Rick'ardo Debiazze Nunes Vieira

Privacidade de localização geográfica em Consultas a Serviços Públicos Web de Localização: Uma abordagem baseada em médias aleatórias

Vitória - ES, Brasil

18 de agosto de 2010

Rick'ardo Debiazze Nunes Vieira

Privacidade de localização geográfica em Consultas a Serviços Públicos Web de Localização: Uma abordagem baseada em médias aleatórias

Dissertação apresentada para obtenção do Título de Mestre em Informática pela Universidade Federal do Espírito Santo.

Orientador:
Magnos Martinello

Co-orientador:
Cesar Augusto C. Marcondes

Departamento de Informática
Centro Tecnológico
Universidade Federal do Espírito Santo

Vitória - ES, Brasil

18 de agosto de 2010

Dissertação de Mestrado sob o título *"Privacidade de localização geográfica em Consultas a Serviços Públicos Web de Localização: Uma abordagem baseada em médias aleatórias "*, defendida por Rick'ardo Debiazze Nunes Vieira e aprovada em 18 de agosto de 2010, em Vitória, Estado do Espírito Santo, pela banca examinadora constituída pelos professores:

Prof. Dr. Magnos Martinello
Orientador

Prof. Dr. Cesar Augusto C. Marcondes
Co-orientador
Universidade Federal de São Carlos

Prof. Dr. José Gonçalves Pereira Filho
Universidade Federal do Espírito Santo

Prof. Dr. Artur Ziviani
Laboratório Nacional de Computação Científica

Resumo

A crescente demanda por Serviços Baseados em Localização (SBLs) tem despertado a necessidade de soluções que considerem a privacidade das consultas efetuadas a SBLs públicos. As abordagens atualmente propostas apoiam-se, principalmente, na construção de uma região espacial de anonimização (ASR) baseada em um retângulo que contenha o usuário consultante e, pelo menos, outros $K - 1$ usuários, estabelecendo, assim, um nível K de anonimidade para a consulta realizada.

Paralelamente, infraestruturas P2P têm sido elaboradas para suportar a carga de mobilidade que vem caracterizando boa parte dos usuários de SBLs, haja vista o notável avançado tecnológico que proporcionou a disponibilização de modernos recursos (como capacidade de processamento do sinal GPS) em aparelhos portáteis, tais como os *smartphones*.

Contudo, as respostas obtidas dos SBLs, considerado o contexto das pesquisas realizadas, limitam-se à indicação dos pontos de interesse mais próximos da localização do usuário consultante, não sendo capazes de indicar uma rota para o destino apontado.

O presente trabalho visa apresentar a proposta *AnoniMobi*, cujo objetivo principal é fornecer rotas para pontos de interesse como resposta a consultas anônimas efetuadas a SBLs públicos, garantindo o nível de privacidade requerido pelo usuário consultante. A abordagem desenvolvida desvincula-se de propostas tradicionais para formação de região espacial de anonimização (ASR) ao considerar o uso de pontos (médias aleatórias) que permitem agrupar coordenadas de modo não-determinístico. Adicionalmente, uma infraestrutura P2P baseada na arquitetura CAN é elaborada como suporte à auto-organização necessária para estruturar e agrupar os usuários do sistema.

Abstract

The increasing demand for Location-Based Services (LBSs) has increased the need for solutions that consider the privacy of queries in public LBSs. The currently proposed approaches rely mainly on the construction of an anonymizing spatial region (ASR) based on a rectangle containing a user consultant and at least other $K-1$ users, thereby setting a K level of anonymity for the query performed.

In parallel, P2P infrastructures have been designed to bear the burden of mobility that has characterized much of LBSs users, given the remarkable technological advance that provided the availability of modern resources (such as processing of the GPS signal) on portable devices, such as smartphones.

However, the answers obtained from LBSs, considering the context of research conducted, is limited to the indication of points of interest near the consultant user's location, not being able to indicate a route to the destination indicated.

This paper aims at presenting the proposal *AnoniMobi*, whose main objective is to provide routes to points of interest in response to anonymous queries made to public LBSs, ensuring the privacy level requested by the consultant user. The approach separates itself from traditional proposals for formation of spatial region anonymization (ASR) when considering the use of points (random means) that allow you to group coordinates in a non-deterministic way. Additionally, an infrastructure based on CAN P2P architecture is designed to support self-organization needed for structuring and grouping users of the system.

Dedicatória

Dedico este trabalho à minha esposa, Ediane de Sousa Lima, companheira fiel que concordou em suportar minha ausência para que eu pudesse dar vida a este trabalho.

Agradecimentos

Agradeço a Deus pela oportunidade de aprendizado técnico através das pesquisas realizadas. Agradeço mais ainda pelo aprendizado moral e ético obtido através do convívio com as pessoas.

Agradeço ao professores Magnos Martinello e Cesar Augusto C. Marcondes por terem aceitado a tarefa de orientar-me e, sobretudo, fazer-me compreender o sentido de um Mestrado.

Agradeço aos amigos e parentes, que entenderam a necessidade de ausência (e, de certo modo, de isolamento) para a conclusão das tarefas.

Sumário

Lista de Figuras

Lista de Tabelas

1 Introdução

"*Não há fé inabalável senão aquela que pode encarar a razão face a face, em todas as épocas da Humanidade.*"
Allan Kardec

1.1 Contexto

Os equipamentos eletrônicos portáteis têm gradativamente incorporado novas funcionalidades anteriormente restritas aos computadores *desktops* (aumento na capacidade de armazenamento/processamento, ubiquidade de acesso a Internet, etc). Devido ao seu baixo custo, esses equipamentos estão ganhando popularidade e podem em pouco tempo tornar-se o modo dominante pelo qual os usuários acessem a Internet, conforme aponta [Buford, Heather e Lua 2009]. Consequentemente, os aparelhos móveis terão um papel importante nas futuras redes P2P.

Boa parte dos estudos sobre mobilidade em sistemas P2P assume que a população de pares consiste primariamente de nós não-móveis [Androutsellis-Theotokis e Spinellis 2004]. A tendência no uso de equipamentos eletrônicos e aparelhos celulares sugere que no futuro a maioria dos equipamentos conectados à Internet poderão estar operando em nós móveis. Supondo que se concretize essa perspectiva, as técnicas baseadas em redes sobrepostas [1] (como é o caso dos sistemas P2P) que "acomodarem" mobilidade utilizando recursos de pares estacionários podem ter seu desempenho degradado.

