics. The results showed that k-DynMix outperformed the prediction engines in predicting privacy. Further to maximizing privacy, it achieved performance similar to the best result of classic mix-zones.*

Resumo. *Mix-zones é mecanismo proteção de privacidade baseado de anonimização usados em diversos contextos contra ataques de rastreamento. No entanto, as mix-zones dependem de fatores que afetam o seu desempenho, e.g., a definição de níveis de privacidade justos. Este trabalho propõe o k-DynMix, um esquema de mix-zone dinâmica que ajusta o nível de privacidade ao longo do tempo em modo online e complexidade linear, de acordo com fluxo de veículos, para alcançar maior anonimização. Nos experimentos, analisamos conjuntos de dados reais e sintéticos comparando o k-DynMix com dois mecanismos de predição para estimar a privacidade ao longo do tempo e com mix-zones clássicas usando métricas de cobertura e Anonymization Quality de mix-zones. Os resultados mostraram que o k-DynMix superou os mecanismos de predição em prever privacidade. Além de maximizar a privacidade, ele obteve um desempenho semelhante ao melhor resultado das mix-zones clássicas.*

*Este trabalho foi realizado com apoio da Fundação de Amparo a Pesquisa do Estado de Minas Gerais (FAPEMIG) (APQ-00426-22); Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) (#2023/00721-1 e #2018/23064-8); Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (#312682/2021-2); e Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) (Finance Code 001).

1. Introdução

Na era da computação pervasiva, tecnologias como *Internet of Things* (IoT) e *Internet of Vehicles* (IoV) facilitaram a interconexão de dispositivos que ampliaram o compartilhamento de serviços de localização nas cidades inteligentes [Paiva et al. 2020]. No entanto, estes serviços geram dados de localização massivos e irrestritos relativos aos cidadãos, levantando a certas preocupações em termos de privacidade. Ao analisar os dados de localização, é possível descobrir informações sobre pontos de interesse, comportamento individual e coletivo a partir da mobilidade dos cidadãos, inclusive a sua identidade [de Mattos et al. 2022]. Para mitigar a identidade dos usuários, *Location Privacy Protection Mechanism*s (LPPMs) baseadas em anonimização foram propostas, como as *mix-zones* – áreas designadas de anonimização definidas por um raio r, onde as entidades mudam seus pseudônimos a partir do nível de privacidade k ao atingir um mínimo de k entidades simultaneamente dentro dela [Freudiger et al. 2007, Domingues et al. 2021].

Apesar das mix-zones serem amplamente utilizadas em muitas áreas, incluindo *Vehicular ad-hoc network*s (VANETs), elas ainda apresentam algumas limitações. Notavelmente, as mix-zones dependem de fatores como posicionamento, geometria, padrões de mobilidade, densidade de veículos e taxas de chegada, que podem afetar o seu desempenho. Parâmetros mal configurados de mix-zones degradam a taxa de anonimização ou protegem dados com um baixo nível de privacidade habilitando ataques de rastreamento e inferência [de Mattos et al. 2022, Khodaei and Papadimitratos 2020]. Embora existam propostas para estas questões, pouco tem sindo explorado sobre como melhorar o desempenho de níveis de privacidade e o comportamento de qualidade das mix-zones.

Este trabalho propõe o k-Dynamic Mix-zone (k-DynMix), um mecanismo de mix-zone dinâmica que ajusta o nível de privacidade k ao longo do tempo em modo online e complexidade linear, de acordo com eventos como flutuações no tráfego de veículos para alcançar a noção de maior anonimização. Inspiramos o k-DynMix em métricas de qualidade de mix-zones e alguns conceitos de mecanismos de controle de congestionamento TCP. Nos experimentos, analisamos dois conjuntos de dados real e sintético, comparando o k-DynMix com dois mecanismos de predição para estimar a privacidade ao longo do tempo e com mix-zones clássicas em relação às métricas de cobertura e *Anonymization Quality* (AQ). Os resultados mostraram que o k-DynMix superou os mecanismos de predição em prever privacidade. Além de maximizar a privacidade, obteve um desempenho semelhante ao melhor resultado das mix-zones clássicas. Ao contrário das mix-zones estáticas, o k-DynMix obteve AQ para todas as mix-zones, mostrando melhor comportamento das mix-zones, inclusive para mix-zones com baixo tráfego. Até onde sabemos, esta é a primeira abordagem de privacidade dinâmica ao longo do tempo que considera as flutuações do tráfego de veículos para maximizar a privacidade, a eficácia e a qualidade de anonimização em mix-zones. As contribuições deste trabalho são:

- k-Dynamic Mix-zone (k-DynMix): uma abordagem dinâmica de mix-zones em tempo real baseados no fluxo de veículos, visando alcançar maior anonimização;
- Introdução a noção de *Higher Anonymization*;
- Integração de métricas de *Anonymization Quality* e conceitos de mecanismos de controle de congestionamento TCP para o projeto de LPPMs dinâmicas;
- Avaliação do k-DynMix com mix-zones estáticas em termos de métricas de cobertura e de qualidade de anonimização com dados reais de táxis. Análise do k-DynMix em relação a métodos de predição com dados sintéticos e reais de táxis.

A Seção 2 discute os trabalhos relacionados; A Seção 3 aborda conceitos e a definição do problema; A Seção 4 introduz o k-DynMix. A Seção 5 descreve a avaliação dos experimentos. As Seções 6 e 7 detalham os resultados e as considerações finais.

2. Trabalhos Relacionados

As mix–zones são dependentes de vários fatores, tais como o posicionamento, geometria, padrões de mobilidade, densidade de veículos, taxas de chegada e ajustes de seus parâmetros [de Mattos et al. 2022, Khodaei and Papadimitratos 2020]. As mix-zones veiculares se diferem das mix-zones tradicionais devido às suas restrições espaciais e temporais, trajetórias exclusivamente em estradas, direção, adesão às regras e condições de trânsito e das estradas [Chow and Mokbel 2011]. A seguir, apresentamos algumas propostas que abordam essas questões.

[Freudiger et al. 2007] abordaram aspectos espaço-temporais e de segurança para VANETs. Eles propuseram o protocolo CMIX que criptografa mensagens de mudanças de pseudônimos nas mix-zones posicionadas em cruzamentos. [Palanisamy and Liu 2014] propuseram resolver as limitações nos padrões de movimento e comportamentos estatísticos dos usuários com mix-zones adaptativas de formato não retangular, onde o comprimento é determinado pela velocidade média do segmento da estrada, a janela de tempo e o limite mínimo de entropia em pares de usuários.

Para mitigar ataques de inferência facilitados por limitações de velocidade em cruzamentos e semáforos onde estão situadas mix-zones, [Zhou and Zhang 2019] introduziram uma abordagem que incorpora troca de pseudônimos e adiciona ruído às trajetórias que passam por mix-zones. Sobre a infraestrutura e posicionamento de mix-zones, [Yamazaki et al. 2021] propuseram um esquema de mix-zones centrado em veículos para anonimizar dados em vez de *Road Side Unit*s (RSUs). Eles também propuseram resolver o atraso na comunicação entre veículos com mix-zones de duas camadas, onde uma delas desconsidera os veículos que causarão atrasos.

Esforços foram feitos para compreender o comportamento das mix-zones e a utilidade dos dados anonimizados. Em estudos anteriores, evidenciamos que a mobilidade pode impactar as abordagens de privacidade de localização, como mix-zones, no contexto da mobilidade inteligente [de Mattos et al. 2022]. Mostramos os efeitos colaterais da privacidade e utilidade de dados quando os parâmetros das mix-zones são configurados incorretamente em um ambiente multimodal. Além disso, estudamos como o comportamento das mix-zones afetam a *Anonymization Quality* (AQ) para o projeto e seleção de LPPMs robustas [Mattos et al. 2022, de Mattos et al. 2023]. AQ é um conceito relacionado à proteção, eficácia e funcionamento interno de LPPMs. Abordamos como identificar aplicações e serviços de cidades inteligentes que possam melhor aproveitar os dados de mobilidade anonimizados por mix-zones. Para isso, propusemos uma metodologia e posteriormente um *framework* que avalia a utilidade em diversos aspectos com métricas relacionadas à aspectos sociais, de privacidade e mobilidade de trajetórias anonimizadas produzidas por mix-zones [Mattos et al. 2023, de Mattos et al. 2024].

