PrivLBS: Uma Abordagem para Preservação de Privacidade de Dados em Serviços baseados em Localização

Eduardo R. D. Neto[1], André L. C. Mendonça[1], Felipe T. Brito[1], Javam C. Machado[1]

[1]Laboratório de Sistemas e Banco de Dados (LSBD)
DC/UFC – UFC – CEP 60440-900 – Fortaleza – CE – Brazil

{eduardo.rodrigues,andre.luis,felipe.timbo,javam.machado}@lsbd.ufc.br

Abstract. *Location based services have been increasingly integrated into people's daily activities. However, some of these services may not be trustworthy and lead to serious privacy breaches. This work proposes a new technique for privacy preserving data, named PrivLBS, which ensures that individual's location will not be easily re-identified by malicious services. Experimental results show that, for euclidean distance-based attacks, individual's probability of location re-identification, after using PrivLBS, is around 11.4%, whereas in existing work, this probability reaches 59.2%.*

Resumo. *Serviços baseados em localização têm sido integrados às atividades diárias das pessoas. Entretanto, alguns desses serviços podem não ser confiáveis e levar a sérios riscos de violação de privacidade. Este trabalho propõe uma nova técnica de preservação de privacidade de dados, denominada PrivLBS, capaz de assegurar que as localizações dos indivíduos não serão facilmente reidentificadas por serviços mal intencionados. Resultados de avaliação experimental demonstram que, para ataques baseados em distância euclidiana, a probabilidade de reidentificação das localizações de um indivíduo, após utilização do PrivLBS, é em torno de 11.4%, enquanto que, em trabalhos já existentes na literatura, essa probabilidade chega a 59.2%.*

1. Introdução

Serviços baseados em localização (*Location-Based Services, LBS*) são serviços que proveem recursos adicionais a dispositivos móveis baseado em suas localizações geográficas. Esses serviços têm sido integrados às atividades diárias das pessoas, permitindo que elas utilizem sua localização atual para diversos fins, tais como navegação, rastreamento, recomendação, entre outros. Em geral, para que serviços baseados em localização sejam utilizados, os usuários enviam ao provedor de serviço (provedor de LBS) sua identidade e localização geográfica real, definida pela latitude e longitude, além de consultas que se desejam obter respostas, como o shopping mais próximo, supermercado, restaurante [Niu et al. 2014]. Dessa forma, os usuários obtém os locais relativos à consulta realizada.

Por outro lado, a utilização de serviços baseados em localização pode levar a sérios riscos de violação de privacidade devido a provedores de serviços mal intencionados ou não confiáveis [Li et al. 2014, Niu et al. 2015]. Provedores de LBS não confiáveis são capazes de expor dados de localização de seus usuários ou até mesmo vender informações de localizações a terceiros [Zhu et al. 2013]. De posse dessas informações, os dados obtidos por terceiros são utilizados para descoberta de padrões de movimento do usuário,

podendo revelar informações sensíveis sobre ele. Por exemplo, se um usuário, ao utilizar um serviço baseado em localização, geralmente exibe sua localização próximo a um hospital, as informações de localização poderiam ser utilizadas para inferir que aquele usuário provavelmente possa ter algum problema de saúde.

Para que seja mantida a privacidade dos usuários na utilização desses serviços, várias técnicas de preservação de privacidade em LBS foram propostas nos últimos anos [Niu et al. 2016, Tsoukaneri et al. 2016, Ullah and Shah 2016, Sun et al. 2017b]. Algumas dessas técnicas são baseadas em métodos de camuflagem, os quais empregam o modelo de privacidade k-anonimato [Sweeney 2002] para proteger a privacidade dos locais percorridos por um usuário. Este modelo garante que um usuário só poderá ser reidentificado com probabilidade $\frac{1}{k}$, onde k é o grau de privacidade especificado pelo usuário. Quanto maior o valor de k, menor a probabilidade de reidentificação das localizações de um indivíduo.

Uma forma de camuflar as localizações de um usuário, utilizando o modelo de privacidade k-anonimato, é por meio da técnica de *"dummy locations"* [Kido et al. 2005]. Nessa abordagem, $k-1$ localizações falsas são geradas e adicionadas à consulta realizada pelo usuário ao provedor do LBS, a fim de confundir a localização real do indivíduo que realizou a consulta. Por exemplo, no momento em que um usuário desejar obter o shopping mais próximo de sua localização atual, ao especificar o valor de k, outras $k-1$ localizações falsas serão geradas e enviadas ao provedor de serviço. O provedor retornará ao usuário os shoppings mais próximos para cada umas das $k-1$ localizações falsas, como também o shopping mais próximo da localização real do usuário. Contudo, trabalhos existentes na literatura [Kido et al. 2005, Vu et al. 2012, Niu et al. 2014, Sun et al. 2017a] não levam em consideração qualquer métrica de distância física das localizações no momento da geração, o que as tornam vulneráveis a ataques que exploram essa deficiência. Assim, localizações falsas geradas podem não ser coerentes com a distância percorrida pelo usuário durante o intervalo de realização de duas consultas consecutivas.