A mobilidade tem gerado demandas por Serviços Baseados em Localização (SBLs)[2], impulsionados, em grande parte, pelo barateamento de equipamentos com capacidade de recepção e processamento dos sinais GPS. Os smartphones com tais funcionalidades de georeferenciamento são um exemplo de aparelhos portáteis que proporcionam aos seus usuários desfrutarem de SBLs para monitorarem a localização de uma frota de veículos, para encontrarem amigos em uma região geográfica, ou simplesmente para descobrirem o restaurante mais próximo.

No que concerne ao uso de SBLs, [Kütter 2005] informa que a privacidade é uma preocupação recorrente e, na prática, há uma relutância em acessar serviços que possam revelar afiliações políticas/religiosas ou estilos de vida alternativos. Pesquisar a localização geográfica de escritórios de um partido político pode indicar um vínculo partidário que se deseje manter em segredo, por exemplo. Semelhantemente, uma consulta pelo hospital para tratamento de pacientes soropositivos mais próximo pode revelar a existência de um familiar portador do vírus HIV, e o receio de despertar discriminações pode impedir o usuário de usufruir de um SBL para tal finalidade. Além disso, usuários podem hesitar em efetuar consultas aparentemente inofensivas como "qual a farmácia mais próxima de minha atual localização?" desde que, uma vez revelada sua identidade, passem a receber

[1] Redes *overlay*.
[2] Vide Capítulo 3.

propagandas e promoções não-solicitadas (spam).

Neste contexto, a identidade do usuário de um SBL precisa ser preservada. Apesar de um usuário ser capaz de ocultar dos SBLs sua identidade pessoal utilizando serviços de navegação anônima, as consultas efetuadas possuem informações de localização geográfica. Tais informações podem, em muitos casos, ser relacionadas ao usuário consultante e, desta forma, revelar indiretamente a sua identidade.

A relação entre coordenadas geográficas e indivíduos pode se dar, por exemplo, através de triangulação do sinal de telefones celulares, método funcional tanto em áreas externas como em ambientes internos. Um atacante (terceiro não-confiável), de posse da informação de localização enviada ao SBL e da coordenada geográfica obtida do sistema de telefonia celular, pode identificar o autor da consulta por mera comparação de dados.

Por meio de observação física também é possível determinar a coordenada de um indivíduo. Um atacante, de posse de um aparelho de GPS, pode perseguir sua vítima, monitorando constantemente a sua própria coordenada, e, consequentemente, deduzir a coordenada do usuário do SBL.

Determinadas situações propiciam maior praticidade ainda ao atacante. Se o usuário efetua consultas a partir de um local que o caracterize bem (sua residência, por exemplo), a informação de localização fornecida ao SBL pode ser correlacionada a bancos de dados públicos (como o GoogleMaps), revelando sua identidade.

Assumindo-se que SBLs públicos podem ser vulneráveis e, consequentemente, entidades não-confiáveis, atacantes têm a capacidade de comprometer o serviço oferecido, neutralizando as tentativas de garantia de privacidade durante consultas realizadas por terceiros. Desta forma, o problema abordado neste trabalho pode ser descrito como se segue:

> **Realizar, em tempo satisfatório, consultas anônimas a SBLs públicos, garantindo o nível de privacidade desejado com o menor consumo de recursos possível (processamento, armazenamento, largura de banda).**

As abordagens tradicionais enfrentam o problema considerando a construção de uma região espacial de anonimização (ASR) baseada em retângulos contendo um mínimo de usuários que garanta um nível K de anonimidade.

A abordagem desenvolvida no sistema *AnoniMobi* desvincula-se de propostas tradi-

cionais para formação de região espacial de anonimização (ASR) ao considerar o uso de pontos (médias aleatórias) que permitem agrupar coordenadas de modo não-determinísco. Essa propriedade leva à construção de uma ASR com garantia do nível desejado de privacidade. Além disso, o uso de pontos como unidade básica de consulta proporciona não apenas uma semântica mais adequada para o processamento dos SBLs mas, essencialmente, provê rotas que tendem a aumentar consideravelmente a qualidade da resposta.

Para a construção da ASR sugerida, o sistema *AnoniMobi* faz uso de um algoritmo de aglomeração (algoritmo de Lloyd), com baixíssimo consumo de recursos, executado no próprio nó consultante. As informações de coordenadas vizinhas (usuários mais próximos, segundo a métrica de Hilbert) são coletadas na infraestrutura P2P subjacente (baseada na arquitetura CAN).

Os resultados (teóricos e simulados) obtidos neste trabalho indicam o potencial do sistema *AnoniMobi*, comparativamente às abordagens tradicionais. Algumas propostas para implementação do protocolo de comunicação entre os nós é apresentada como indicação de trabalhos futuros.

1.2 Estrutura da Dissertação

Além deste capítulo introdutório, o presente trabalho é composto por um capítulo referente aos trabalhos relacionados (Capítulo 2), quatro capítulos de fundamentação teórica (Capítulos de 3 a 6), dois capítulos sobre a abordagem proposta (Capítulos 7 e 8) e um capítulo de conclusão (Capítulo 9).

As ideias-chave das atuais abordagens vinculadas ao universo da anonimização de informações de localização são apresentadas no Capítulo 2, "Trabalhos Relacionados". Em seguida o capítulo 3, "Serviços Baseados em Localização", expõe fundamentos relativos aos SBLs e introduz conceitos relacionados a essa tecnologia e a outras afins. Subsídios para a compreensão do funcionamento dos Sistemas P2P são fornecidos no capítulo 4, "Redes Par a Par e Mobilidade". O capítulo 5, "Curvas de Peano", traz conceitos básicos relacionados às curvas de preenchimento espacial. Por sua vez, conceitos básicos relacionados ao universo dos problemas de agrupamento são fornecidos no capítulo 6, "Algoritmos de Agrupamento". A abordagem proposta, batizada de *AnoniMobi*, é formalmente apresentada no capítulo 7, "*AnoniMobi*, a Abordagem Proposta", e os resultados de experimentos estão descritos no capítulo 8, "Resultados e Análises". Por fim, o capítulo 9, "Conclusões e Trabalhos Futuros", aponta as conclusões referentes aos estudos realizados

e indica caminhos de continuidade para as pesquisas desenvolvidas.

O leitor que se sinta à vontade com os temas abordados na fundamentação teórica poderá concentrar-se nos capítulos 2, 7, 8 e 9 sem prejuízo de compreensão das ideias discutidas. A Figura 1 ilustra a dependência dos assuntos abordados em cada capítulo, servindo como orientação para uma leitura mais direcionada.

