Privacidade de Localização: Uma abordagem baseada em médias aleatórias

Rick'ardo D. N. Vieira[1], **Magnos Martinello**[1], **Ramon M. Ramos** [1], **Cesar A. C. Marcondes**[2]

[1]Departamento de Informática – Universidade Federal do Espírito Santo (UFES)
Av. Fernando Ferrari, S/N, 29060-970 – Vitória – ES, Brasil

[2]Departamento de Computação – Universidade Federal de São Carlos (UFSCar)
Rod. Washington Luís, Km 235, 13565-905 – São Carlos – SP, Brasil

{rick,magnos,ramon}@inf.ufes.br, marcondes@inf.ufscar.br

Resumo. *Este trabalho apresenta uma abordagem para garantir a privacidade de localização geográfica durante consultas públicas a Serviços Baseados em Localização (SBL). A ideia-chave apoia-se em médias aleatórias que permitem gerar regiões de anonimização integrando aspectos de aglomeração e aleatoriedade em uma única abordagem. Tal abordagem é avaliada considerando-se grau de anonimidade, tempo de consulta e precisão de localização. Simulações são executadas assumindo-se que os usuários são distribuídos sobre uma região de acordo com i) uma função de distribuição Uniforme; ii) uma função de distribuição Gaussiana; e iii) traços de mobilidade. Os resultados indicam que a abordagem propicia a obtenção de rotas com precisão aceitável, mantendo nível de anonimização e tempo de consulta satisfatórios.*

Abstract. *This work presents an approach to guarantee geographic location privacy for public queries in Location-Based Services (LBS). The key-idea relies on random means which allow to generate anonymous regions by integranting clustering and randomizing aspects in an unique approach. Such approach is evaluated considering anonimity level, query time and location precision. This study is conducted distributing randomly a set of users over a region. Simulations are executed assuming that users are distributed according i) Uniform distribution; ii) Gaussian distribution; and iii) mobility tracings. The results suggest that the proposed approach provides routes with acceptable precision, keeping level of anonimity and query time satisfactory.*

1. Introdução

Os modernos aparelhos celulares têm se tornando populares e sofisticados de tal modo que tem sido cada vez mais comum encontrar informação de localização geográfica global (GPS) embutida nos aparelhos. Este fato contribui para o desenvolvimento de uma ampla gama de aplicações que fazem uso de Serviços Baseados em Localização (SBL). Por exemplo, serviços de navegação em veículos [Keegan 2007], alertas de condições de tempo e de tráfego baseados em localização, grupo de amigos pertencente a uma região [de Oliveira Santos et al. 2008], etc. Essas aplicações, aliadas a disseminação da tecnologia de rede sem fio, parecem preparar o terreno da visão de computação ubíqua vislumbrada por Mark Weiser [Weiser 1991], onde a computação é invisível e permeia todos os aspectos do cotidiano.

Entretanto, toda vez que uma aplicação efetua uma consulta em um SBL, tal consulta contém a coordenada exata do usuário. Esta informação privada de localização pode ser correlacionada com outra informação pública (i.e. páginas amarelas) ou mesmo serviço de localização público, por exemplo *GoogleMaps*. Se as consultas forem efetuadas continuamente, então alguém não autorizado pode traçar os movimentos ou mesmo determinar caminhos percorridos pelo usuário móvel. Na prática, ao utilizar estes serviços, os usuários podem revelar informações relacionada às suas vidas privadas.

Um cenário ilustrativo é uma pesquisa pelas clínicas de tratamento de pacientes soropositivos mais próximas de uma determinada localização. Tal consulta revelaria a existência de um familiar portador do vírus da Aids, o que poderia despertar o preconceito para com o indivíduo. Assim, apresenta-se o seguinte problema: efetuar consultas a SBLs públicos de forma anônima, isto é, sem revelar a identidade/coordenada do usuário, obtendo resultados suficientemente precisos para uma eficaz orientação geográfica.