Assumindo que o provedor do LBS não é confiável, neste trabalho propomos uma nova técnica baseada no modelo de privacidade k-anonimato, denominada PrivLBS, capaz de assegurar que as localizações dos indivíduos que utilizam serviços baseado em localização não serão facilmente reidentificadas. Para isso, propomos um novo tipo de ataque baseado em distância que busca revelar a localização real do usuário, considerando a distância euclidiana entre as localizações de consultas consecutivas enviadas ao provedor de serviço. Demonstramos, através de simulações, que nosso modelo de ataque possui uma alta taxa de reidentificação das localizações reais dos usuários quando aplicado sobre a estratégia DLP (*Dummy Location Privacy-preserving*) [Sun et al. 2017a], proposta recentemente na literatura. Por outro lado, PrivLBS assegura que provedores de serviços não confiáveis, que utilizam ataques baseado em distância, não são capazes de violar a privacidade dos usuários com probabilidade média maior que $\frac{1}{k}$, onde k é o grau de privacidade.

O restante do artigo está organizado da seguinte forma. A Seção 2 apresenta os trabalhos relacionados ao tema de preservação de privacidade em serviços baseados em localização. Na Seção 3 apresentamos o nosso modelo de ataque baseado em distância euclidiana. Em sequência, na Seção 4, apresentamos o método PrivLBS como solução para o problema e, em seguida, o avaliamos experimentalmente na Seção 5 utilizando um

conjunto de dados real. Por fim, a Seção 6 conclui o trabalho e apresenta os direcionamentos futuros de pesquisa.

2. Trabalhos Relacionados

Diversas soluções foram propostas com o objetivo de garantir a privacidade de usuários ao utilizarem serviços baseados em localização e, assim, impedir que suas informações sensíveis sejam descobertas. Em sua grande maioria, as soluções são divididas em abordagens baseadas em anonimização de localizações [Gedik and Liu 2008, Ying and Makrakis 2014], criptografia [Lu et al. 2014] ou seleção de *dummy locations* [Niu et al. 2014, Niu et al. 2015, Sun et al. 2017a].

O trabalho em [Gedik and Liu 2008] propõe um modelo personalizado do *k*-anonimato utilizando a estratégia de camuflagem. Nesse trabalho, os autores utilizaram um servidor de anonimização confiável que considera o *trade-off* entre a privacidade da localização e a qualidade do serviço para anonimizar a localização dos usuários. Na solução, uma região de camuflagem contendo outros $k-1$ usuários, geograficamente distribuídos, é formada e, somente então, a consulta é submetida ao serviço baseado em localização. Também utilizando a estratégia de camuflagem, o trabalho em [Ying and Makrakis 2014] assegura a privacidade dos usuários ao construir uma região de camuflagem contendo, pelo menos, *k* usuários e *l* segmentos de rua.

O trabalho proposto em [Lu et al. 2014] apresenta um *framework*, denominado PLAM, para a preservação de privacidade em redes sociais de área local. Esse *framework*, além de atender ao modelo de privacidade *k*-anonimato, também assegura o modelo *l*-diversidade [Machanavajjhala et al. 2006], considerando casos em que um adversário pode inferir informações sensíveis sobre indivíduos mesmo sem identificá-los. Entretanto, o servidor de anonimização confiável é substituído por uma técnica de criptografia, denominada pseudo-ID, a qual não mantém a utilidade dos dados para fins de análise.

[Niu et al. 2014] propõem o DLS (*Dummy Location Selection*), um algoritmo de seleção de *dummy locations* baseado em entropia, o qual mede o grau de incerteza sobre um conjunto de localizações selecionadas. Nesse trabalho, os autores apresentaram um modelo de LBS no qual o provedor do serviço é responsável por coletar e disponibilizar aos usuários dados estatísticos sobre as consultas. Tais dados dizem respeito às probabilidades nas quais requisições são demandadas ao LBS. Assim, o DLS assegura a privacidade dos usuários, garantindo as propriedades do modelo *k*-anonimato, ao submeter uma consulta contendo a localização real do usuário e de outras $k-1$ localizações falsas escolhidas utilizando como critério de seleção localizações que tenham uma probabilidade de ser enviada ao LBS semelhante a da localização real.