Até onde sabemos, nenhuma proposta anterior de mix-zones foi focada em ajustar a privacidade ao longo do tempo. Diferente de trabalhos anteriores, avançamos no estado da arte propondo o k-DynMix: uma mix-zone dinâmica que ajusta a privacidade ao longo do tempo conforme as flutuações do tráfego veicular para obter maior anonimização.

3. Mix-zones e Definição do Problema

Nesta seção apresentamos o esquema de mix-zones e descrevemos o problema das mix-zones relacionado ao nível de privacidade.

Uma mix-zone M é uma área geográfica de anonimato k pela qual os veículos passam, fazendo com que seus pseudônimos sejam modificados [Chen et al. 2018]. Quando um veículo em movimento entra em M com raio r, sua trajetória será dividida em duas sub-trajetórias delimitadas por pseudônimos diferentes - um correspondente à parte antes e o outro à parte após a mix-zone. Os veículos mudam de pseudônimo dentro da mix-zone se houver *Mix-zone Activation* (MA), i.e., $MA \leftarrow |A| \geq k$, onde A é um conjunto de veículos simultaneamente presentes em M, denotado como conjunto de anonimato, e k é o nível de privacidade de M [Beresford and Stajano 2003, Beresford and Stajano 2004]. A configuração o nível de privacidade de um k justo, é um fator crítico que afeta o desempenho das mix-zones [de Mattos et al. 2022, Khodaei and Papadimitratos 2020]. Quanto maior o valor de k, maior é a privacidade. Em contrapartida, quanto menor k, menor é a privacidade. Em ambos os casos, o anonimato ocorre se houver MA na mix-zone M. Ou seja, o número de veículos (*Number of Cars on Mix-zone* (NCM)) deve ser maior ou igual a k, i.e., $MA \leftarrow NCM \geq k$.

Algorithm 1: k-DynMix

Data: k_t; k_b; NCM; t; τ_{lst}
Result: κ
$\eta, \tau_{lst} \leftarrow$ timeoutArrCar(t, τ_{lst}, NCM)
if $(\eta) \vee (NCM < k_t)$ **then**
  $\rho \leftarrow \frac{k_t}{2}$
  $\kappa \leftarrow k_b$
end
else
  if $k_t < \rho$ **then**
    $\kappa \leftarrow$min$(2k_t, \rho)$
  end
  else
    $\kappa \leftarrow k_t + 1$
  end
  if $(\kappa > NCM) \wedge (\kappa > k_b)$ **then**
    $\kappa \leftarrow \kappa - c$
  end
end

Algorithm 2: TAC alg.

Data: t; τ_{lst}; NCM
Result: η; τ
$\eta \leftarrow False$
$\delta \leftarrow$ computeITM(t, NCM)
if $(\delta > \tau_{lst}) \vee (|\tau_{lst} - \delta| > \tau_{thsh})$ **then**
  $\delta_{est} \leftarrow (1 - \alpha)\delta_{est} + \alpha\delta$
  $\delta_{dev} \leftarrow (1 - \beta)\delta_{dev} + \beta|\delta - \delta_{est}|$
  $\tau \leftarrow \delta_{est} + u\delta_{dev}$
  if $\delta > \tau_{lst}$ **then**
    $\eta \leftarrow True$
  end
end

Se M não atingir MA, os veículos não serão anonimizados, isto degrada as taxas de eficácia e anonimização de M permitindo ataques de ligação e inferência com altas taxas de sucesso. A degradação da eficácia da mix-zone é mais evidente ao usar uma configuração estática de k. Em um cenário onde M é configurada com um alto k mas baixo tráfego de veículos (baixo NCM), M não irá atingir MA [de Mattos et al. 2023]. Já em um cenário de alto tráfego de veículos e M configurada com baixo k, o nível de privacidade pode não ser eficiente. Em condições ideais, o melhor cenário de anonimização para M é o – *Optimal Anonymization (OA)* – quando a distribuição k é igual a NCM ao longo do tempo t, ou seja, $NCM_t = k_t$. No entanto, prever k para alcançar OA ao longo do tempo é complexo e requer o mecanismo de predição ideal. Assim, apresentamos um relaxamento de OA chamada – *Higher Anonymization (HA)* – onde a distribuição de k ao longo do tempo segue o NCM tendo-o como limite inferior o NCM, o que satisfaz a anonimização do MA o mais próximo possível do fluxo de tráfego. Ou seja, quanto mais

próximo de k se aproximar do NCM por um limite inferior, maior será o nível de privacidade possível. Ter trajetórias com *HA* torna o problema de re-identificação mais difícil porque a re-identificação dos usuários é um problema intrinsecamente combinatório, onde um atacante tenta vincular as trajetórias dos usuários cortadas por pelo menos k veículos. Portanto, quanto maior o k, maior será a combinação a ser feita pelo atacante para re-identificar os usuários. Resumimos esta discussão com a definição de *HA*:

► **Def. 1 -** *Higher Anonymization (HA)*. Seja $\mathcal{M}$ um conjunto de mix-zones e $M_{k(t)} \in \mathcal{M}$ uma mix-zone configurada com privacidade k no tempo t. Para $M_{k(t)}$ obter maior anonimização (HA), deve-se $(NCM_t \geq k_t \wedge k_t \approx NCM_t) \implies MA_t^H$, onde k_t, NCM_t e MA_t^H são níveis de privacidade, *Number of Cars on Mix-zone* (NCM) é uma variável aleatória e *Mix-zone Activation* (MA) em um *HA* no tempo t, respectivamente.

4. O Mecanismo k-Dynamic Mix-zone (k-DynMix)

A ideia do k-DynMix é controlar o nível de privacidade ao longo do tempo com base em eventos que ocorrem nas mix-zones. A partir desses eventos, a estratégia é ajustar k como um limite inferior, mas próximo de NCM o mais rápido possível para atender a (Def.1). Assim, o k-DynMix deve decrementar e aumentar exponencial e linearmente o valor de k a partir de eventos em M. Identificamos eventos relacionados ao fluxo de tráfego veicular que ocorre em M:

- ***Mix-zone Activation* (MA)**: sinaliza que M está ativo, ou seja, $NCM \geq k$, indica a presença de fluxo de veículos em M, portanto é possível anonimizar veículos;
- ***Mix-zone Deactivation* (MD)**: complemento de MA, sinaliza que M está desativado, ou seja, $NCM < k$, não sendo possível anonimizar os veículos;
- ***Timeout of Arriving Cars in Mix-zones* (TAC)**: evento em que os veículos não chegam a M no tempo limite τ definido. Pode indicar ausência de carros em M para atingir o MA, consequentemente não gerando anonimização. TACs podem ocorrer em períodos com baixas taxas de chegada de veículos, como ao amanhecer. TAC deve acompanhar as oscilações do tráfego ao longo do tempo em M [de Mattos et al. 2023] e considerar as taxas de entrada de veículos recentes que refletem o fluxo atual em M.

Experimentos empíricos foram feitos sobre o crescimento de k. Observamos que o tempo para atender a Def. 1 é maior no crescimento linear do que no crescimento exponencial. Porém, usar crescimento exponencial de k, supera o limite NCM rapidamente e produz eventos MD. A estratégia de previsão do k-DynMix é que para eventos MD e TAC em M a privacidade diminui significativamente para a privacidade inicial k_b para atingir MA. Definir k para k_b é uma abordagem para atender situações de baixo tráfego de veículos em k_b e também para atingir MA. Em seguida, define-se um limite de privacidade ρ como metade de k, o ρ orquestra o crescimento de k. Quando MA ocorre, k cresce exponencialmente, até atingir ρ. Quando k for igual a ρ, k aumenta linearmente, para evitar perda de privacidade com MD, até atingir o NCM.