Para preservar a privacidade, ao invés de contactar diretamente um serviço de localização, o usuário poderia passar por um servidor confiável intermediário que esconderia sua identificação de origem (endereço IP). Estes serviços de navegação anônima na web estão disponíveis através de protocolos de anonimização tais como Crowds [Reiter and Rubin 1998] ou TOR [Syverson 2006] que proporcionam a anonimidade do emissor. É importante observar que o serviço de comunicação anônima permite preservar, à primeira vista, a identidade do emissor. No entanto, a coordenada exata do usuário permanece exposta nesta consulta ao SBL. A partir dessa coordenada, é possível inferir a identidade do usuário em diversas situações, como, por exemplo, quando o usuário efetua a consulta do seu endereço residencial.

As abordagens tradicionais encontradas na literatura buscam ocultar a localização do usuário construindo uma região de anonimização ao seu redor [Gedik and Liu 2005, Ghinita et al. 2007b, Zhong and Hengartner 2008], e enviando as coordenadas dessa região ao SBL. Partindo desses clássicos algoritmos de disfarce espacial, a proposta apresentada neste trabalho inova ao propor uma escolha aleatória de k pontos de anonimização sem geometria pré-definida.

A seção 2, descreve os fundamentos dos algoritmos de disfarce espacial de localização e alguns trabalhos relacionados. Na seção3, descreve-se a proposta de anonimização de coordenadas baseda em técnicas de médias aleatórias. Na seção 4, um estudo baseado em simulação é apresentado para avaliar a abordagem, onde podem-se verificar compromissos entre grau de anonimidade, maior precisão do destino de interesse e tempo de consulta anônima. Também, pode-se verificar que a quantidade de pontos de anonimização tem impacto significativo na precisão da localização. O artigo é finalizado com uma visão geral dos pontos investigados e com propostas de trabalhos futuros.

2. Fundamentação Teórica para Disfarce Espacial de Localização

A anonimidade de localização no caso de dispositivos móveis equipados com GPS, é definida como a propriedade que visa garantir que a coordenada do usuário não possa ser identificada com esforço computacional razoável [Gruteser and Grunwald 2003]. Em algoritmos de disfarce espacial, um dos princípios adotados consiste em manter um grau de anonimidade em qualquer região diminuindo o nível de precisão das informações de localização [Gruteser and Grunwald 2003, Ghinita et al. 2007b].

Para tratar consultas anônimas a SBLs públicos sem comprometimento da privacidade do usuário, muitas abordagens se apoiam no conceito de região espacial de anonimização de nível K (*K-ASR: K-Anonymous Spatial Region*) [Gruteser and Grunwald 2003, Gedik and Liu 2005, Zhong and Hengartner 2008]. A partir de um servidor centralizado de anonimização, os usuários precisam atualizar constantemente as suas respectivas localizações (coordenadas). Antes de efetuar uma pesquisa anônima, o "usuário consultante" solicita ao servidor as $K-1$ coordenadas dos usuários mais próximos. De posse de K coordenadas (incluindo a sua própria), o usuário gera uma região espacial de anonimização contendo todos os K usuários participantes do processo de anonimização. Para otimizar a resposta do SBL, a menor região espacial é gerada, o que comumente se dá com a definição de um retângulo cujas coordenadas sejam as coordenadas extremas dos usuários (Figura 1).

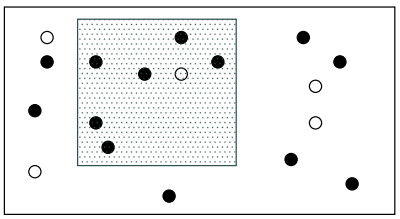


Figura 1. Usuários: pontos pretos; estabelecimentos: pontos brancos

Na sequência, os dados da consulta por um determinado tipo de estabelecimento, bem como as coordenadas do retângulo obtido, são enviados ao SBL que retorna, como resposta, a(s) coordenada(s) do(s) estabelecimento(s) de interesse mais próximo(s) do retângulo. Como coordenadas de K usuários são enviadas ao SBL, existe uma probabilidade 1/K de o usuário ter efetuado a consulta, o que determina um grau K de anonimidade, sendo esta, a vantagem desta abordagem. Quanto maior a quantidade de usuários envolvidos, menor a probabilidade de identificação do usuário consultante.