Por fim, o trabalho em [Sun et al. 2017a] propõe o algoritmo DLP, que assim como o DLS utiliza a técnica de *dummy locations* e a probabilidade das localizações sobre as consultas feita ao LBS como critério de seleção, porém alcançando um grau de entropia superior aos trabalhos anteriores, isto é, uma maior incerteza sobre um conjunto de localizações selecionadas. Os autores propõem um algoritmo de ataque desenvolvido especificamente para revelar a localização real do usuário quando a anonimização utiliza como critério de seleção das $k-1$ localizações falsas a probabilidade destas nas consultas enviadas e coletadas pelo LBS.

Ao contrário das soluções anteriores, este artigo propõe uma técnica baseada na

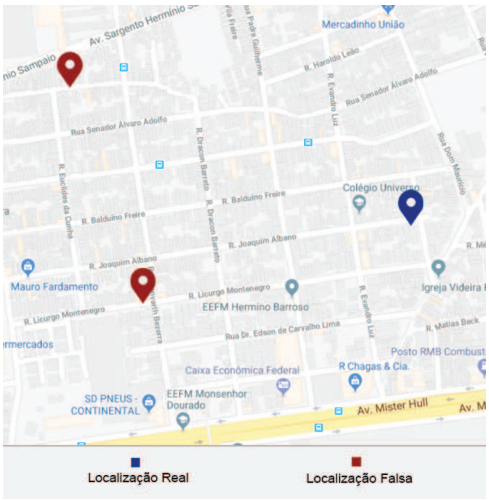
seleção de *dummy locations* cujas localizações falsas são selecionadas utilizando critérios de distância euclidiana e probabilidade de suas ocorrências com base em informações coletadas pelo LBS, garantindo, assim, uma maior privacidade aos usuários contra ataques que explorem esses critérios sobre as consultas enviadas ao LBS.

3. Ataque baseado em Distância Euclidiana

Quando lidamos com serviços baseados em localização, pode-se realizar dois tipos de requisições (consultas) ao provedor de serviço: consultas simples e contínuas. Uma consulta simples consiste em uma requisição realizada pelo usuário antes mesmo dele obter um novo identificador, por exemplo, quando um usuário solicita o shopping mais próximo da localização informada e, após receber o conteúdo requisitado, encerra a conexão com o LBS. Caso seja realizada uma nova requisição ao LBS, o cliente já é visto como um novo usuário. Consultas contínuas tratam-se de múltiplas consultas realizadas por um usuário em um determinado intervalo de tempo por meio de um mesmo identificador. Por exemplo, quando um usuário solicita o tempo estimado para se chegar a um destino, várias vezes em um determinado intervalo de tempo, até que o mesmo encerre a requisição. Para qualquer tipo de consulta, o LBS recebe a requisição e retorna a informação requerida de acordo com seu conteúdo. Este trabalho visa preservar a privacidade de indivíduos que realizam tanto consultas simples quanto consultas contínuas a provedores de LBS.

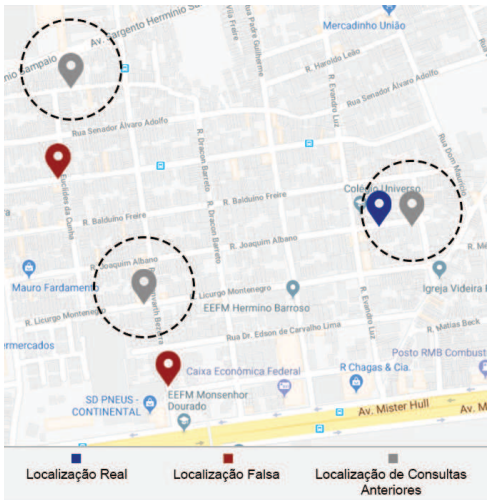
Para demonstrar a eficiência do PrivLBS em relação aos modelos existentes de geração de localizações falsas, propomos um algoritmo de ataque baseado em distância que visa revelar a localização real do usuário utilizando a métrica de distância euclidiana entre as localizações de duas consultas consecutivas enviadas pelo mesmo usuário. Vale ressaltar que o algoritmo de ataque proposto pode ser adaptado para utilizar qualquer tipo de função de distância, não apenas a euclidiana.