4.1. Definição do k-DynMix

Seja M uma mix-zone posicionada num cruzamento rodoviário e implantada em uma RSU habilitado com um sensor veicular que periodicamente detecta o *Number of Cars on Mix-zone* (NCM) em M. A cada iteração, M invoca o k-DynMix (Algo. 1), seus parâmetros são a privacidade k_t e *NCM* no tempo t; k_b é a privacidade inicial; e o último

tempo limite τ_{lst}. A saída de Algo. 1 é a privacidade esperada κ. A linha 1 calcula o *timeout* da estimativa de carros que chegam em M, com a função *timeoutArrCar*, definido em Algo. 2, onde verifica se ocorreu *timeout* η e o cálculo do último tempo limite τ_{lst}. As linhas 2-5 de Algo. 1 calculam o decremento κ nos eventos que significam baixo tráfego em M – *timeout* e MD – e κ deve ser definido como k_b, ρ é definido como metade k_t. As linhas 6-12 são sobre o MA e calculam o incremento κ sendo linear e exponencial. Se k_t for menor que ρ indica que k_t está longe do valor *NCM*, então κ aumenta exponencialmente até atingir ρ (linhas 7-9). Em contraste, quando k_t é grande ou igual a ρ significa que k_t está próximo de *NCM*, o κ aumenta linearmente em uma unidade até chegar ao *NCM*. Desta forma, buscamos atender HA, com nível máximo de privacidade igual a *NCM*. Na mesma iteração do algoritmo, κ pode ser um limite superior, mas próximo de *NCM*, resultando em MD. Então, ajusta-se κ subtraindo-o a constante c, onde $\kappa \geq c$ (linhas 13-15).

O algoritmo TAC (Algo. 2) é inspirado no mecanismo Temporizador de Retransmissão do TCP, que calcula e gerencia o temporizador de retransmissões (*timeouts*) no host remetente [Paxson et al. 2011]. Em nosso contexto, o TAC favorece a eventos recentes de *Interval of Arrival Time between Cars on Mix-zones* (ITM) em detrimento dos antigos para calcular o *timeout* dos veículos em M. O ITM é uma métrica de qualidade de mix-zone que mede o intervalo de tempo em segundos entre os veículos que entram na mix-zone [Mattos et al. 2022, de Mattos et al. 2023].

Os parâmetros do Algo. 2 são o último TAC (τ_{lst}) e o NCM no tempo t, e a saída são dois valores: η representa se houve evento de *timeout*; τ é o TAC para a próxima iteração. η é inicializado com False (linha 1), indicando que não obteve um *timeout*. A linha 2 calcula a métrica de qualidade ITM δ. Em cada evento de entrada de veículo em M, o ITM é calculado. As linhas 3-10 calculam o τ e se ocorreu *timeout* η em M. O *timeout* será estimado em dois casos (linha 3). No primeiro caso, se o ITM for maior que o último *timeout* estimado τ_{lst}. O segundo caso, para prevenção de *timeout* altos, controlado se a diferença entre τ_{lst} e ITM for maior que o *timeout* τ_{thsh}. A linha 4 calcula o ITM estimado (δ_{est}) que é uma combinação ponderada entre os valores anteriores de ITMs (δ_{est}) e δ, α é o peso para favorecer uma das duas variáveis. O δ_{est} favorece os valores recentes de ITMs. A linha 5 calcula o desvio entre ITMs δ_{est} e δ para suavizar o cálculo do *timeout*. Esta equação baseia-se em média exponencial ponderada para considerar o δ_{dev} mais recente. O *timeout* é calculado na linha 6 com a soma de δ_{est} mais uma margem de erro, que tem valor alto com a alta variação de δ_{dev} e baixa em caso contrário, com um fator de margem u, onde $u > 0$. As linhas 7-9 retornam se houve evento de *timeout* η.

▶ *Análise da Complexidade de Tempo.* Considerando que M tem um loop de tamanho n, que detecta NCM e então executa o k-DynMix com uma complexidade de $\Theta(n)$. O Algo. 1 tem $\Theta(1) + 2\Theta(1)$, sendo o Algo. 2 tem $\Theta(1)$. Portanto, a complexidade de tempo completa é: $T(\textit{k-DynMix}) = \Theta(n).(\Theta(1) + 2\Theta(1)) = \Theta(n)$

5. Avaliação dos Experimentos

Nesta seção apresentam as análises realizadas para a avaliação do k-DynMix. A avaliação do k-DynMix é feita por três análises.

▶ *Análise da Acurácia das Técnicas de Predição de k.* A primeira análise, investigaremos a capacidade do k-DynMix prever k para atingir MA ao longo do tempo. Iremos comparar o k-DynMix com duas abordagens amplamente utilizadas de média móvel. O

Simple Moving Average (SMA) calcula a média dos n pontos de dados anteriores, onde cada ponto de dados tem igual importância. O *Weighted Exponential Moving Average (WEMA)* permite usar funções de ponderação específicas, como atribuir peso elevado a pontos recentes. Mais detalhes sobre essas abordagens, veja o Apêndice A. Utilizaremos a métrica ***Accuracy in k's Predictions* (ACC_{pred})** para avaliar as abordagens de predição, representado por $ACC_{pred} = |MA|/(|MA| + |MD|)$, onde $|MA|$ é o número de tentativas que uma abordagem estimou k e obteve MA, pelo total de tentativas ($|MA| + |MD|$).

► *Análise de Cobertura das Mix-zones.* A segunda análise, exploramos as métricas de cobertura de mix-zones extraídas de configurações estáticas e do k-DynMix. As métricas de cobertura da mix-zone M são:

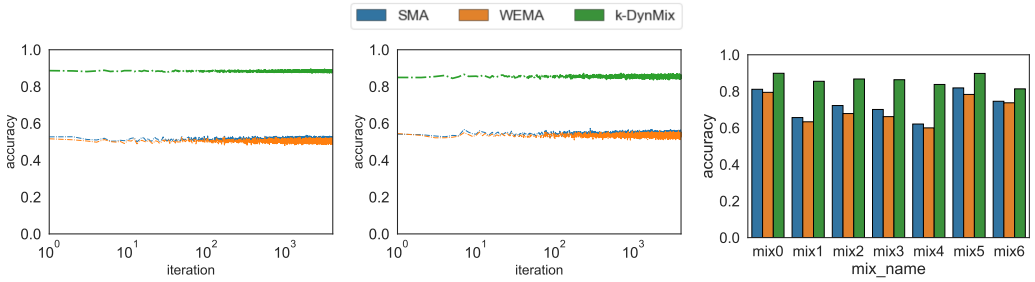
- ***Anonymization Rate* (AR)**: AR_M é o número de trajetórias que passaram e foram anonimizadas por M;
- ***Non-Anonymization Rate* (NAR)**: NAR_M é o número de trajetórias que passaram e não foram anonimizadas por M;
- ***Mix-zone Efficacy* (ME)**: é a proporção entre um número de usuários anonimizados por M, AR_M, e a população P_M que cruzou M, ou seja, $ME_M = |AR_M|/|P_M|$.