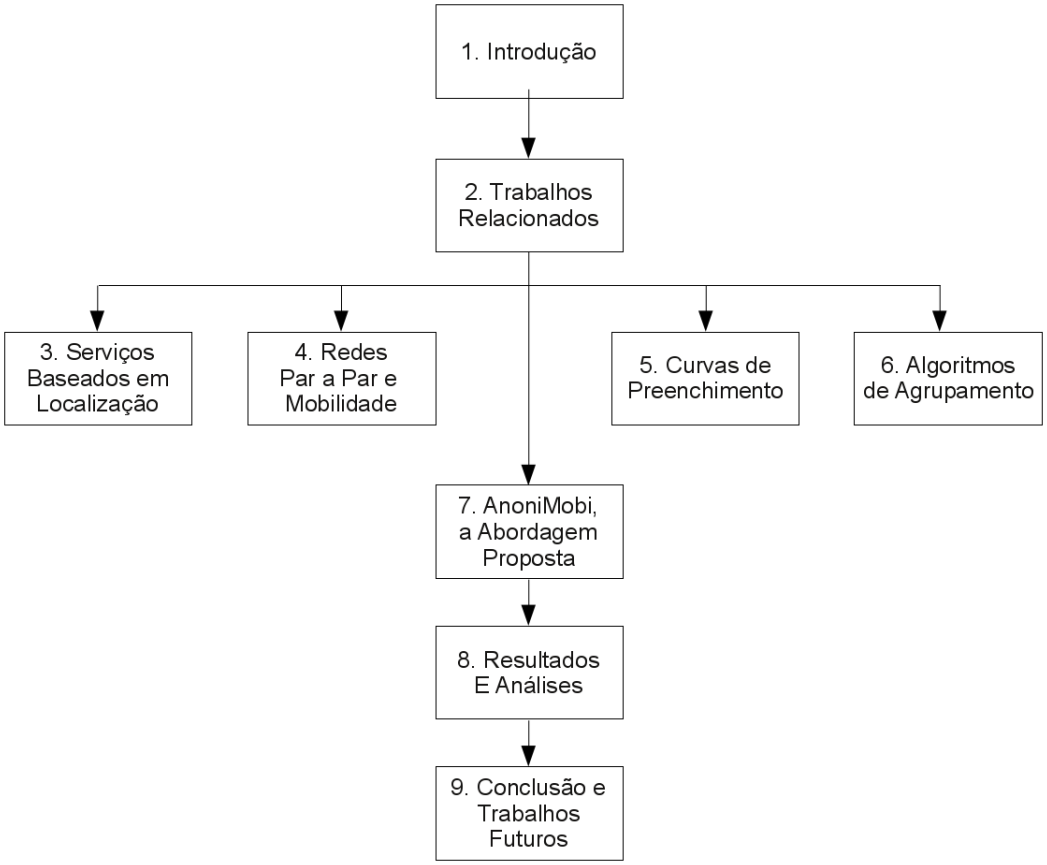


Figura 1: Dependência de Leitura

2 Trabalhos Relacionados

"*Na natureza, nada se cria, nada se perde, tudo se transforma.*"
Antoine Laurent Lavoisier (1743–1794)

A proliferação de aparelhos móveis com capacidades de localização vem pressionando o desenvolvimento de aplicações de SBL. Contudo, para que tais aplicações tenham sucesso, a privacidade e a confidencialidade são aspectos essenciais a serem considerados.

O presente capítulo relaciona alguns dos trabalhos mais relevantes no contexto da pesquisa desenvolvida, trazendo as ideias-chave das propostas e abordagens vinculadas ao universo da anonimização de informações de localização.

Preliminarmente, o conceito de *anonimidade-K* é explicitado para adequada compreensão das técnicas citadas.

2.1 Anonimidade-*K*

Seja um cenário com uma determinada base de dados, como a de um hospital ou a de um banco, que possui uma coleção privada de dados estruturados sobre pessoas específicas. Suponha-se que exista a necessidade de compartilhar uma *versão* dos dados. A questão natural que surge é como definir esta versão dos dados privados com garantia comprovada de que os indivíduos (sujeitos dos dados) não poderão ser identificados, mantendo, ao mesmo tempo, alguma utilidade prática para a versão fornecida para pesquisa. A proposta em [Sweeney 2002], conhecida como *anonimidade-K*, busca resolver essa problemática apresentando um modelo de proteção acompanhado de um conjunto de políticas de implementação.

Uma versão da base de dados, contendo informações sobre *raça*, *data de nascimento*, *sexo* e *código de endereçamento postal* (CEP), não consegue, em muitos casos, garantir a privacidade dos dados. Tais informações podem ser relacionadas a outras informações publicamente disponíveis para (re)identificar os indivíduos e inferir informações que deveriam ser ocultadas.

Resumidamente, um conjunto de dados é dito K-anônimo se cada um de seus registros é indistinguível dentre, pelo menos, outros $K - 1$ registros com relação a determinados atributos.

[Samarati 2001] apoia-se na definição da anonimidade-K para desenvolver uma solução baseada em técnicas de *generalização* e *supressão* de informações que consegue solucionar o problema de proteger a privacidade em versões de informação de base de dados. A abordagem revela *microdados* de tal forma que as identidades dos indivíduos não podem ser descobertas.

2.2 Consultas anônimas a SBLs públicos

O conceito de anonimidade-K também vem sendo utilizado no contexto de SBLs. A definição de [Pfitzmann e Köhntopp 2001] para anonimidade diz que "Anonimidade é a qualidade de um elemento estar não identificável em meio a um conjunto de elementos, o *conjunto de anonimidade*". Apoiado nessa definição e inspirado nos trabalhos de [Samarati e Sweeney 1998], [Gruteser e Grunwald 2003] considera que um usuário está K-anônimo, com respeito a informações de localização, se e somente se a informação de localização apresentada é indistiguível de outras $K - 1$ informações de localização de

usuários. O trabalho desenvolve um algoritmo baseado em *quadtree* que subdivide a área ao redor da posição do usuário em quadrantes até que o número de usuários dentro da (sub)área atinja um valor inferior a um k_min (parâmetro global que indica o nível de anonimidade). O quadrante que mantém o nível k_min é então retornado.

Em seguida, as coordenadas do quadrante (retângulo) são enviadas ao SBL que retorna ao sistema, como resultado, as coordenadas dos pontos de interesse mais próximos. A Figura 2 ilustra uma pesquisa feita pelo usuário u_1 em busca do hospital mais próximo de sua atual localização, requerendo um nível $K = 3$ de anonimidade. O quadrante definido engloba u_1, u_2 e u_3. A resposta do SBL indica h_3 (interno ao retângulo), h_2 e h_4.