Utilizaremos a Figura 1 para ilustrar como o ataque é realizado sobre uma nova requisição feita ao LBS quando a anonimização não leva em consideração a distância euclidiana entre os pontos no momento da escolha de suas localizações falsas. Os pontos em azul e vermelho representam as localizações reais e falsas, respectivamente. Os pontos em cinza são as localizações da última consulta enviada ao LBS. Na Figura 1(a) observamos a requisição no primeiro momento t_0 anonimizada com grau de privacidade $k = 3$, escolhido especificamente para simplificar o exemplo. Já a Figura 1(b) apresenta o momento em que o usuário, após um intervalo de tempo t, realiza a consulta seguinte em uma nova localização. A circunferência ao redor dos pontos em cinza representam a área contendo todos os pontos alcançáveis a partir dele. A Figura 1(c) exibe as localizações selecionadas no processo de anonimização e enviadas na requisição ao LBS pelo usuário no tempo t_1. O algoritmo de ataque, tendo obtido o domínio da consulta anterior enviada ao LBS, verifica quais pontos da nova consulta estão dentro de uma das áreas de alcance dos pontos da consulta anterior. As localizações da nova consulta que estiverem dentro dessas áreas são as localizações candidatas, visto que as outras localizações devem ser ignoradas por não serem alcançáveis pelo usuário no intervalo de tempo de realização de consultas consecutivas. O algoritmo de ataque seleciona como localização real uma das localizações dentre as candidatas. Podemos observar pela Figura 1(c) que apenas um ponto da nova consulta está dentro de uma dessas áreas, sendo assim identificada pelo algoritmo de ataque como a localização real do usuário.



(a) Requisição em t_0.

(b) Nova localização em t_1.



(c) Requisição anonimizada em t_1.

Figura 1. Anonimização de localizações sem critério de distância.

O Algoritmo 1 refere-se à nossa proposta de ataque, que possui como entrada os parâmetros R', R, denotando respectivamente o conjunto das localizações contidas na requisição anterior e atual. O parâmetro P, contendo a lista de localizações atendidas pelo LBS e suas respectivas probabilidades. Além disso, o algoritmo tem como parâmetro de entrada um $limite$, estabelecido pelo atacante, que representa a distância máxima permitida entre as localizações. Esse parâmetro é calculado pela função $limite = v * t$, onde v é a velocidade média do usuário estimada pelo LBS, e t é o tempo decorrido entre uma requisição e outra.

O algoritmo atua da seguinte forma: para cada localização r_i da nova requisição R, o algoritmo calcula a distância euclidiana entre r_i e cada uma das localizações r'_j da requisição anterior R'. Se essa distância for menor ou igual ao $limite$, então a localização r_i é adicionado ao conjunto das localizações candidatas C. Quanto mais preciso o $limite$, mais eficaz é o algoritmo, visto que ele define os elementos do conjunto das localizações candidatas C à localização real. Um $limite$ alto implica em um relaxamento da condição

de alcançabilidade de um ponto a outro, aumentando a probabilidade de localizações que não são realmente alcançáveis serem adicionadas ao conjunto C e, portanto, diminuindo a precisão do algoritmo. De maneira análoga, um limite baixo implica em uma restrição maior na escolha dos elementos do conjunto C, o que leva a um conjunto com poucos elementos. Por fim, o algoritmo de ataque proposto retorna, como localização real, a localização l com maior probabilidade, conforme P, dentre aquelas do conjunto C.

Algoritmo 1: ATAQUE BASEADO EM DISTÂNCIA EUCLIDIANA

Entrada: R' , R, $limite$, P
Saída: l
para cada *localização* $r_i \in R$ **faça**
  para cada *localização* $r'_j \in R'$ **faça**
    se $Distância(r_i, r'_j) <= limite$ **então**
      Insere r_i *em* C;
    fim
  fim
fim
$l = maxProb(r \in C)$;
retorna l

4. PrivLBS

Para contornar o problema da preservação de privacidade de dados de um usuário, que utiliza um serviço baseado em localização, adotamos um modelo semelhante ao proposto em [Sun et al. 2017a]. Primeiramente, o LBS é responsável por coletar, para cada localização l_i, a probabilidade q_i de uma consulta sobre ela. Tal probabilidade é definida pela Equação 1, denominada informação complementar (*side information*).

$$q_i = \frac{\textit{número de consultas sobre } l_i}{\textit{número total de consultas}} \tag{1}$$

A Figura 2 ilustra o fluxo da abordagem proposta. Inicialmente, o usuário inicia a sessão requisitando ao LBS a informação complementar coletada. Após obter essa informação, a anonimização da consulta é realizada utilizando o Algoritmo 2, que seleciona $k - 1$ localizações falsas a serem adicionadas à consulta. Dessa forma, o provedor do LBS irá responder conforme o conteúdo da requisição. Por fim, o usuário filtra aquela informação que é de seu interesse.