► *Análise Qualidade da Anonimização das Mix-zones.* Analisamos o k-DynMix em termos de *Anonymization Quality* (AQ) na terceira análise. AQ são métricas que mensuram o funcionamento da mix-zones e refletem na anonimização. As mix-zones com AQ apresentam elevada eficácia e níveis de privacidade consideráveis, anonimizando os dados de mobilidade no momento em que são ativados. As métricas de AQ são:

- ***Number of Cars on Mix-zone* (NCM)**: Número de veículos que cruzam M ao longo do tempo;
- ***Interval of Arrival Time between Cars on Mix-zones* (ITM)**: Intervalo de tempo em segundos entre a entrada de veículos em M, possibilitando a medição do volume de tráfego dentro de M;
- ***Interval of Departure Time between Cars on Mix-zones* (IDM)**: Intervalo de tempo, em segundos, entre a saída dos veículos de M, possibilitando a medição do volume de tráfego dentro de M;
- ***Number of Trips Completed within the Mix-zone (NTC)***: Número de veículos que terminam suas viagens dentro de M ;
- ***Activation Time of the Mix-zone* (ATM)**: Período de tempo em que M está ativada para anonimização: quando a quantidade de veículos dentro dela for maior ou igual ao parâmetro k. Um ATM alto denota mix-zones com anonimização por um longo período de tempo.

A partir das métricas de qualidade, calcula-se a função AQ - descreve os aspectos considerados de privacidade e comportamento da mix-zone ao longo do tempo [de Mattos et al. 2023]. O AQ de um período diário s é denotado pela distribuição de Boltzmann, representada por $AQ_s = H_s(u')\phi_s\epsilon^{-(\mathcal{I}_s+\mu)/\lambda}$. Os $H_s(u')$, ϕ_s, $\mathcal{I}_s$ e μ representam o peso da privacidade, o número total de ATMs, a taxa de transferência $\mathcal{I}_s = IDM/ITM$ e NTC na mix-zone M, no período s, respectivamente. O λ é a constante de Boltzmann. O $\epsilon^{-(\mathcal{I}_s+\mu)/\lambda}$ é usado para penalizar o resultado quando muitos veículos entram nas mix-zones e não saem delas. Consequentemente, esses veículos não são anonimizados. Especificamente, para a relação $\mathcal{I}_s + \mu > 1$ então $\epsilon^{-(\mathcal{I}_s+\mu)/\lambda}$ tende a zero. Caso contrário, se $\mathcal{I}_s + \mu \leq 1$ então $\epsilon^{-(\mathcal{I}_s+\mu)/\lambda}$ tende para um. $AQ \in [0, 1]$ onde 1 é o valor máximo para a qualidade de anonimização.

6. Resultados e Discussão



(a) Gaussian distr. D_c **(b) Unknown distr. D_d** **(c) Cabs distr. D_b**

Figura 1. Precisão (ACC_{pred}) das técnicas de predição de k usando NCM de distribuições gaussianas (1a), aleatórias (1b) e reais (1c). A Fig. 1c: NCM extraído do conjunto de dados Cabspotting, anonimizados com mix-zones posicionadas com o FPMT.

Esta seção apresenta os resultados e discussões sobre o desempenho do k-DynMix em prever privacidade em relação à dois mecanismos de predição. Além disso, avaliamos o k-DynMix com mix-zones com configurações estáticas em termos de cobertura e AQ.

6.1. Coleção de Dados

Neste estudo, dois conjuntos de dados foram analisados. O primeiro deles, denominado *Cabspotting* com dados de mobilidade reais de aproximadamente 500 taxis coletados em São Francisco, EUA, durante 25 dias [Piorkowski et al. 2009]. Coletado em 2008, possui informações sobre a localização dos taxis, amostradas periodicamente por um sensor GPS embarcado. O *Cabspotting* possui cerca 440.000 viagens, com uma média de 17.600 viagens por dia e cobertura acima de 70% das ruas da cidade, com aproximadamente 400.000 contatos entre veículos por dia, o que comprova o potencial significativo do conjunto de dados para a nossa análise. Duas amostras de períodos de tráfego intenso foram extraídas do *Cabspotting* para a análise. A primeira, D_a, refere-se ao dia 18 de Maio de 2008 (domingo), contendo 366.951 registros, 422 usuários distintos e 1.770 viagens; a utilizamos para analisar a qualidade da anonimização dentre outras métricas das mix-zones. A segunda amostra, D_b, refere-se ao dia 19 de Maio de 2008 (segunda), correspondendo a 417.781 registros, 454 usuários distintos, e 2.036 viagens que utilizamos para a análise da predição de k baseadas no NCM das mix-zones posicionadas na área da cidade.

Além disso, para validar as técnicas de predição de k para as mix-zones, utilizamos dois conjuntos de dados sintéticos de series temporais que simulam o fluxo de trafego em uma mix-zone. Especificamente, geramos NCM aleatórios usando uma distribuição Gaussiana (D_c) e aleatória (D_d) de tamanho 2.000, variando a amostragem entre 1 segundo e 30 minutos. Em relação a distribuição Gaussiana (D_c), utilizamos a media e o desvio padrão de uma distribuição NCM iguais a 10 e 2,5, respectivamente. Os limites inferior e superior da distribuição aleatória de valores de NCM foram 0 e 21.

As mix-zones foram selecionadas com um algoritmo de posicionamento denominado FPMT, que gera uma lista de mix-zones candidatas em ordem decrescente, baseado a frequência de veículos nas interseções [Mattos et al. 2019]. Somente sete mix-zones foram escolhidas da lista, baseadas em dois critérios. Primeiro, as mix-zones precisam anonimizar trajetórias em pelo menos dois de três cenários de configuração para k. Segundo, não pode haver sobreposição topológica entre as mix-zones escolhidas. Devido a essas condições e o tamanho limitado da região de São Francisco, o montante de sete mix-zones foi definido como o orçamento de privacidade.

6.2. Análise de Mecanismos de Predição de Nível de Privacidade

O primeiro passo para projetar mix-zones dinâmicas capazes de ajustar o valor de k ao longo do tempo é identificar um mecanismo de predição eficiente para estimar o valor ideal para k. Particularmente, analisamos a ACC_{pred} do k-DynMix e das técnicas de predição definidas na Seção 5 utilizando dois conjuntos de dados sintéticos de NCM com distribuições Gaussiana (D_c) e aleatória (D_d). Utilizamos *bootstrapping* para gerar os conjuntos de dados re-amostrando um total de 4.000 iterações. Para cada iteração, calculamos ACC_{pred}. A distribuição aleatória resultou em uma media de 9,2% de NCM nas amostras, nas quais NCM é menor que o valor mínimo de k, indicando um comportamento mais realístico que a distribuição Gaussiana, que obteve valor próximo de zero. A Fig. 1a apresenta o ACC_{pred} para a distribuição de NCM Gaussiana, na qual o k-DynMix superou outros métodos de predição, alcançando um ACC_{pred} médio de 88,3%. SMA e WEMA tiveram resultados piores, com *avg(ACC_{pred})* de 51,4% e 50,4%, respectivamente. O mesmo comportamento pode ser observado para a distribuição aleatória (Fig. 1b), em que o k-DynMix teve um *avg(ACC_{pred})* igual a 85,5%, seguido por SMA e WEMA, com valores iguais a 54,3% e 53,5%, respectivamente.

Em relação à análise do NCM com o fluxo real de tráfego de táxis (D_b), analisamos o NCM de sete mix-zones previamente implantadas por um algoritmo de posicionamento [Mattos et al. 2019] (veja Tabela 1). Esta análise apresentou um comportamento diferente dos conjuntos de dados sintéticos (ver Fig. 1c). Pois D_b teve cerca de 44,5% das amostras de NCM, cujo NCM é inferior ao mínimo do k, indicando um conjunto de dados ainda mais realista que os sintéticos D_c e D_d. As abordagens de média móvel tiveram um melhor desempenho do que com conjuntos de dados sintéticos, mas foram inferiores ao k-DynMix, que teve melhor desempenho em todas as mix-zones. O SMA e WEMA tiveram *avg(ACC_{pred})* de 72,5% e 69,8%, mas o k-DynMix obteve 86,2%, destacando um pico nas mix0 e mix5 com ACC_{pred} até 89,8%. Os resultados sugerem que o k-DynMix superou os mecanismos de predição em prever privacidade ao longo do tempo.