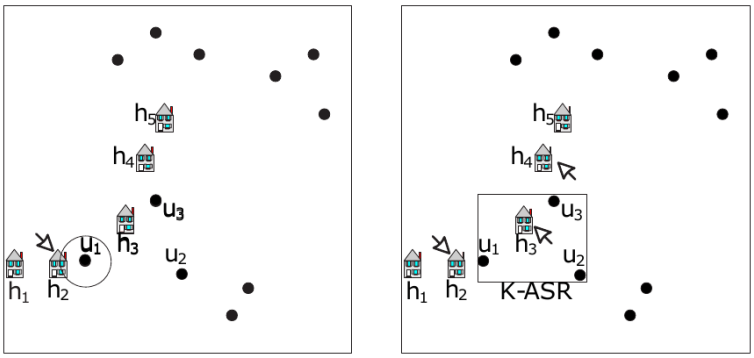


Figura 2: Pesquisa pelo hospital mais próximo

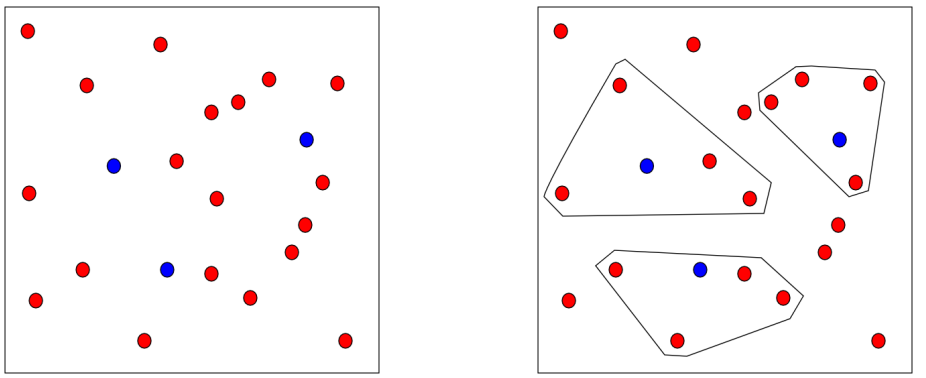
Este pioneiro trabalho no uso do conceito de anonimidade-K no contexto de pesquisas a SBLs tem uma desvantagem digna de nota. Ao assumir um valor global para k_min, o tempo de atendimento a requisições (qualidade de serviço) tem seu limite inferior "pré-fixado", mesmo para o caso dos usuários móveis cujos requisitos de privacidade sejam atendidos por valores inferiores a k_min. Em situações mais realísticas, os usuários tendem a definir diferentes níveis de anonimidade para pesquisas distintas.

Em [Gedik e Liu 2005], o nível de anonimidade-K é uma escolha do usuário e a localização dos usuários é ocultada através da construção de uma *Região Espacial de Anonimização-K* (*Anonymous Spatial Region*, K-ASR) que engloba a localização do usuário consultante mais $K - 1$ localizações de outros usuários. Um servidor confiável (*anonimizador*), conhecendo a localização de todos os usuários do sistema, consegue determinar a K-ASR relativa a um usuário e encaminhar sua consulta anonimamente ao SBL público. Quando o SBL retorna o resultado, o servidor reencaminha-a ao usuário de origem.

Apesar de possibilitar que o usuário escolha o nível de anonimidade desejado, os algoritmos desenvolvidos não são escaláveis, sendo adequados apenas para pequenos valores de K (entre 5 e 10), o que limita muito sua aplicabilidade.

[Mokbel, Chow e Aref 2006] propõe uma abordagem mais escalável, cuja arquitetura do servidor *anonimizador* é baseada em uma pirâmide que armazena uma lista de localizações dos usuários do sistema. O usuário consultante fornece ao anonimizador o valor de K e, adicionalmente, um valor A_{min} que indica a resolução mínima aceitável para a região de anonimização (útil, segundo os autores, em casos de regiões densamente populadas). Novamente, o anonimizador intermedia a comunicação com o SBL público, retornando ao usuário o resultado obtido.

Contudo, a forma de estabelecimento da K-ASR considerando usuários efetivamente mais próximos do usuário consultante (empregada em todas as abordagens citadas) falha ao tentar ocultar sua localização em muitos casos, pois há uma tendência muito forte de que o usuário situe-se no centro dos demais (vide Figura 3). Sendo este um subsídio para determinar a localização do usuário, outras métricas de proximidade precisam ser consideradas. Acrescente-se a este, o fato de que há demasiado esforço computacional para o cálculo dos vizinhos mais próximos de um usuário.



(a) Usuários consultantes em azul

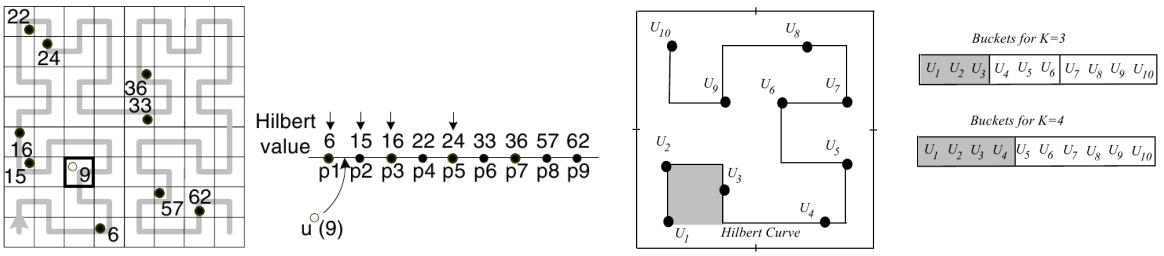
(b) Agrupamentos de usuários mais próximos para $K = 5$

Figura 3: Tendência de centralização do usuário consultante

Em [Kalnis et al. 2006], uma métrica de proximidade alternativa é adotada. Baseada na curva de preenchimento espacial de Hilbert (vide capítulo 5), os valores de coordenadas dos usuários são mapeados para um valor no espaço de Hilbert e agrupados adequadamente para garantir o nível K de anonimidade solicitado.

A Figura 4(a) apresenta usuários distribuídos sobre um espaço bidimensional mapeado (curva de Hilbert em cinza). A lista de proximidade ao lado da figura ilustra, por exemplo, a situação do usuário $p1$ em relação ao usuário $p8$: ambos encontram-se geograficamente próximos; porém, seus valores de Hilbert (6 e 57, respectivamente) distanciam-nos na lista. Na óptica da curva de preenchimento, $p1$ (6) está mais próximo de $p3$ (16) do que de $p8$ (57). Ao ser inserido no contexto, o usuário u (9) irá situar-se entre $p1$ e $p2$, dado

seu valor de Hilbert.