Detalhando o processo de anonimização. PrivLBS recebe como parâmetros de entrada o grau de privacidade k, a informação complementar P, contendo a lista de localizações atendidas pelo LBS e suas respectivas probabilidades, a localização real l_r e a última requisição enviada R'. O usuário armazena, em seu histórico, a última requisição enviada ao LBS. Caso o histórico do usuário esteja limpo, isto é, o usuário está fazendo a sua primeira ou única consulta, o parâmetro R' será nulo. Neste caso, a seleção das localizações falsas é feita utilizando o próprio algoritmo DLP, proposto por [Sun et al. 2017a]. Nesta situação o algoritmo de ataque baseado em distância não é aplicável, já que não há uma consulta anterior. Caso R' não seja nulo, para cada

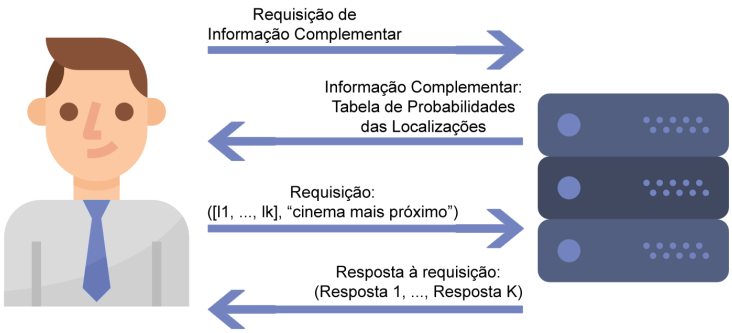


Figura 2. Fluxo de informações do PrivLBS.

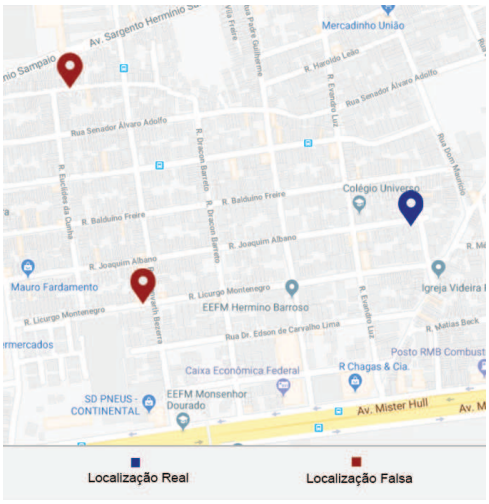
localização falsa $r'_i \in R'$, é construído um conjunto A_i, contendo todas as localizações alcançáveis a partir de r'_i, obtidas através da função *BuscarDistância(P,r'_i)*. Essa função calcula a distância euclidiana entre as localizações e, caso a distância seja menor que a distância máxima possível de ser percorrida pelo usuário, ela é adicionada ao conjunto. De cada conjunto A_i é selecionada a localização cuja probabilidade mais se aproxima da localização real l_r, adicionando-as ao conjunto L, formando $k-1$ localizações falsas. Tais localizações são adicionadas à localização real l_r em L e enviadas ao servidor do LBS.

Algoritmo 2: PRIVLBS

Entrada: k, P, l_r, R'
Saída: L
1 **se** *R' == vazio* **então**
2 $L \leftarrow DLP(k, l_r)$;
3 **senão**
4 **para cada** $r'_i \in R'$ **faça**
5 $A_i \leftarrow BuscarDistância(P, r'_i)$;
6 *Insere em L a localização contida em* A_i *cuja probabilidade seja a mais próxima de* l_r;
7 **fim**
8 *Insere* l_r *em L*
9 **fim**
10 **retorna** L

A Figura 3 ilustra como funciona o algoritmo PrivLBS. Novamente, os pontos em azul e vermelho representam as localizações reais e falsas, respectivamente. Os pontos em cinza são as localizações das consultas anteriores. A Figura 3(a) representa o momento inicial, onde o usuário realiza a primeira consulta anonimizada com grau de privacidade $k = 3$. A Figura 3(b) apresenta o momento seguinte, onde o usuário se desloca para uma nova posição após um intervalo de tempo t e realiza uma nova consulta. As circunferências ao redor dos pontos em cinza representam as áreas contendo todos os pontos alcançáveis a partir dos vértices e possíveis candidatos a serem selecionados pelo PrivLBS como localizações falsas. A Figura 3(c) mostra os pontos selecionados pelo algoritmo PrivLBS que irão fazer parte da requisição junto à localização real a ser enviada ao LBS. Como o PrivLBS seleciona, para cada localização da consulta anterior, uma localização que seja alcançável por ela, dado a velocidade do usuário e o intervalo

de tempo decorrido entre as consultas, um possível ataque que visa explorar esse critério observa cada uma das localizações na nova consulta como deslocamentos possíveis do usuário. Dessa forma, o atacante não é capaz de reidentificar a localização real com probabilidade superior a $\frac{1}{k}$. Isso garante o modelo de privacidade k-anonimato. Além disso, o algoritmo procura escolher localizações alcançáveis que tenham um probabilidade de consulta ao LBS semelhante à localização real, o que protege também o usuário contra ataques probabilísticos sobre o teor da consulta, isto é, ataques que visam identificar, na consulta, uma localização que tenha uma probabilidade maior que as outras.