Tabela 1. Mix-zones posicionadas, localizações, locais e coef. variação.

Nome	Latitude	Longitude	Coef. Var. (%)	Localização	Locais Cobertos pela Mix-zone
mix0	37.714801	-122.397982	74,13	101 express way, Visitacion Valley.	cafes, subways station, restaurants.
mix1	37.724830	-122.400157	74,43	101 express way, Bay View.	supermarkets, restaurants, clinics.
mix2	37.735133	-122.404532	82,49	101 express way, Bernal Heights.	road interchange, gas station, supermarket.
mix3	37.676005	-122.391491	75,95	101 express way, Firth Park.	tourist area, hotels, docks.
mix4	37.615315	-122.393566	65,68	entrance road to the airport.	cabs park, bus stop, restaurant, garages
mix5	37.774378	-122.401540	72,10	downtown, near 101 express way.	exit to Oakland city, shopping, church.
mix6	37.768990	-122.419450	**93,11**	downtown, Mission District.	tourist area, museums, subways station.

Tabela 2. (Non) Anon. and Efficacy Geral: mix-zones estáticas e o k-DynMix.

	No Anon	Anon	Total	Efficacy
$k = 2$	3048	5421	8469	0,640
$k = 4$	7482	987	8469	0,117
$k = 6$	8289	180	8469	0,021
k-DynMix	3332	5137	8469	**0,607**

6.3. Caracterização e Mix-zones com Métricas de Cobertura

Após as mix-zones serem posicionadas com o algoritmo de posicionamento definido em [Mattos et al. 2019], anonimizamos o conjunto de dados D_a com configurações estáticas de mix-zones k = 2, 4, 6 e o k-DynMix com um raio igual a 500 metros, conforme proposto em [de Mattos et al. 2023]. A Tabela 1 mostra a localização das mix-zones implantadas, locais cobertos por elas e o Coeficiente de Variação (CV)[1] de veículos

[1]O Coeficiente de Variação (CV) mede a dispersão de uma distribuição de probabilidade usada para identificar a porcentagem de variação sobre a média central de uma amostra.

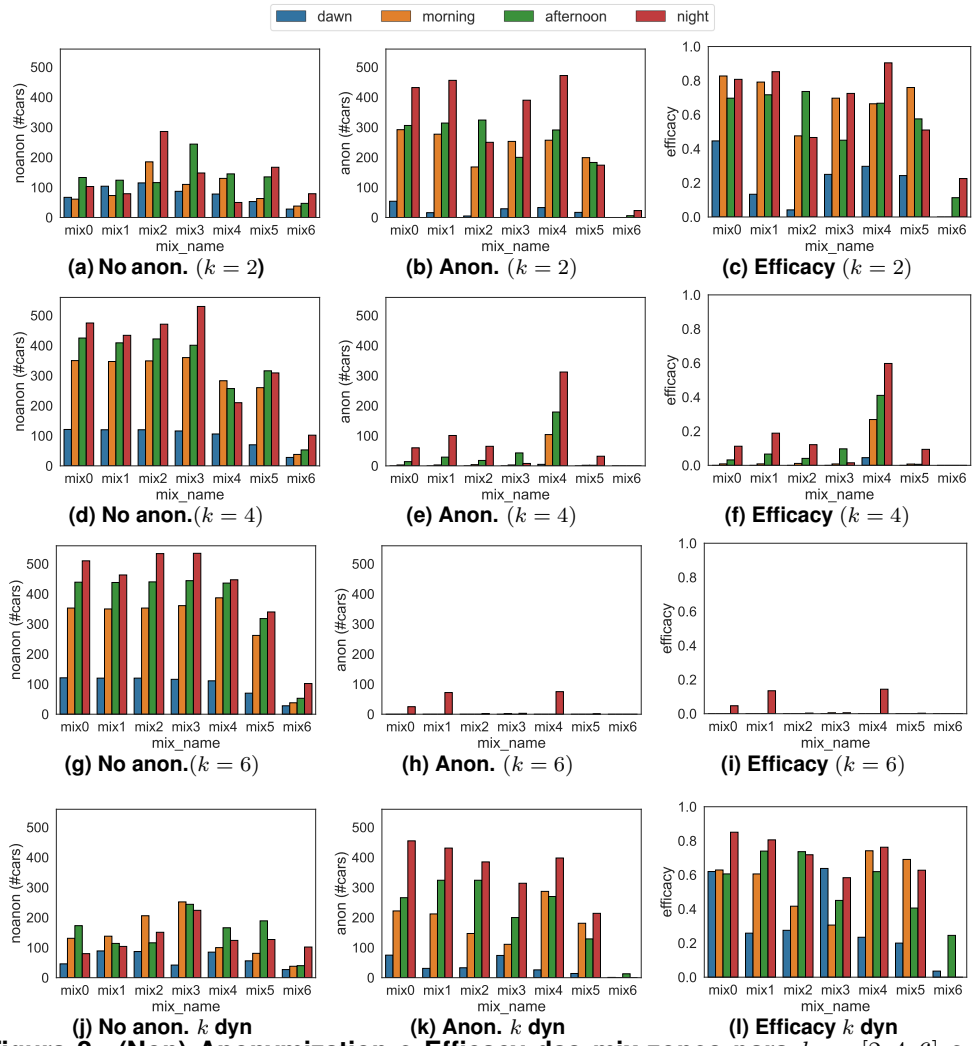


Figura 2. (Non) Anonymization e Efficacy das mix-zones para $k = [2, 4, 6]$ e k-DynMix posicionadas com FPMT.

para cada mix-zone (ou seja, o NCM). Em relação ao CV, as mix6, mix2 e mix1 tiveram maior variação de NCM em relação à média central do que as demais, destacando a mix6 com o CV de 93,11. Essas mix-zones tiveram uma oscilação maior na densidade de veículos ao longo do tempo, o que pode impactar negativamente as métricas AR e ME.

A Tabela 2 mostra o total de viagens anonimizadas, não anonimizadas e a eficácia geral de mix-zones para cada configuração. Para mix-zones estáticas, k é inversamente proporcional à taxa de anonimização e à eficácia geral. Ou seja, à medida que k cresce, o conjunto de dados protegidos torna-se mais vulnerável a ataques de rastreamento. A melhor eficácia foram as mix-zones $k = 2$ com 64%, mas a privacidade foi baixa. Já o k-DynMix, teve uma taxa de anonimização igual a 5.421 viagens e uma eficácia geral de 60,7%, próximo a $k = 2$, mas com níveis de privacidade superior.

Para a análise das mix-zones estáticas e o k-DynMix ao longo do tempo, definimos uma janela de tempo W de quatro períodos [de Mattos et al. 2023]: madrugada (*dawn*), manhã (*morning*), tarde (*afternoon*) e noite (*night*). No geral, foi à noite e ao amanhecer que as mix-zones tiveram AR máximo e mínimo, respectivamente. Particularmente, à noite, a mix4 – implantada no Aeroporto de São Francisco – obteve o maior AR e ME do que outras mix-zones para todas as configurações, indicando um alto tráfego (um alto

NCM). Já as configurações $k = 4$ e $k = 6$ tiveram um NCM insuficiente para ativar as mix-zones, ocasionando um NAR alto e ME próximo a zero. Em uma comparação com configurações estáticas, o ME do k-DynMix teve um desempenho próximo à configuração $k = 2$ e melhor que $k = 4$ e $k = 6$, mas com a possibilidade de ter maior privacidade do que $k = 2$. Por exemplo, à noite para as mix0, mix1 e mix4, os ME do k-DynMix foram respectivamente 0,85, 0,80 e 0,76 contra 0,11, 0,18 e 0,59 para $k = 4$, e 0,04, 0,13, 0,14 para $k = 6$. Além disso, o k-DynMix melhorou o desempenho de mix-zones com baixo tráfego, como nas mix2 e mix3 para $k = 6$, cujo AR foram de 2 e 5 respectivamente, e com o k-DynMix aumentaram para 998 e 889.