(a) A curva de Hilbert altera a definição de proximidade entre usuários

(b) Agrupamentos de usuários próximos, conforme valores de Hilbert

Figura 4: A curva de Hilbert como métrica de proximidade

Além de diminuir o tempo de processamento na definição da K-ASR, esta abordagem contorna elegantemente o problema de centralização do usuário consultante em relação aos demais usuários que participam do processo de anonimização (grupo de anonimidade). A Figura 4(b) ilustra a definição de grupos de anonimização com $K = 3$ e $K = 4$ considerando a curva de Hilbert como métrica de proximidade.

Nota-se que estas abordagens tornam o servidor confiável um gargalo para o sistema, já que todas as requisições (e atualizações de posições dos usuários) passam por ele. Além disso, este servidor central é um ponto extremo de falha: se comprometido, revela a posição de todos os usuários, bem como suas consultas.

Em [Ghinita, Kalnis e Skiadopoulos 2007], é apresentado o sistema Privé que busca resolver as problemáticas apresentadas. Utilizando-se da curva de Hilbert para estabelecimento de uma nova métrica de proximidade, a abordagem desvencilha-se do servidor anonimizador distribuindo as informações de localização entre os diversos usuários P2P do sistema. Os usuários auto-organizam-se em grupos cujos líderes mantêm as informações de localização de seus integrantes (a liderança é alternada periodicamente para efeito de balanceamento de carga de processamento).

Os líderes, por sua vez, trocam informações entre si para estabelecer os grupos de anonimização quando uma consulta deve ser realizada. Entretanto, em vez de utilizar um *anonimizador*, o usuário consultante coleta as coordenadas de seus vizinhos com o auxílio de seu líder e, por conta própria, constrói a K-ASR desejada. Em seguida, envia as coordenadas da mesma para o SBL, obtendo o resultado esperado sem quaisquer intermediários.

Um conceito bastante pertinente que é discutido em Privé diz respeito à *Reciprocidade* das consultas. A Figura 5 ilustra a ideia. Supondo-se uma pesquisa com nível $K = 3$

de anonimidade, o conjunto de usuários a participarem do processo de anonimização caso o usuário u_1 efetue a consulta será u_1, u_2, u_3. O mesmo grupo será definido caso a pesquisa seja efetuada pelos usuários u_2 e u_3. Entretanto, se o usuário u_4 for o usuário consultante, o grupo de anonimização será u_1, u_3, u_4. Nota-se que para os três primeiros casos, o retângulo que consitui a K-ASR será o mesmo e envolverá apenas os usuários já indicados. No quarto caso o retângulo será diferente. Nesta situação, supondo o pior caso, em que o atacante conhece as coordenadas de todos os usuários do sistema (vide seção 1.1), será possível concluir que o autor da consulta é u_4.

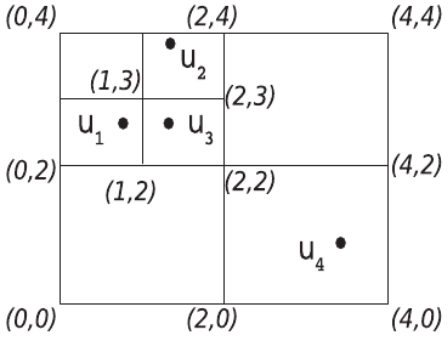


Figura 5: Ausência de *reciprocidade* permite inferência do usuário consultante

Dito isso, Privé define a *Reciprocidade* K-ASR da seguinte forma:

Definição 2.1 *Seja usuário u_q realizando uma consulta e seja A_q sua K-ASR associada. A_q satisfaz a propriedade de reciprocidade caso exista um conjunto AS de usuários internos a A_q, tais que (i) $|AS| \geq K$, (ii) $u_q \in AS$ e (iii) todo usuário $u \in AS$ é interno à K-ASR de todos os outros usuários em AS.*

A Figura 6 ilustra o exemplo de reciprocidade para as K-ASR A_1 e A_2 construídas segundo esse critério, considerando $K = 5$.

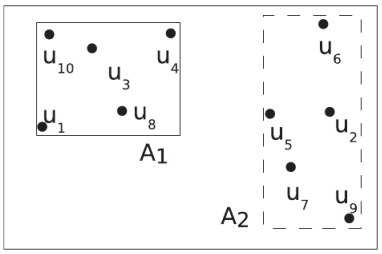


Figura 6: Definição de K-ASR com *reciprocidade* entre usuários

Em Privé, os líderes de cada grupo organizam-se hierarquicamente em supergrupos com o objetivo de compartilharem as informações necessárias para a construção das K-ASR solicitadas. Essa auto-organização em árvore B+ prossegue até a definição de um

líder *raiz*. Dependendo do nível K solicitado e da quantidade de usuários no sistema, a troca de mensagens entre líderes pode atingir a raiz.

A organização dos usuários P2P em árvore B+, apesar de elegante, tem a desvantagem de sobrecarregar os líderes, principalmente o líder *raiz*, elevando o nível de rejeição de consultas por *buffer overflow*. Com cerca de 10.000 usuários, o tempo de resposta pode chegar a 10 minutos [Ghinita, Skiadopoulos e Kalnis 2007], o que torna impraticável a efetuação de consultas corriqueiras.

A proposta de [Ghinita, Skiadopoulos e Kalnis 2007] busca resolver este problema de sobrecarga. Adotando os mesmos elementos anteriormente mencionados (curva de Hilbert e distribuição de informações de localização entre usuários P2P), é definido o sistema MobiHide para consultas anônimas. Contudo, a infra-estrutura de organização dos nós é baseada no protocolo Chord (vide 4.3.2). A chave para a organização do anel Chord que armazena os valores da tabela de Hash distribuída (DHT) é o próprio valor de Hilbert. Esta organização proporciona maior escalabilidade para o sistema. Uma consulta-K é atendida em $O(K)$ saltos (entre líderes).

Há três operações recorrentes na estrutura. A latência [1] da operação de *entrada/saída* de um nó com relação a um determinado grupo tem ordem de complexidade de $O(\log N - \log \alpha)$, com α sendo um parâmetro que define a quantidade máxima de usuários por grupo e com N indicando a quantidade de usuários do sistema. O custo de comunicação para a operação de entrada/saída é $O(\log N - \log \alpha + \alpha)$. A operação de *realocação* possui os mesmos valores de complexidade. A complexidade da *consulta-K*, tanto em termos de latência quanto de custo de comunicação, é de $O(K/\alpha)$.