(a) Requisição em t_0

(b) Nova localização em t_1

(c) Requisição anonimizada em t_1

Figura 3. Anonimização de localizações utilizando o algoritmo PrivLBS.

5. Experimentos

Foram realizados experimentos a fim de avaliar a eficácia do algoritmo PrivLBS frente a ataques baseados em distância. Nossa análise foi realizada com base na taxa de reconhecimento da localização real quando aplica-se o ataque baseado em distância sobre a consulta. Nós também mensuramos o grau de privacidade da requisição, denotado por sua entropia, que consiste na incerteza de identificação da localização real dentre as localizações

falsas selecionadas [Serjantov and Danezis 2003], independente da distância. Quanto maior a entropia, mais incerta é a informação acerca das localizações.

5.1. Conjunto de dados

Utilizamos um conjunto de dados real disponibilizado pela CTA[1] (*Chicago Transit Authority*), responsável por operar o segundo maior sistema de transporte público dos Estados Unidos, atendendo toda a cidade de Chicago e 35 subúrbios na periferia dessa cidade. Esse conjunto de dados foi escolhido por conter, para cada uma das 11.593 estações de ônibus, além da latitude e longitude, a média de embarque em um dia de semana do mês de Outubro de 2012. Isso nos permitiu estimar a probabilidade de requisições sobre cada uma das estações de ônibus com base na média de embarque, formando assim a informação complementar sobre as localizações utilizada tanto nos algoritmos de anonimização DLP (nosso *baseline*) e PrivLBS, como também no algoritmo de ataque baseado em distância e no algoritmo de ataque baseado em probabilidade, proposto em [Sun et al. 2017a].

5.2. Simulação

Foram simulados dez mil usuários realizando consultas consecutivas ao LBS. Para cada usuário foi selecionada uma posição inicial aleatória do conjunto de dados. Cada usuário realizou três consultas consecutivas utilizando os algoritmos de anonimização PrivLBS e DLP. Cada consulta foi realizada em um momento temporal (e.g. t_0, t_1 e t_2). Entre os momentos t_0 e t_1, t_1 e t_2, foi simulado um deslocamento para alguma localização aleatória que se encontra dentro de um raio de 1 km da localização anterior. Estabeleceu-se esse limite considerando um intervalo de 1 minuto entre uma consulta e outra, e uma velocidade média de 60 km/h do usuário, resultando em um deslocamento de até 1 km. Ao término de cada consulta calculamos a entropia sobre o conjunto de localizações selecionadas e aplicamos o algoritmo de ataque baseado em distância e o algoritmo de ataque baseado na probabilidade de execução das consultas sobre cada localização para simular um ataque do provedor de serviço ao tentar violar a privacidade de localização do usuário.

5.3. Resultados

Para demonstrar o grau de privacidade alcançado pelo PrivLBS, uma série de mil simulações foram realizadas, onde foram medidas a entropia sobre os conjuntos de localizações selecionadas e a probabilidade de reindentificação da localização real do usuário sobre vários graus de privacidade (valores de k).

Entropia	Grau de anonimização k				
	2	**4**	**8**	**16**	**32**
DLP	0,69	1,38	2,07	2,77	3,46
PrivLBS	0,58	1,26	1,92	2,60	3,28

Tabela 1. Comparação da entropia entre os algoritmos DLP e PrivLBS.

A Tabela 1 mostra a entropia média obtida utilizando os algoritmos DLP e PrivLBS na seleção das localizações falsas da consulta, variando o grau de privacidade k. Observa-se um comportamento constante, no qual o DLP apresentou uma entropia um

[1]http://www.transitchicago.com

pouco maior em todos os graus analisados. Isso implica que o DLP é menos suscetível a ataques que exploram a probabilidade das localizações contidas nas consultas realizadas ao provedor do LBS. Este comportamento já era esperado, visto que o PrivLBS constrói um subconjunto, baseado na distância, das localizações disponíveis, diminuindo a probabilidade de selecionar localizações com probabilidade semelhante à localização real.

Apesar disso, quando vamos calcular a taxa de reconhecimento da localização real, obtida utilizando o algoritmo de ataque proposto por [Sun et al. 2017a], percebe-se, conforme Figura 4, que tanto o algoritmo DLP como o algoritmo PrivLBS são robustos para este tipo de ataque, garantindo a propriedade do modelo *k*-anonimato, uma vez que, para qualquer grau de privacidade k no gráfico, a taxa de reconhecimento ficou abaixo de $\frac{1}{k}$.