Entretanto, o uso de um servidor privado/centralizado para gerenciar a localização dos usuários tem seus inconvenientes. A centralização de informações pode se tornar um gargalo para o sistema à medida que o mesmo cresce (problema de escalabilidade). Além disso, ataques bem-sucedidos ao servidor podem comprometer tanto a privacidade dos usuários (informações centralizadas) quanto a continuidade no fornecimento do serviço.

Como o servidor de anonimização irá retornar ao usuário consultante as coordenadas dos usuários efetivamente mais próximos, existe a forte tendência de que este usuário localize-se no centro dos demais usuários, o que oferece subsídios para revelar a localização (e consequente identidade) de quem efetua a consulta. Além disso, para efetiva utilização da abordagem, é necessário que a semântica do SBL trate "coordenadas de retângulos" em vez de coordenadas pontuais. Adicionalmente, há uma limitação no fornecimento de respostas do serviço baseado em retângulos, pois não existe a possibilidade de um retorno contendo rotas.

Para solucionar os problemas de escalabilidade e de vulnerabilidade dos servidores centralizados de anonimização, foram propostas abordagens que distribuem as informações de coordenadas dos usuários em uma rede P2P formada pelos próprios usuários [Ghinita et al. 2007b, Ghinita et al. 2007a]. Nessas abordagens, os usuários

habilitam-se no sistema através de um servidor de autenticação, cuja tarefa é simplesmente a identificação dos usuários, não acumulando qualquer tarefa relacionada ao gerenciamento de coordenadas. As informações de localização dos usuários estão distribuídas entre os usuários, que trocam informações entre si constantemente para atualizar dados e fornecer coordenadas de usuários mais próximos sem, contudo, revelar identidades.

Adicionalmente, o problema da tendência de centralização do usuário consultante em relação aos demais usuários foi solucionada nas abordagens indicadas com a redefinição da métrica de proximidade. Em vez de considerar usuários mais próximos segundo uma métrica de distância Euclideana, por exemplo, as coordenadas dos usuários são mapeadas para um espaço de Hilbert [Hilbert 1891], no qual uma curva de preenchimento espacial 2-D é definida sobre a região geográfica considerada. As informações 2-D de localização dos usuários (coordenadas) são, então, mapeadas para informações 1-D de valores sobre a curva de Hilbert. Quanto mais próximos os valores de Hilbert, mais próximos estão os usuários uns dos outros. Como pode ser observado na Figura 2, os usuários efetivamente mais próximos (métrica Euclideana) nem sempre são tidos como tais segundo a lógica de Hilbert. Isso evita satisfatoriamente a tendência de centralização do usuário consultante. Todavia, tais abordagens não tratam das questões relativas à adequação do SBL a coordenadas de retângulos e muito menos fornecem rotas como retorno.

A abordagem TPA (Triângulo Pontualizado de Anonimização) [Vieira et al. 2009] apresenta uma proposta de solução para os problemas de adequação do SBL a coordenadas de retângulos, oferecendo a possibilidade de obtenção de rotas como retorno. Conforme esse estudo, para que o SBL entregue rotas, é de grande utilidade alguma orientação adicional a respeito da localização do usuário. Contudo, esta orientação adicional não deve comprometer o nível de anonimização desejado para as pesquisas efetuadas. Em vez de construir uma região espacial de anonimização baseada em retângulos (ou quaisquer figuras planas), o TPA constitui-se de um conjunto de 4 pontos (os 3 vértices de um triângulo mais o seu baricentro) dispostos conforme média e desvio das coordenadas dos K usuários participantes do processo de anonimização (mantendo, assim, o nível K de anonimidade) (vide Figura 3). As coordenadas desses 4 pontos são enviadas ao SBL, que fornece rotas para os mesmos como retorno.