2.3 Posicionamento

Todas as abordagens pesquisadas apresentam como retorno um conjunto de pontos de interesses próximos à K-ASR definida. Contudo, saber a posição de um *alvo* nem sempre é suficiente para conseguir alcançá-lo com eficiência. Em muitos casos, faz-se necessário o fornecimento de orientações adicionais (rotas).

A proposta do presente trabalho busca atingir esta meta, sem contudo incorrer em problemas de latência, falta de reciprocidade, sobrecarga de servidor centralizado e fornecimento de subsídios a atacantes internos e externos ao sistema.

[1] Neste contexto, a latência é avaliada em termos da quantidade de nós pelos quais as mensagens precisam trafegar.

3 Serviços Baseados em Localização

"Navegar é preciso, viver não."
Luís de Camões

"Serviços Baseados em Localização (SBLs) podem ser definidos como serviços de TI que criam, compilam, selecionam ou filtram informação considerando as localizações dos usuários ou de outras pessoas ou objetos móveis" [Küpper e Treu 2010]. Podem atuar em conjunto com serviços convencionais de telefonia, por exemplo, para realizar roteamento de chamadas baseado em localização. Um dos atrativos dos SBLs resulta do fato de que seus usuários não precisam inserir manualmente a informação de localização, mas são automaticamente "apontados" e "rastreados".

Uma vez obtida a informação de localização, ela deve ser tratada de vários modos, incluindo a adequação de seu formato ao de algum outro sistema de referência espacial, a correlação com outros conteúdos geográficos, a geração de mapas, ou o cálculo de instruções de navegação, dentre outros. Em geral, tais tarefas não são processadas nos aparelhos móveis. Desta forma, o sucesso no processo de utilização de SBLs depende da interoperação cooperativa de diversos atores, tais como operadoras de redes e provedores de serviço e conteúdo, atuando em uma infraestrutura distribuída. Esta situação impõe uma série de desafios, seja na troca em tempo-real das informações de localização entre os atores, seja na garantia da privacidade dos indivíduos a que se referem essas informações.

Este capítulo expõe fundamentos relativos aos SBLs e introduz conceitos relacionados à essa tecnologia e a outras afins, proporcionando uma compacta e compreensiva visão geral dos métodos, protocolos de localização e plataformas de serviços existentes.

3.1 Breve Introdução

Apesar de os SBLs já serem um tema explorado no campo das comunicações móveis há alguns anos, ainda não existe uma terminologia comum para defini-los. Os termos *serviço baseado em localização*, *serviço consciente de localização*, *serviço referente a localização*, e *serviço de localização* [1] são comumente utilizados intercambiavelmente.

> *"Uma razão para este dilema pode repousar no fato de que o caráter e a aparência de tais serviços têm sido determinados por diferentes comunidades, especificamente o setor de telecomunicações e a área de computação ubíqua"* [Kütter 2005].

Uma das principais origens dos SBLs é o documento E911 (*Enhanced 911*) [FCC 2010], emitido pela *Federal Communications Commission* (FCC) em 1996, que obrigou as operadoras das redes de celular a localizar as chamadas telefônicas destinadas aos serviços de emergência (911) e direcioná-las para pontos de resposta geograficamente mais próximos da origem das chamadas. Como consequência, as operadoras de celular em todo o mundo passaram a oferecer uma série de SBLs na forma de buscadores de serviços (como hospitais, postos de abastecimento e terminais eletrônicos de auto-atendimento).

Com o barateamento na produção de receptores de sinais GPS, esta tecnologia passou a dominar o mercado de geoposicionamento, estando comumente embutida em aparelhos celulares de baixo custo. Este cenário torna o uso de SBLs ainda mais promissor.

3.2 Uma Taxonomia

De um ponto de vista funcional, os SBLs podem ser classificados de acordo com os seguintes critérios [Küpper e Treu 2010]:

Auto-referência *versus* **Referência Cruzada.** Este critério diz respeito ao papel que o usuário e o alvo participantes de um SBL adotam durante sua execução. O *usuário* é a pessoa que solicita algum SBL, ao passo que o *alvo* é o indivíduo a ser localizado ou rastreado. Em SBLs de auto-referência, usuário e alvo são o mesmo indivíduo. Um exemplo é a busca por pontos de interesse próximos a localização atual do usuário. Em SBLs de referência cruzada, os papéis de usuário e alvo são exercidos por indivíduos

[1] Do inglês *location-based service*, *location-aware service*, *location-related service* e *location service*, respectivamente.

distintos. O usuário solicita a localização do alvo (ou seu rastreamento). Esse tipo de serviço é uma demanda, por exemplo, de empresas de transporte, que precisam saber constantemente a posição de seus veículos.

Alvo-único *versus* **Multi-alvos.** SBLs de alvo-único focam-se no rastreamento de um único alvo. Normalmente, a posição do alvo é inter-relacionada com conteúdo geográfico para, por exemplo, criar dados de rota para navegação. Neste caso, o principal objetivo é converter uma posição geográfica (latitude, longitude e altitude) em localização descritiva (ex.: endereços de ruas) relacionada ao local onde se encontra o alvo. Um exemplo típico é o serviço de rastreamento de crianças. Em SBLs de multi-alvo, o foco está mais direcionado para o inter-relacionamento das posições de vários alvos entre si. O estabelecimento de redes sociais baseadas em proximidade geográfica é uma das aplicações deste tipo de SBL, em sintonia com a filosofia *webware* (ou Web 2.0) [Anderson 2007], [Sá e Bertocchi 2007] e [Wikipédia 2010].

Reatividade *versus* **Proatividade.** Um SBL é classificado como reativo se a informação baseada em localização é entregue ao usuário quando uma consulta é formalmente efetuada. Neste cenário, um usuário efetua uma consulta ao SBL que, após o devido processamento, retorna o resultado baseado em localização. As solicitações por informações relacionadas ao trânsito em rodovias são situações facilmente tratadas por esta categoria de SBLs. Por outro lado, SBLs proativos são acionados automaticamente na presença de algum evento pré-definido. Se o alvo aproxima-se de uma determinada posição, por exemplo, o SBL reage produzindo algum outro evento (um guia turístico eletrônico pode notificar via SMS um turista tão logo este se aproxime de algum ponto de interesse).