6.4. Mix-zones com *k* Estático vs. *k*-DynMix em termos das métricas de AQ

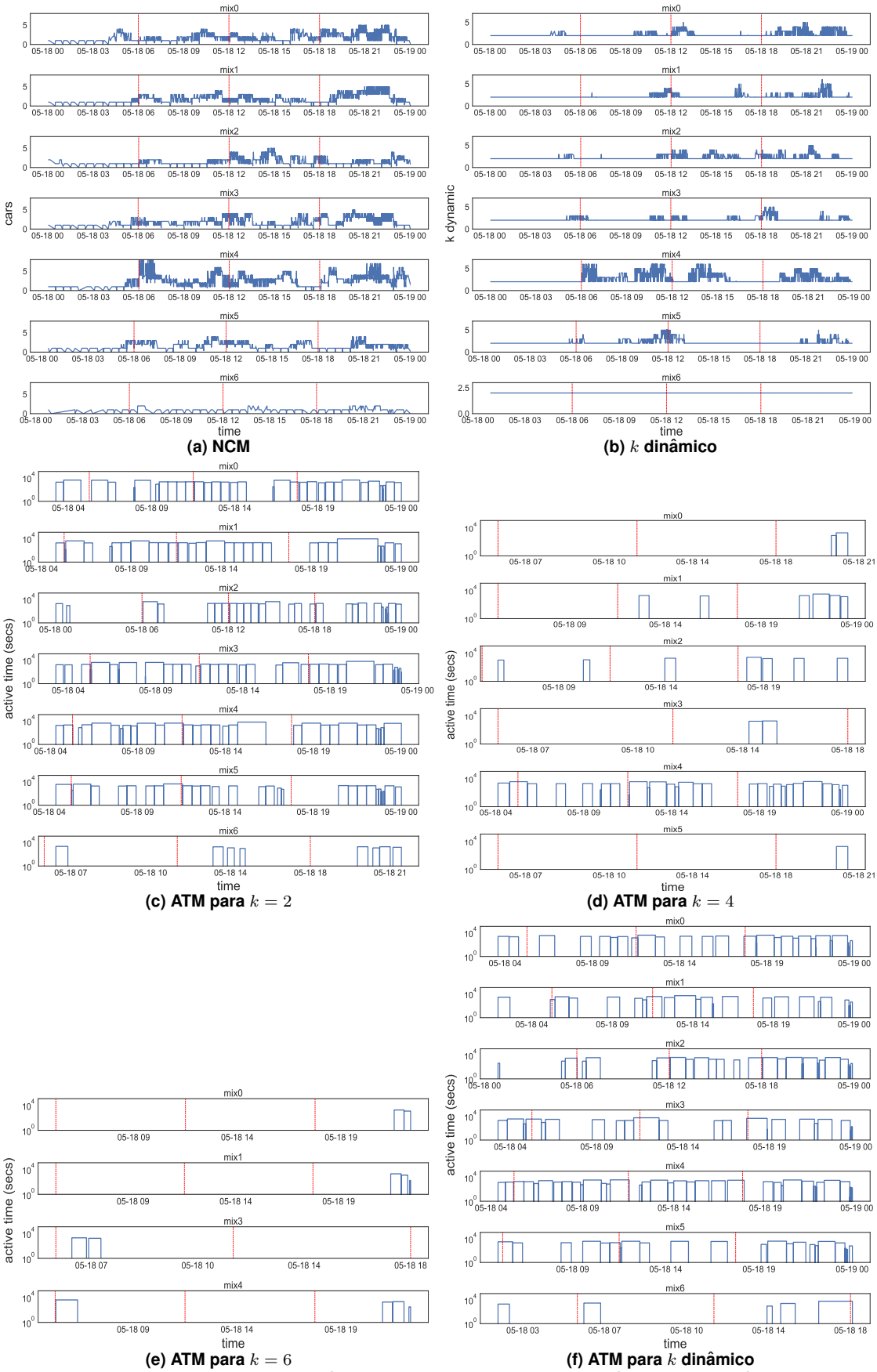
6.4.1. *Number of Cars on Mix-zone* (NCM) e k Dinâmico ao Longo do Tempo

Calculamos o NCM em cada mix-zone para avaliar as tendências do AR ao longo do tempo (ver Fig. 3a). Durante a madrugada, poucos táxis passaram pelas mix-zones, resultando em baixas taxas de AR (Figs. 2b e 2k). O volume de veículos e AR aumentaram gradativamente ao longo do dia. Esta tendência foi destacada para $k = 4$ (ver Figs. 2e e 2k). À noite, as mix0, mix1 e mix2 tiveram os níveis mais altos de tráfego de veículos, com NCM chegando a cinco, tendo picos significativos de anonimização. A mix4 obteve picos de NCM pela manhã e à noite com valores iguais a 8 e 7, respectivamente (ver Fig. 3a). Notavelmente, a mix4 está localizada na região do Aeroporto Internacional de São Francisco, onde permanecem muitos táxis para realizar suas viagens, exibiu um padrão NCM distinto em relação às outras mix-zones. A mix6, localizada na região central de São Francisco, apresentou um menor volume de tráfego, registrando um NCM médio de 2 veículos ao meio-dia, com o mesmo período de pico para o AR.

A Fig. 3b detalha o k ao longo do tempo produzido pelo k-DynMix para as sete mix-zones. Todas as mix-zones tiveram um pico de privacidade superior a 2 das mix-zones mix0 à mix5 obtiveram um k de 5, 6, 5, 5, 6 e 5, respectivamente. Exceto a mix6 que obteve ($k = 2$) para todos os períodos devido ao baixo NCM. Notamos também que a distribuição k segue a variação de NCM ao longo do tempo para cada mix-zone. Por exemplo, o k mínimo ocorre nos períodos do amanhecer, e picos ocorrem durante a manhã, tarde e noite (Fig. 3b). É o caso da mix4, que teve mais NCM pela manhã e tarde com picos de 6, e à noite a com o k igual a 5 (Fig. 3a). As Mix0, mix1 e mix2 tiveram picos de k à noite de 5, 6 e 5. Esta análise sugere que o k dinâmico teve melhor desempenho que o k estático em termos de privacidade.

6.4.2. *Activation Time of the Mix-zone* (ATM) e k Dinâmico

Um k alto significa um alto nível de privacidade, mas anonimizar com um k alto ocorre se e somente se *NCM* $\geq k$, o que implica que as mix-zones estejam ativas (MA). Para AR com um k alto, os ATMs mais longos são melhores que os ATMs mais curtos e adjacentes, que por sua vez são melhores que os ATMs curtos e dispersos [de Mattos et al. 2023]. Nesta análise, podemos notar picos de AR em períodos de ATMs mais longos, como o $k = 2$ à noite para as mix1, mix3 e mix4 que tiveram o ATMs mais longo (ver Fig. 3c) e consequentemente um alto AR (Fig. 2b). Contudo, os ATMs curtos e adjacentes, como os gerados pelo k-DynMix na Fig. 3f, podem ter AR com melhor privacidade do que as abordagens estáticas. Por exemplo, a mix4 com $k = 2$, que tem probabilidade de re-identificar uma trajetória de $1/2$ nos períodos manhã, tarde e noite, tiveram AR de 257 , 291 e 472, respectivamente. Porém, o k-DynMix nos mesmos períodos tiveram viagens