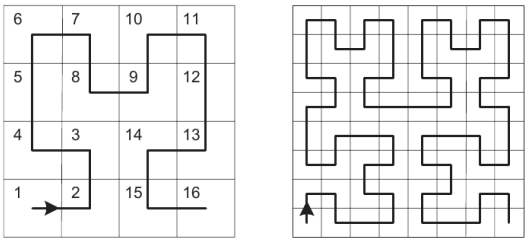


Figura 2. Curvas de Hilbert

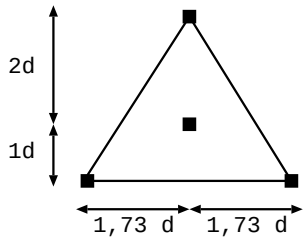


Figura 3. TPA: desvio "d"

É interessante desvincular a definição da região de anonimização de uma figura de geometria facilmente dedutível para evitar o fornecimento de subsídios a atacantes. Uma abordagem alternativa para enfrentar este problema de geometria seria a escolha aleatória de k pontos de anonimização. Entretanto, uma total aleatoriedade poderia ser ineficiente

no fornecimento de rotas "úteis" ao usuário (isto é, rotas que estivessem a no máximo alguns poucos cruzamentos de distância dos K usuários).

Rotas úteis poderiam ser fornecidas com o uso de k pontos "ótimos", isto é, pontos cujo somatório de menores distâncias até os usuários mais próximos seja o menor possível (problema de aglomeração ou clusterização em k-médias). Contudo, este problema de otimização é extremamente custoso, sem solução prática na literatura para valores substanciais de K (número de usuários) e de k (quantidade de médias). Medidas aproximadas envolvendo a quantidade de pontos "ótimos" e o "erro" (definido como somatório de menores distâncias) produzido por tais pontos revelaram que a relação entre essas duas grandezas (erro $versus$ número de pontos) tem um decaimento exponencial.

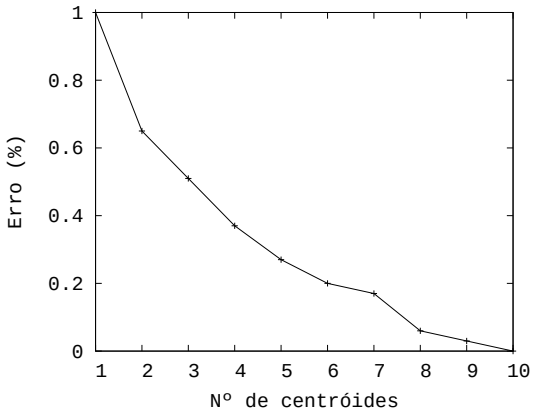


Figura 4. Impacto da quantidade de pontos no percentual de erro

3. Privacidade de localização baseada em Médias Aleatórias

A ideia chave para garantir privacidade, chamada aqui de Médias Aleatórias - MA, apoia-se na combinação de aglomeração e aleatoriedade para gerar a região de anonimização. Sua estrutura é composta dos seguintes elementos: *i*) definição do grau de anonimidade; *ii*) obtenção das coordenadas de usuários mais próximos; *iii*) cálculo de k centróides; *iv*) solicitação de k rotas ao SBL; *vi*) exibição para o usuário consultante.

Assim, definido o nível K de anonimidade desejado [1], a abordagem MA consiste, primeiramente, na obtenção de K-1 coordenadas dos usuários mais próximos do usuário consultante (segundo a distância de Hilbert, por exemplo), no posterior cálculo dos k centróides do problema de aglomeração através do algoritmo de Lloyd [Lloyd 1982], no envio de consultas ao SBL solicitando k rotas dos centróides aos estabelecimentos de interesse (por exemplo, clínica de tratamento de pacientes soropositivos) e na efetiva exibição para o usuário da rota mais próxima de sua atual localização.

Tal abordagem torna-se interessante por possuir potencial próximo ao da solução ótima, sem, contudo, onerar o cálculo de centróides, com a vantagem adicional de possuir um nível de anonimização relativamente superior ao das abordagens de geometria fixa devido à escolha aleatória de pontos iniciais.

Formalmente, o problema de aglomeração (clusterização) em k-médias (*k-means clustering problem*) apresenta-se da seguinte forma:

[1] Na notação adotada, K (maiúsculo) = nº de usuários e k (minúsculo) = nº de centróides

Dados um conjunto real d-dimensional de entidades $x_1, x_2, ..., x_K$ e um inteiro $k < K$, encontrar o conjunto $S = S_1, S_2, ..., S_k$ que minimize a seguinte expressão:

$$\sum_{i=1}^{k} \sum_{x_j \in S_i} \|x_j - \mu_i\|^2, \tag{1}$$

onde μ_i é o centróide (ponto central) de S_i.