Localização de Alvos *versus* **Alvos em uma Localização.** Duas ópticas "simétricas" de localização podem ser definidas para SBLs com relação aos alvos. Na primeira, ocorre um mapeamento do conjunto de alvos para o conjunto de localizações, quando se diz que o SBL fornece um serviço de *localização de alvos* (dado um alvo, retorna-se sua localização). Na outra, o mapeamento é reverso, isto é, de conjuntos de localizações para conjuntos de alvos. Diz-se, nesse caso, que o SBL fornece um serviço de informações sobre *alvos em uma localização* (dada uma localização, retorna-se o conjunto de alvos mais próximos). Um exemplo de localização de alvos é o rastreamento de um indivíduo, onde a posição do alvo é requerida. A pesquisa pelas farmácias mais próximas de uma determinada localização é um exemplo de pesquisa por alvos em uma localização.

Ambiente Externo *versus* **Ambiente Interno.** A distinção entre essas duas classes é essencial para a definição da tecnologia de suporte a ser utilizada. Os SBLs de

ambiente externo fazem uso de tecnologias de posicionamento via satélites ou celulares, cujas precisões de resultado podem variar de 10 m (GPS)[2] a algumas centenas de metros (ex.: métodos da rede celular). Já os SBLs de ambiente interno baseiam-se na observação de padrões de rádio e na sua comparação com padrões pré-gravados e bem-definidos de localização, chegando a obter precisões na ordem de centímetros [3].

3.3 Algumas Aplicações

Existe uma série de diferentes Serviços Baseados em Localização. Sem pretender esgotar as possibilidades, eis alguns exemplos categorizados [Steiniger, Neun e Edwardes 2006]:

- Navegação
  - Direções
  - Roteamento em ambiente interno
  - Orientações de estacionamento
  - Gerenciamento de tráfego
- Informação
  - Serviços de Informação
  - Guias turísticos e de viagens
  - Planejamento de viagens
  - Páginas amarelas móveis
  - Guias de compras
- Rastreamento
  - Rastreamento de pessoas/veículos
  - Rastreamento de produtos
- Jogos
  - Jogos móveis

[2] O valor de 10 m é uma aproximação do valor de referência estabelecido no documento *GPS SPS Performance Standard* [NavStar 2008] de erro $\leq$ 12,8 m 95%.

[3] [Kolodziej e Hjelm 2006] apresenta um estudo pormenorizado sobre SBLs de ambiente interno e [Kütter 2005] aprofunda o estudo sobre SBLs de ambiente externo.

  - *Geocaching* [4]

- Emergência
  - Chamadas de emergência
  - Assistência automotiva
- Propaganda
  - *Banners*
  - Alertas de propagandas
- Bilhetagem
  - Pedágio de avenidas
  - Bilhetagem sensível à localização
- Gerenciamento
  - de facilidades
  - de infraestrutura
  - de relacionamento com clientes
  - de frota (agendamento)
  - de ambientes
  - de segurança (polícia, ambulância, ...)
- Lazer
  - Busca por amigos [Santos et al. 2008]
  - Mensagens instantâneas

Uma situação típica é um serviço que localiza os restaurantes próximos a localização atual do usuário. Este tipo de serviço pode levar em consideração as preferências desse usuário, como o tipo de comida favorita, a faixa de preço adequada, etc.

A Tabela 1 apresenta algumas ações comumente executadas por usuários móveis, bem como as relações dessas ações com as ações de contexto espacial [Reichenbacher 2004]. A

[4] *Geocaching* é uma espécie de "Caça ao Tesouro Geodésica", com diversos adeptos no mundo inteiro [Geocaching 2010].

Tabela 1: Ações elementares de usuários móveis e suas relações espaciais

	Questões	**Objetivo**	**Operações**	**Serviço**	**Parâmetros**	**Suporte**
Orientação e Localização *Localização*	Onde eu estou? Onde está { alguém | algo }?	Localizar pessoas e objetos	- posicionamento - geocodificação - geodecodificação	entrega da posição de pessoas e objetos	coordenadas objetos endereços nome de lugares	orientação no espaço
Navegação *Navegação, Planejamento de rotas*	Como chego a { nome de lugar | endereço | *xy* }?	Encontrar o caminho para um destino	- posicionamento - geocodificação - geodecodificação - roteamento	entrega de rotas e instruções de navegação	ponto inicial, ponto final, e pontos intermediários	definição de caminhos através do espaço
Busca *Busca por pessoas e objetos*	Onde está o { objeto | indivíduo & } mais { próximo | relevante }?	Procurar por pessoas e objetos segundo um critério	- posicionamento - geocodificação - cálculo de distância/área - descoberta de relações	descoberta de serviços disponíveis; descoberta de pessoas/objetos	localização área/raio objeto/categoria	descoberta de objetos relevantes; descoberta de pessoas
Identificação *Identificação e reconhecimento de pessoas*	{ O que | Quem | Quanto } está { aqui | lá }?	Identificar pessoas e objetos; Quantificar objetos	- listagem - seleção - busca temática/ espacial	entrega de informação sobre pessoas/ objetos	objeto	informação sobre objetos do mundo real no contexto apresentado
Verificação de Evento *Verificação de eventos; determinação do estado de objetos*	O que acontece { aqui | lá }?	Saber o que acontece; Conhecer o estado de objetos		entrega de informação sobre objetos e eventos	tempo localização objeto	descoberta de eventos; informação sobre o estado de objetos no contexto

coluna "Operações", que resume algumas das atividades que o provedor do SBL deve suportar, faz referência constante à atividade de localizçaão geográfica. Apesar de muitas pessoas afirmarem estar familiarizadas com o conceito de localização, é útil analisar mais profundamente o assunto e distinguir as diferentes categorias de informação de localização existentes. A próxima seção aborda esta questão.

3.4 A Respeito da Localização

Basicamente, o termo "localização" está associado a um certo lugar no mundo real. Este tipo de localização pertence à classe de *localizações físicas*. Entretanto, a Internet criou uma série de novas aplicações para as quais o termo "localização" muitas vezes adquire um novo significado, referindo-se a lugares virtuais. Tem-se, assim, o domínio das *localizações virtuais* [5]. Esta última classe de localizações, em geral, não é abordada por SBLs. Três subcategorias de localizações físicas podem ser distinguidas.

Localizações Descritivas. Localizações descritivas são um conceito fundamental utilizado pelas pessoas para agendar compromissos e receber/enviar correspondências, por exemplo. Uma localização descritiva sempre refere-se a objetos geográficos naturais como territórios, montanhas, rios e lagos, ou a objetos geográficos produzidos pelo homem como cidades, rodovias, edifícios e salas. Estes objetos (naturais ou artificias) são citados através de descrições, isto é, nomes, identificadores ou números.