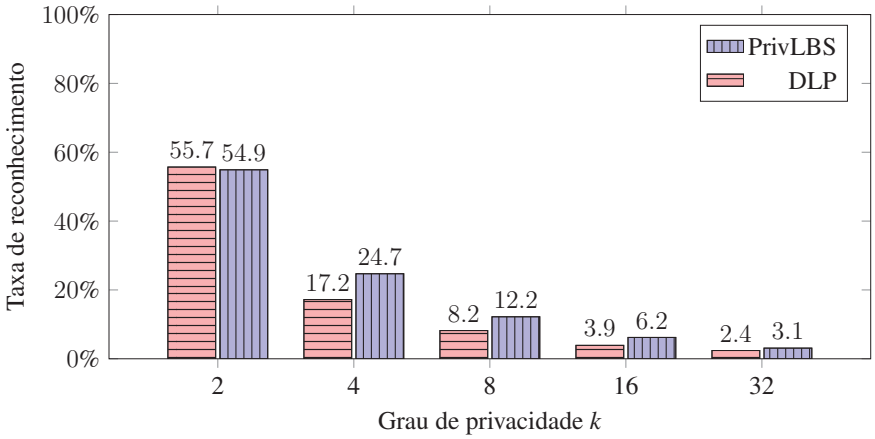


Figura 4. Taxas de reconhecimento da localização real para o ataque baseado na probabilidade de execução das consultas.

Já em relação a taxa de reconhecimento da localização real quando a consulta realizada sobre o provedor do LBS recebe o ataque baseado em distância euclidiana, podemos observar, conforme Figura 5, que quanto maior o grau de privacidade k, menor é a taxa de reconhecimento nas consultas realizadas utilizando o DLP como algoritmo de anonimização. Vale ressaltar que o parâmetro k representa, também, a quantidade de localizações que serão enviadas na consulta, aumentando a chance de mais localizações alcançáveis serem selecionadas como localizações falsas pelo algoritmo DLP. Este comportamento pode estar relacionado diretamente com o tamanho do conjunto de dados, já que isto aumentaria as chances de serem escolhidas localizações não alcançáveis pelo DLP. Apesar disso, podemos perceber que, para qualquer grau de privacidade k, o DLP não garante um *k*-anonimato, uma vez que a taxa de reconhecimento se manteve acima de $\frac{1}{k}$ para o ataque baseado em distância euclidiana.

Em contrapartida, nas requisições que utilizam o PrivLBS como algoritmo de anonimização, a taxa de reconhecimento se manteve sempre abaixo de $\frac{1}{k}$ para todos os graus de privacidade observados. Além disso, quando comparado ao algoritmo DLP, a probabilidade de reidentificacação das localizações de um usuário utilizando o PrivLBS é, em média, quatro vezes menor que o algoritmo DLP para os valores de $k = \{2, 4, 8, 16, 32\}$. Essa probabilidade chega a ser até cinco vezes menor que o algoritmo DLP quando $k = 8$, diminuindo o valor, que antes era de 59,2%, para 11,4%.

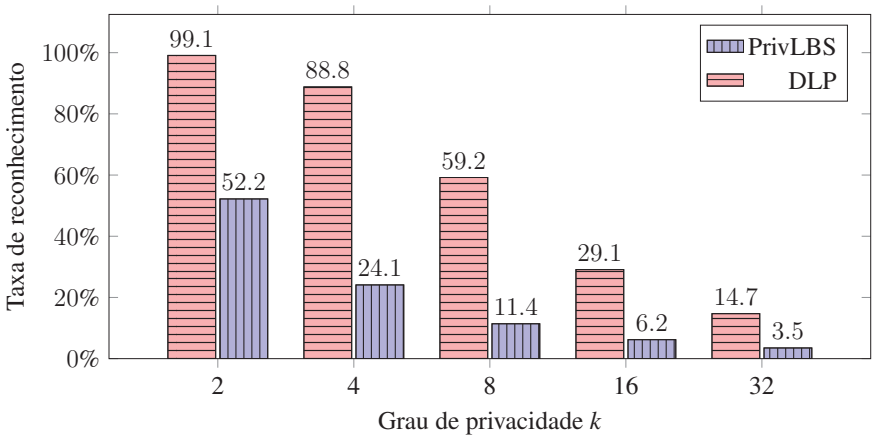


Figura 5. Taxas de reconhecimento da localização real para o ataque baseado em distância euclidiana.

6. Conclusão e Trabalhos Futuros

Neste trabalho apresentamos o PrivLBS, uma abordagem para preservação de privacidade de dados em serviços baseados em localização. Inicialmente propomos um modelo de ataque baseado na distância euclidiana entre as localizações contidas em requisições consecutivas ao LBS. Mostramos que esse tipo de ataque apresenta uma alta taxa de reidentificação quando aplicado sobre requisições consecutivas, que não foram anonimizadas considerando a distância euclidiana entre as localizações selecionadas. Demonstramos também que o PrivLBS, por ponderar tanto critérios de distância euclidiana como de probabilidade entre as localizações da consulta, apresenta uma baixa taxa de reidentificação ao sofrer ataques baseado em distância ou probabilísticos, garantindo as propriedades do modelo de privacidade k-anonimato.