Observa-se que o objetivo do problema é minimizar a variância total intra-aglomerados (*intra-clusters*), minimizando a soma de erros quadráticos. Conforme o trabalho ([Inaba et al. 1994]) para n entidades [2] a serem aglomeradas, o problema pode ser resolvido no tempo polinomial $O(n^{dk+1} \log n)$. Sendo um problema NP-difícil, em geral são utilizados algoritmos heurísticos para solucioná-lo. Em particular, o trabalho [Lloyd 1982] apresentou um algoritmo simples e elegante para encontrar uma solução aproximada para este problema de aglomeração, conhecido também como Iteração (ou Relaxação) de Voronoi. O algoritmo de Lloyd tem seu uso tão disseminado que é citado frequentemente na literatura como o "Algoritmo de k-médias" (*k-means algorithm*).

O algoritmo começa com um conjunto inicial de k médias (os centróides) definidas aleatoriamente ou segundo alguma heurística específica. Aglomerados são definidos associando-se cada um dos K pontos do conjunto de entidades ao "centróide" mais próximo, usualmente utilizando uma função de distância Euclideana. Em seguida, recalculam-se os centróides de cada aglomerado por meio de alguma métrica (em geral, médias dimensionais do espaço Euclideano em questão). Novos centróides são recalculados sucessivamente até que uma das condições de convergência seja atingida, o que é obtido quando nenhum dos K pontos troca de centróide ou, alternativamente, quando nenhum centróide é recalculado. Matematicamente, estes dois passos de *Associação* e de *Atualização* podem ser expressos da seguinte forma.

Dado um conjunto inicial de k médias ${m_1}^{(1)}, ...{m_k}^{(1)}$, o algoritmo prossegue alternando entre dois passos:

Passo de Associação: *Associe cada entidade ao aglomerado cujo centróide seja o mais próximo.*

$$S_i^{(t)} = \left\{ x_j : \left\| x_j - m_i^{(t)} \right\| \leq \left\| x_j - m_{i^*}^{(t)} \right\| \forall\ i^* = 1, \ldots, k \right\} \tag{2}$$

Passo de Atualização: *Para cada aglomerado, calcule as novas médias e atualize os centróides para estes valores.*

$$m_i^{(t+1)} = \frac{1}{|S_i^{(t)}|} \sum_{x_j \in S_i^{(t)}} x_j \tag{3}$$

Apesar do algoritmo convergir, um critério de parada comumente utilizado é a adoção de um número máximo de iterações. Como os k pontos iniciais podem ser definidos aleatoriamente, observa-se que soluções diferentes são obtidas a cada execução do algoritmo, pois há, em geral, mais de um conjunto quasi-ótimo de aglomerados.

[2] Na notação adotada, K (maiúsculo) é equivalente ao n

4. Experimentos e Análise de Sensibilidade

Essa seção avalia a abordagem proposta levando em conta duas métricas-chaves: tempo de consulta e obtenção de rotas úteis (precisão). A quantidade k de pontos a serem utilizados no processo de consulta ao SBL é um parâmetro que tem especial interesse para aplicações práticas, já que o tempo total da consulta ao SBL é função da quantidade de rotas solicitadas. Este tempo é afetado pelos seguintes componentes:

- tempo para consulta na rede P2P;
- tempo de cálculo dos k pontos (centróides) de anonimização;
- tempo de transmissão dos dados do usuário para o SBL;
- tempo de processamento, no SBL, para obtenção de rotas;
- tempo de transmissão dos dados do SBL para o usuário;
- tempo de processamento, no usuário, para filtragem de resultados (exibição apenas da rota mais próxima);

A metodologia de avaliação é conduzida através de uma bateria de experimentos para determinar o comportamento do algoritmo face a um conjunto de 100 usuários localizados aleatoriamente sobre o plano. A distribuição dos usuários no plano é efetuada através de 3 simulações distintas: *i*) considerando usuários distribuídos de acordo com uma função de distribuição Uniforme; *ii*)usuários distribuídos segundo uma função de distribuição Normal e *iii*) uma simulação baseada em traços de mobilidade [Nagel 2010].