(a) NCM **(b) k dinâmico**

(c) ATM para $k = 2$ **(d) ATM para $k = 4$**

(e) ATM para $k = 6$ **(f) ATM para k dinâmico**

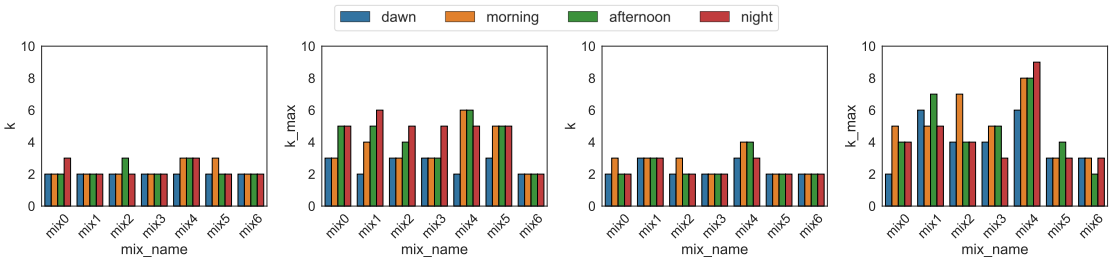
Figura 3. NCM, ATM, e Análise de k para mix-zones estáticas e k dinâmico.

anonimizadas com picos de k de até 6, ou seja, a probabilidade de re-identificação cerca de $1/6$, e AR próximos ao de $k = 2$ iguais a 287, 270 e 398. O mesmo comportamento tiveram as mix0, mix1 e mix2, com os ATMs curtos e adjacentes à tarde e à noite. Nesta análise o k-DynMix também obteve um melhor desempenho que $k = 4$ e $k = 6$, porque algumas mix-zones não produziram anonimização para essas configurações, como é o caso da mix6 com $k = 4$ e as mix2, mix5 e mix6 com $k = 6$ (ver Figs. 3d e 3e).

6.4.3. k Médio e Máximo por Período

Nesta análise exploramos a capacidade do k-DynMix de estimar o valor de k. Analisamos o valor médio e máximo de k por período com dois cenários: o primeiro com menor fluxo de veículos corresponde a amostras de trajetórias D_a de domingo, 18/05/2008; e o segundo com maior fluxo de veículos corresponde a amostras de trajetórias D_b de segunda-feira, 19/05/2008 (ver subseção 6.1 para detalhes sobre D_a e D_b).

Para D_a, o k-DynMix produziu k médio igual a 3 por pelo menos um período para as mix-zones mix0, mix2, mix4 e mix5 (ver Fig. 4a). Destacamos a mix4, em que teve k médio de 3 para três períodos. Para a amostragem D_b, o k médio foi igual a 3 para mix0, mix1, mix2 e mix4. Destacamos a mix4, que obteve um k médio de 4 para os períodos da manhã e tarde (ver Fig. 4c). Analisamos também a anonimização em relação ao valor máximo de k estimado por período. Para D_a, das sete mix-zones, seis delas sendo as mix0, mix1, mix2 mix3, mix5 e mix4 atingiram picos iguais ou superiores a 5 nos períodos tarde e noite (Fig. 4b); destacando a mix4 que obteve k máximo igual a 6 durante a tarde e a noite. Para D_b, todas as mix-zones tiveram picos de k iguais ou superiores a 3; e cinco mix-zones tiveram um k máximo iguais ou superiores a 4 (Fig. 4d). Mais uma vez, a mix4 teve k máximo igual ou superior a 8 nos períodos manhã, tarde e noite.



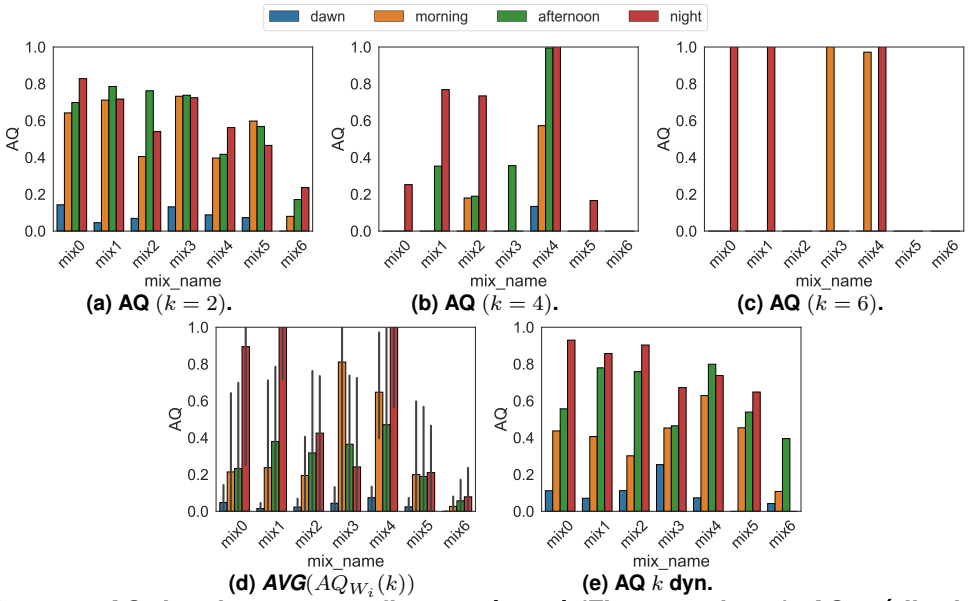
(a) avg(k), dia 18^o (b) max(k), dia 18^o (c) avg(k), dia 19^o (d) max(k), dia 19^o

Figura 4. k médio e máximo por período calculado pelo k-DynMix. Figs. 4a e 4b: Amostragem D_a: Domingo, 18 de maio de 2008; - Figs. 4c e 4d: Amostragem D_b: segunda-feira, 19 de maio de 2008

6.4.4. *Anonymization Quality* (AQ) e k-Dinâmico

Nas mix-zones estáticas, os maiores valores de AQ para $k = 2$ foram para mix0, mix1 e mix3 que tiveram altas taxas de AR, ME e baixo NAR durante à tarde e à noite (ver Fig. 5a). Este comportamento se reflete nos mesmos períodos para as métricas NCM e ATM, nas quais as mix0 e mix1 tiveram mais destaque que as demais, particularmente a mix1 que obteve NCM e ATM até 5 e 1832 segundos, respectivamente. Conforme o k cresce, as métricas AR e ME decrescem, assim, a mix4 se destaca das demais mix-zones pelo seu alto fluxo de tráfego, resultando em um AQ próximo do valor máximo. Para $k = 6$, a mix4 tem AQ de 0,97 e 1 de manhã e à noite, porém o AQ é zero em todos os períodos para as mix2, mix5 e mix6 (Fig. 5c).

Diferente das configurações estáticas, o k-DynMix obteve AQ para todas as mix-zones. Particularmente, na mix4 o k-DynMix teve AQ maior que $k = 2$ com AQ de 0,62,



(a) AQ $(k = 2)$**.** **(b) AQ** $(k = 4)$**.** **(c) AQ** $(k = 6)$**.**

(d) ***AVG***$(AQ_{W_i}(k))$ **(e) AQ** k **dyn.**

Figura 5. AQ de mix-zones config. $k = [2, 4, 6]$ **(Figs. 5a, 5b, 5c), AQ médio dos níveis de privacidade por período (Fig. 5d) e o** k**-DynMix (Fig. 5e).**

0,80 e 0,73 para os períodos manhã, tarde e noite, enquanto que $k = 2$ obteve o AQ de 0,39, 0,41, e 0,56 para estes períodos. A variação do k ao longo do tempo foi o principal fator para este desempenho, atingindo uma média do k igual a três nestes períodos (ver Fig. 4a). Além disso, o AQ do k-DynMix foi superior ao de $k = 2$ para as mix0, mix1, mix2 e mix4 no período noturno. O k-DynMix obteve o AQ igual a 0,93, 0,85, 0,90 e 0,73 contra $k = 2$ com AQ de 0,82, 0,71, 0,54 e 0,56 para estas mix-zones (ver Figs. 5e e 5a).