Como trabalho futuro pretendemos realizar uma análise do impacto do tamanho do conjunto de dados sobre o PrivLBS, além de propor um modelo completo e dinâmico, buscando uma solução alternativa de seleção das localizações alcançáveis, que garanta uma maior entropia e produza o menor *overhead* possível, adotando, por exemplo, distância de rede de ruas para definir as localizações alcançáveis.

Agradecimentos

Os autores agradecem à CAPES, ao CNPq (132614/2017-0) e ao LSBD/UFC pelo financiamento parcial deste trabalho.

Referências

Gedik, B. and Liu, L. (2008). Protecting location privacy with personalized k-anonymity: Architecture and algorithms. *IEEE Transactions on Mobile Computing*, 7(1):1–18.

Kido, H., Yanagisawa, Y., and Satoh, T. (2005). An anonymous communication technique using dummies for location-based services. In *ICPS '05. Proceedings. International Conference on Pervasive Services, 2005.*, pages 88–97.

Li, H., Sun, L., Zhu, H., Lu, X., and Cheng, X. (2014). Achieving privacy preservation in wifi fingerprint-based localization. In *INFOCOM, 2014 Proceedings IEEE*, pages 2337–2345. IEEE.

Lu, R., Lin, X., Shi, Z., and Shao, J. (2014). Plam: A privacy-preserving framework for local-area mobile social networks. In *INFOCOM, 2014 Proceedings IEEE*, pages 763–771. IEEE.

Machanavajjhala, A., Gehrke, J., Kifer, D., and Venkitasubramaniam, M. (2006). l-diversity: Privacy beyond k-anonymity. pages 24–24.

Niu, B., Gao, S., Li, F., Li, H., and Lu, Z. (2016). Protection of location privacy in continuous lbss against adversaries with background information. In *2016 International Conference on Computing, Networking and Communications (ICNC)*, pages 1–6.

Niu, B., Li, Q., Zhu, X., Cao, G., and Li, H. (2014). Achieving k-anonymity in privacy-aware location-based services. In *INFOCOM, 2014 Proceedings IEEE*, pages 754–762. IEEE.

Niu, B., Li, Q., Zhu, X., Cao, G., and Li, H. (2015). Enhancing privacy through caching in location-based services. In *Computer Communications (INFOCOM), 2015 IEEE Conference on*, pages 1017–1025. IEEE.

Serjantov, A. and Danezis, G. (2003). Towards an information theoretic metric for anonymity. In Dingledine, R. and Syverson, P., editors, *Privacy Enhancing Technologies*, pages 41–53, Berlin, Heidelberg. Springer Berlin Heidelberg.

Sun, G., Chang, V., Ramachandran, M., Sun, Z., Li, G., Yu, H., and Liao, D. (2017a). Efficient location privacy algorithm for internet of things (iot) services and applications. *Journal of Network and Computer Applications*, 89:3 – 13. Emerging Services for Internet of Things (IoT).

Sun, G., Liao, D., Li, H., Yu, H., and Chang, V. (2017b). L2p2: A location-label based approach for privacy preserving in lbs. *Future Generation Computer Systems*, 74:375 – 384.

Sweeney, L. (2002). k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05):557–570.

Tsoukaneri, G., Theodorakopoulos, G., Leather, H., and Marina, M. K. (2016). On the inference of user paths from anonymized mobility data. In *2016 IEEE European Symposium on Security and Privacy (EuroS P)*, pages 199–213.

Ullah, I. and Shah, M. A. (2016). A novel model for preserving location privacy in internet of things. In *2016 22nd International Conference on Automation and Computing (ICAC)*, pages 542–547.

Vu, K., Zheng, R., and Gao, J. (2012). Efficient algorithms for k-anonymous location privacy in participatory sensing. In *2012 Proceedings IEEE INFOCOM*, pages 2399–2407.

Ying, B. and Makrakis, D. (2014). Protecting location privacy with clustering anonymization in vehicular networks. In *Computer Communications Workshops (INFOCOM WKSHPS), 2014 IEEE Conference on*, pages 305–310. IEEE.

Zhu, X., Chi, H., Niu, B., Zhang, W., Li, Z., and Li, H. (2013). Mobicache: When k-anonymity meets cache. In *Global Communications Conference (GLOBECOM), 2013 IEEE*, pages 820–825. IEEE.