4.1. Percentual de centróides não-associados e tempo de consulta

Os resultados destes experimentos demonstram que, ao executar o algoritmo de Lloyd, nem todos os centróides iniciais (definidos aleatoriamente) associam-se a usuários, permanecendo sem vínculo até o alcance da condição de parada e tornando-se, portanto, inutilizáveis. Denotam-se esses centróides sem vínculos de "não-associados". À medida que a quantidade k de centróides aumenta, o valor de não-associação altera-se, diferentemente, para cada uma das distribuições utilizadas. A Figura 5 resume os resultados.

Observa-se que com a distribuição Uniforme obtém-se um índice de até 70% de não-associação (isto é, com k = 10, por exemplo, apenas 3 centróides são utilizados), valor que cai vertiginosamente para cerca de 20% com a distribuição Normal (isto é, para k = 10, 8 centróides úteis), atingindo em torno de 50% para valores obtidos com a simulação de traços de mobilidade (isto é, 5 centróides efetivamente utilizados com k = 10).

Para avaliar o tempo de consulta, uma aplicação foi desenvolvida na plataforma Android para servir como ambiente de experimentação. A partir de um aparelho G1 conectado através de uma rede 3G, foram efetuadas pesquisas ao GoogleMaps. O experimento avaliou o tempo de consulta do SBL para obter k (centróides) rotas, com k variando de 1 a 10, cujo resultado é apresentado na Figura 6. Analisando os dois resultados conjuntamente, observa-se um compromisso interessante quando o valor de k é 4 e 5, pois o tempo de consulta médio varia de 4,5 a 5,5 ms e o índice de não-associação é, no máximo, de 50%. Por essa razão, optou-se por estes valores para continuidade dos experimentos.

4.2. Precisão de rotas úteis

Com base no resultado de percentual apropriado de centróides não-associados, é preciso avaliar a precisão de resultados obtidos utilizando a abordagem MA como técnica de

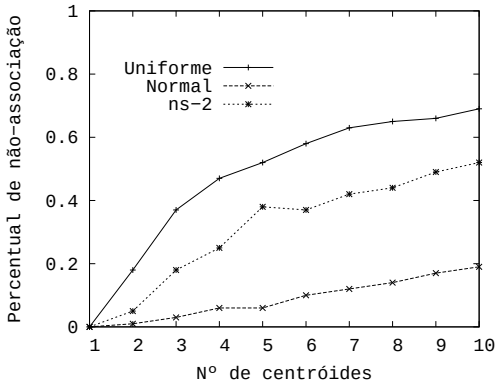


Figura 5. Percentual de usuários não associados

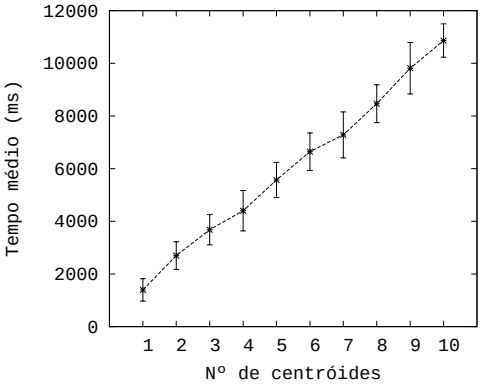
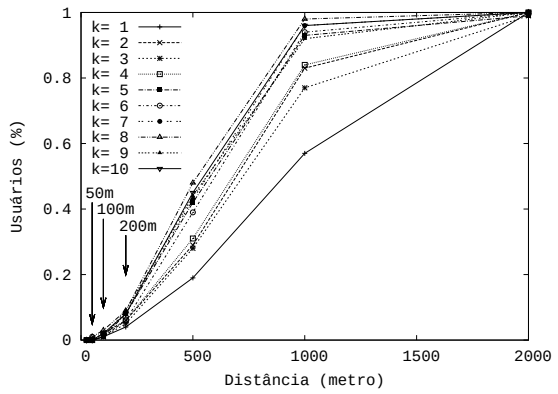
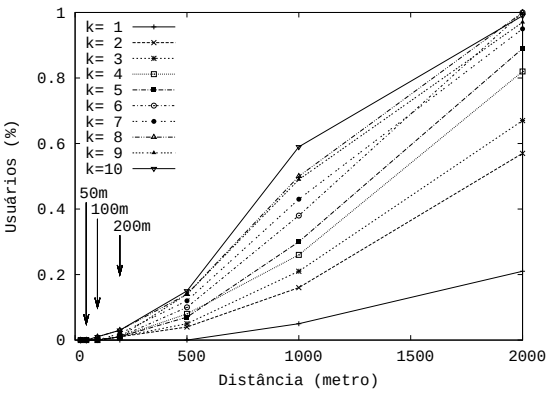