O k-DynMix também apresentou um desempenho superior às mix-zones estáticas em relação a análise entre o AQ médio por período W_i dos níveis de privacidade $k = \{2, 4, 6\}$, $AVG(AQ_{W_i}(k))$, e o AQ estimado pelo k-DynMix (ver Figs. 5d e 5e). Por exemplo, o $AVG(AQ_{W_i}(k))$ da mix0 nos períodos da madrugada, manhã, tarde e noite foram 0,048, 0,21, 0,23 e 0,89, respectivamente, e o AQ do k-DynMix para a mesma foram 0,11, 0,43, 0,55 e 0,93. Podemos observar este mesmo comportamento para as demais mix-zones, inclusive para a mix6 que é caracterizada pelo seu baixo tráfego.

7. Conclusão

Neste trabalho, propusemos o k-DynMix, um esquema de mix-zone dinâmica que ajusta o nível de privacidade k ao longo do tempo em modo online, com complexidade linear, de acordo com eventos como flutuações de tráfego de veículos para obter maior anonimização. O k-DynMix é inspirado nos conceitos de *Anonymization Quality* e controle de congestionamento do TCP. Conduzimos experimentos, analisando conjuntos de dados reais e sintéticos comparando o k-DynMix com dois modelos de previsão para estimar a privacidade ao longo do tempo e com mix-zones clássicas em relação às métricas de cobertura e AQ. Os resultados mostraram que k-DynMix superou os modelos de previsão em estimar a privacidade. Além de obter a eficácia, a taxa de anonimização e a AQ semelhantes ao melhor resultado das mix-zones clássicas, o k-DynMix também maximizou a privacidade ao melhor possível, inclusive para as mix-zones de baixo tráfego, superando as mix-zones clássicas. Por fim, o k-DynMix obteve um AQ para todas as mix-zones, diferente das mix-zones clássicas. Até onde sabemos, esta é o primeira esquema de privacidade dinâmica ao longo do tempo que considera as flutuações do tráfego de veículos para maximizar a privacidade, a eficácia e a qualidade de anonimização em mix-zones.

A. Privacy Level Prediction Approaches

O *Simple Moving Average (SMA)* é um método amplamente utilizado para calcular a média dos n pontos de dados anteriores em um conjunto de dados de série temporal. Cada ponto tem igual importância em SMA sem quaisquer fatores de ponderação aplicados, garantindo um cálculo imparcial. O SMA é definido como $SMA = \frac{P_t + P_{t-1} + \cdots + P_{t-(n-1)}}{n}$, onde P_t é o ponto no tempo t e n representa o número de pontos de dados usados no cálculo.

O *Weighted Exponential Moving Average (WEMA)* é uma variação da Média Móvel Exponencial (EMA) que atribui vários níveis de importância aos pontos de dados. Ao contrário da EMA, que trata todos os pontos de dados igualmente, o WEMA permite esquemas de ponderação personalizados através de funções de ponderação específicas. Isto permite a ênfase de pontos de dados recentes ou significativos em detrimento de outros. O WEMA é denotado por $WEMA_t = \alpha.P_t + (1 - \alpha).WMA_t$, onde P_t é o valor no período t, α representa o grau de diminuição da ponderação como na equação $\alpha = \frac{2}{(n+1)}$, e WMA_t é a média móvel ponderada (WMA) no momento t [Hansun 2013].

Referências

Beresford, A. R. and Stajano, F. (2003). Location privacy in pervasive computing. *IEEE Pervasive computing*, (1):46–55.

Beresford, A. R. and Stajano, F. (2004). Mix zones: User privacy in location-aware services. In *Pervasive Computing and Communications Workshops, 2004. Proceedings of the Second IEEE Annual Conference on*, pages 127–131. IEEE.

Chen, Z., Fu, Y., Zhang, M., Zhang, Z., and Li, H. (2018). A flexible mix-zone selection scheme towards trajectory privacy protection. In *17th IEEE TrustCom*, pages 1180–1186.

Chow, C.-Y. and Mokbel, M. F. (2011). Trajectory privacy in location-based services and data publication. *ACM Sigkdd Explorations Newsletter*, 13(1):19–29.

de Mattos, E. P., Domingues, A. C., Santos, B. P., Ramos, H. S., and Loureiro, A. A. (2022). The Impact of Mobility on Location Privacy: A Perspective on Smart Mobility. *IEEE Systems Journal*.

de Mattos, E. P., Domingues, A. C., Silva, F. A., Ramos, H. S., and Loureiro, A. A. (2023). Slicing who slices: Anonymization quality evaluation on deployment, privacy, and utility in mix-zones. *Computer Networks*, 236:110007.

de Mattos, E. P., Domingues, A. C., Silva, F. A., Ramos, H. S., and Loureiro, A. A. (2024). Protect your data and I'll rank its utility: A framework for utility analysis of anonymized mobility data for smart city applications. *Ad Hoc Networks*, page 103567.

Domingues, A. C., de Mattos, E. P., Silva, F. A., Ramos, H. S., and Loureiro, A. A. (2021). Social Mix-zones: Anonymizing Personal Information on Contact Tracing Data. In *Proceedings of the 18th ACM Symposium on Performance Evaluation of Wireless Ad Hoc, Sensor, & Ubiquitous Networks*, pages 81–88.

Freudiger, J., Raya, M., Félegyházi, M., Papadimitratos, P., and Hubaux, J.-P. (2007). Mixzones for location privacy in vehicular networks. In *Workshop on Wireless Networking for Intelligent Transportation Systems (WiN-ITS)*.

Hansun, S. (2013). A new approach of moving average method in time series analysis. In *2013 conference on new media studies (CoNMedia)*, pages 1–4. IEEE.

Khodaei, M. and Papadimitratos, P. (2020). Cooperative location privacy in vehicular networks: why simple mix zones are not enough. *IEEE Internet of Things Journal*, 8(10):7985–8004.

Mattos, E. P., Domingues, A. C., and Loureiro, A. A. F. (2019). Give Me Two Points and I'll Tell You Who You Are. In *Proceedings of the IEEE Intelligent Vehicles Symposium (IV'19)*. IEEE.

Mattos, E. P. d., Domingues, A. C., Silva, F. A., Ramos, H. S., and Loureiro, A. A. (2022). Behind the Mix-Zones Scenes: On the Evaluation of the Anonymization Quality. In *Proceedings of the 19th ACM International Symposium on Performance Evaluation of Wireless Ad Hoc, Sensor, & Ubiquitous Networks*, pages 133–140.

Mattos, E. P. d., Domingues, A. C., Silva, F. A., Ramos, H. S., and Loureiro, A. A. (2023). Protect your Data and I'll Show Its Utility: A Practical View about Mix-zones Impacts on Mobility Data for Smart City Applications. In *Proceedings of the Int'l ACM Symposium on Performance Evaluation of Wireless Ad Hoc, Sensor, & Ubiquitous Networks*, pages 45–52.

Paiva, S., Ahad, M. A., Zafar, S., Tripathi, G., Khalique, A., and Hussain, I. (2020). Privacy and security challenges in smart and sustainable mobility. *SN Applied Sciences*, 2:1–10.

Palanisamy, B. and Liu, L. (2014). Attack-resilient mix-zones over road networks: architecture and algorithms. *IEEE TMC*, 14(3):495–508.

Paxson, V., Allman, M., Chu, J., and Sargent, M. (2011). Computing TCP's retransmission timer. Technical report.

Piorkowski, M., Sarafijanovic-Djukic, N., and Grossglauser, M. (2009). CRAWDAD dataset epfl/mobility (v. 2009-02-24). Downloaded from `https://crawdad.org/epfl/mobility/20090224`.

Yamazaki, R., Yoshida, M., and Shigeno, H. (2021). A Dynamic Mix-zone Scheme Considering Communication Delay for Location Privacy in Vehicular Networks. In *2021 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*, pages 245–250. IEEE.

Zhou, Y. and Zhang, D. (2019). Double Mix-Zone for Location Privacy in VANET. In *7th Int'l Conf. on Info. Tech.: IoT and Smart City*, pages 322–327.