Figura 6. Tempo médio de consulta ao SBL

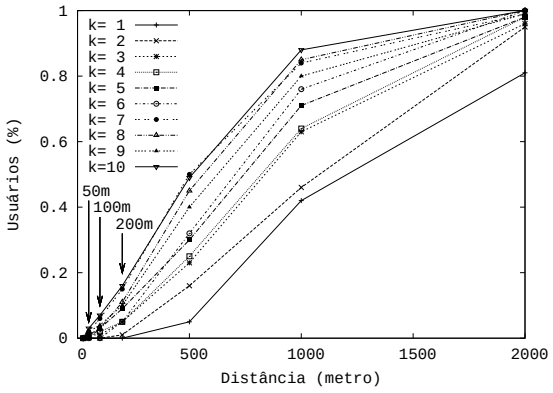
anonimização. Considerando que as rotas fornecidas pelo SBL são dos centróides até os estabelecimentos de interesse, a medida dessa precisão considera a distância D (variável aleatória) que cada usuário encontra-se de seu respectivo centróide. A Figura 7 apresenta a distribuição cumulativa de probabilidade $P[D \leq d]$ em função das distâncias d.



(a) Distribuição Uniforme



(b) Distribuição Normal



(c) Simulação via ns-2

Figura 7. Porcentagem de usuários e distâncias dos centróides

A tabela 1 resume alguns dados relevantes. Observa-se que, para k = 4, tem-se

de 54% a 92% dos usuários atendidos pelos respectivos centróides para distâncias inferiores a 1500 metros, conforme a distribuição utilizada. Para k = 5, esses valores vão de 60% a 96%. Essas distâncias, calculadas segundo a métrica Euclideana (em linha reta), podem ser traduzidas para quadras de 100x100 metros, fornecendo o valor aproximado de 11 quadras. Isto significa que para uma distribuição uniforme dos usuários, mais de 90% destes encontram-se a no máximo 11 quadras de seus respectivos centróides, o que fornece informação efetiva para uma adequada orientação na região do contexto de anonimização. Nota-se que no caso da distribuição uniforme, todos as distâncias tem a mesma probabilidade de ocorrer, enquanto as distâncias médias na distribuição normal ocorrem com maior probabilidade, o que explica as diferenças obvervadas no resultado.

Tabela 1. Porcentagem de usuários

Simulação	***N° de centróides***	***Distâncias em metros***			
		$d < 500$	$d < 1000$	$d < 1500$	$d < 2000$
Uniforme	$k = 4$	$0,31$	$0,84$	$0,92$	$1,00$
	$k = 5$	$0,42$	$0,93$	$0,96$	$0,99$
Normal	$k = 4$	$0,08$	$0,26$	$0,54$	$0,82$
	$k = 5$	$0,07$	$0,30$	$0,60$	$0,89$
Traços	$k = 4$	$0,25$	$0,64$	$0,81$	$0,98$
	$k = 5$	$0,30$	$0,71$	$0,85$	$0,98$

5. Conclusões e Trabalhos Futuros

Os resultados obtidos confirmam o potencial das ideias apresentadas para aplicações práticas envolvendo consultas privadas a SBLs públicos. Os experimentos com o aparelho G1 demonstraram que é possível alcançar um equilíbrio entre precisão de resultados (rotas), nível de anonimização e tempo de respostas (do SBL). Tem-se, portanto, como efetiva contribuição, um eficiente mecanismo para obtenção de rotas em contexto de anonimização.

Em termos de privacidade, obteve-se um segundo nível de anonimização com o uso de centróides como coordenadas de consulta, pois, além do nível K de anonimidade definido pelas coordenadas dos usuários, as k coordenadas dos centróides enviadas ao SBL produzem um novo embaralhamento na localização do usuário consultante.

Apesar de [Ghinita et al. 2007b] apresentar uma abordagem elegante para o problema da obtenção de K coordenadas em uma rede P2P, tal solução possui apenas resultados simulados (p2psim), não trazendo ferramentas efetivas para o enfrentamento de situações tipicamente presentes em implementações, como, por exemplo, a dificuldade de se manter em memória uma tabela para conversão de valores de Hilbert, haja vista ser essa uma curva de Peano (curva de preenchimento) e, como tal, não possuir função algébrica descritiva. Pretende-se estudar essa e outras problemáticas mais a fundo, na tentativa de contornar os problemas existentes.

As ideias apresentadas nesse trabalho precisam de testes exaustivos em situações reais, com mapas e rastros de mobilidade reais, para efetiva validação. Os desenvolvimentos previstos englobam a implementação completa de um sistema que contemple a estrutra integral de definição de pontos de consulta e tratamento do retorno obtido.

Referências

de Oliveira Santos, R., Fabris, F., Martinello, M., and Marcondes, C. (2008). Joinus: Management of mobile social networks for pervasive collaboration. In *SBSC*, pages 224–234. IEEE Computer Society.

Gedik, B. and Liu, L. (2005). Location privacy in mobile systems: A personalized anonymization model. In *Proceedings of IEEE International Conference on Distributed Computing Systems-ICDCS*, pages 620–629.

Ghinita, G., Kalnis, P., and Skiadopoulos, S. (2007a). Mobihide: A mobile peer-to-peer system for anonymous location-based queries. In *SSTD '07: Proceedings of the 10th international symposium on Advances in Spatial and Temporal Databases*, pages 221–238, Berlin, Heidelberg. Springer-Verlag.

Ghinita, G., Kalnis, P., and Skiadopoulos, S. (2007b). Privé: Anonymous location-based queries in distributed mobile systems. *16th international conference on World Wide Web*, (1):371–380.

Gruteser, M. and Grunwald, D. (2003). Anonymous usage of location-based services through spatial and temporal cloaking. *Proceedings of First ACM/USENIX International Conference on Mobile Systems, Applications, and Services (MobiSys)*.

Hilbert, D. (1891). Üeber die stetige Abbildung einer Linie auf ein Flächenstück. In Felix Klein, Walther Dyck, and Adolph Mayer, editors, *Mathematische Annalen*, volume 38, pages 459–460. Springer.

Inaba, M., Katoh, N., and Imai, H. (1994). Applications of weighted voronoi diagrams and randomization to variance-based k-clustering: (extended abstract). In *SCG '94: Proceedings of the tenth annual symposium on Computational geometry*, pages 332–339, New York, NY, USA. ACM.

Keegan, M. (2007). Onstar could thwart car thieves. *The Auto Writer*.

Lloyd, S. P. (1982). Least squares quantization in pcm. *IEEE Transactions on Information Theory*, 28(2):129–137.

Nagel, K. (2010). Multi-agent microscopic traffic simulator. *http://www.lst.inf.ethz.ch/research/ad-hoc/*.

Reiter, M. and Rubin, A. (1998). Crowds: Anonymity for web transactions. *ACM Transactions on Information and System Security*, 1(1).

Syverson, P. (2006). Locating hidden servers. *IEEE Symposium on Security and Privacy*, 1(1).

Vieira, R. D. N., Martinello, M., and Marcondes, C. A. C. (2009). Privacidade de localizacao em servicos moveis: Anonimidade-k baseada em triangulo pontualizado. In *In SBCUP*.

Weiser, M. (1991). The computer for the twenty-first century. *Scientific American*, (94).

Zhong, G. and Hengartner, U. (2008). Toward a distributed k-anonymity protocol for location privacy. *Conference on Computer and Communications Security Proceedings of the 7th ACM workshop on Privacy in the electronic society*, pages 33–38.