Localizações Espaciais. Esta subclasse de localização representa um único ponto no espaço Euclideano representado por coordenadas. Diferentemente das localizações descritivas, as localizações espaciais não são utilizadas frequentemente no cotidiano da maioria das pessoas por preferir-se, em geral, adotar uma orientação em termos de objetos geográficos a uma orientação por coordenadas. Entretanto, o conceito de localização espacial proporciona a base para o mapeamento de localizações descritivas.

Localizações de Rede. As localizações de rede referem-se à topologia de uma rede de comunicações, como a Internet ou o sistema de celular. Tais redes são compostas por *sub-redes* que se interconectam hierarquicamente através de circuitos de comunicação. O endereço de rede fornece informações de roteamento ao provedor de serviço de rede, indicando a localização do usuário dentro da topologia existente. Em redes móveis, uma localização de rede indica a estação-base à qual se está vinculado em determinado instante.

SBLs podem se basear nessas três categorias de localização. O alvo de um usuário de

[5] Um exemplo são as localizações "virtuais" no ambiente *Second Life* [SecondLife 2010]

SBL é indicado pelo processo de posicionamento, para cuja execução existem diferentes métodos. GPS, por exemplo, fornece uma localização espacial, ao passo que o método *Cell-Id* produz uma localização de rede (a combinação de ambas técnicas é possível). Uma vez que a localização de um alvo é obtida, é necessário refiná-la para processamento pelo SBL, o que significa obter uma localização descritiva com sentido prático para ser entregue ao usuário.

3.5 O Problema do Posicionamento

Posicionamento é o processo de obtenção da posição espacial de um alvo. Existem muitos métodos para efetuar esta tarefa e, em geral, consideram-se os seguintes elementos:

- um ou muitos parâmetros obtidos por métodos de medição,
- um método de posicionamento para cálculo de posição,
- um sistema descritivo ou de referência espacial,
- uma infraestrutura, e
- protocolos para coordenar o processo de posicionamento.

A função principal de qualquer sistema de posicionamento é a medição de um ou vários parâmetros observáveis como, por exemplo, ângulos, distâncias, direções ou velocidades. Geralmente, estes parâmetros refletem uma relação espacial do alvo com um ou vários ponto(s) fixo(s) de referência descrito(s) por coordenada(s) bem conhecida(s). São comumente medidos utilizando-se fundamentos físicos de sinais de rádio, infra-vermelhos ou ultra-sonoros (velocidade e atenuação, dentre outros).

Após a obtenção dos valores dos parâmetros, a localização do alvo deve ser calculada considerando-se as coordenadas dos pontos fixos através de algum método de posicionamento. A seguir, uma classficação básica de tais métodos é apresentada [Kütter 2005].

Sensoriamento de Proximidade (*Proximity Sensing*). Neste método, a posição de um alvo é derivada das coordenadas da estação-base que ou recebe o sinal do terminal (por exemplo, um aparelho celular) ou da qual o terminal recebe o sinal. Nos sistemas celulares, este método é conhecido pelos termos *Célula de Origem* (*Cell of Origin - CoO*), *Identidade Global de Célula* (*Cell Global Identity - CGI*) ou simplesmente por *Cell-ID* e se tornou bastante popular graças a sua facilidade de implementação. Entretanto, sua maior

desvantagem é a baixa precisão devido ao raio de alcance das células, que pode variar de 100 m nas áreas urbanas a dezenas de quilômetros nas regiões rurais.

Lateração. Este método mede a distância entre o alvo e um número de pelo menos três estações-base. Tais medidas são inseridas em um sistema de equações não-lineares e a posição do alvo é calculada. O método é conhecido como *trilateração* quando o número de estações-base é exatamente três. Existem, para este método, as variantes *lateração circular* [6] e *lateração hiperbólica* [7]. A questão aqui é como medir as distâncias, já que estas medidas estão sempre sujeitas a erros de diversas naturezas (incluindo a reflexão em obstáculos no caminho entre o terminal e a estação-base), o que promove um certa diferença em relação às distâncias exatas.

Angulação. Também conhecida por Ângulo de Chegada (*Angle of Arrival - AoA*) e Direção de Chegada (*Direction of Arrival - DoA*), este é outro método para estimar a posição do alvo considerando as coordenadas de algumas estações-base. Diferentemente da lateração, o parâmetro observado é o ângulo entre o alvo e uma estação-base. O ângulo do sinal de chegada é medido na estação-base e, assim, a posição do alvo fica restringida ao longo de uma reta que intercepta tanto o alvo quanto a estação-base. Considerando o ângulo de uma segunda estação-base, outra reta é definida, e a interseção das duas retas indica a posição do alvo. Assim, teoricamente, apenas duas estações-base são suficientes para obter uma posição num espaço bidimensional. Entretanto, uma má resolução dos sinais produz uma aproximação grosseira dos valores dos ângulos medidos, forçando a utilização de uma terceira estação-base para definir um resultado adequado. Consequentemente, na prática, medidas de ao menos três estações-base são recomendadas para compensar os erros [8].

Cálculo Deduzido. A posição de um alvo pode ser deduzida ou extrapolada a partir de sua última posição conhecida, assumindo que a direção do movimento e/ou a velocidade do alvo ou a distância percorrida são conhecidos. Este método é conhecido como "Cálculo Deduzido" ou "*Dead Reckoning*" (abreviação do termo original "*Deduced Reckoning*") [9]. Obviamente, o ponto crucial no cálculo deduzido é obter a posição inicial, a direção do movimento bem como a distância e a velocidade do alvo. Para a obtenção de cada uma desta grandezas existem diversos métodos e, atualmente, o método é utilizado

[6] Métodos de posicionamento baseados em lateração circular em combinação com medidas de tempo são usualmente designados pelo termo Tempo de Chegada (*Time of Arrival - ToA*).

[7] Nas redes celulares, a lateração hiperbólica é conhecida pelo termo Diferença de Tempo de Chegada *(Time Diffenrece of Arrival - TDoA).*

[8] Tem-se, com isso, o popular termo *Triangulação.*

[9] Também há referências ao método "*Dead Reckoning*" como "Navegação Inercial" (*Inertial Navigation*) [Kolodziej e Hjelm 2006].

para refinar informações de posição fornecido por GPS para sistemas de navegação veicular com o intuito de manter o fornecimento de instruações de navegação mesmo na ausência do sinal de satélite. No contexto de SBLs, o cálculo deduzido pode ser utilizado, por exemplo, para otimizar a frequência de atualizações de posição a ser processada em um servidor externo, como explica [Walther e Fischer 2002].

Correspondência de Padrão. O processo de posicionamento também pode ser realizado por meio de correspondência de padrões. O objetivo principal dessa técnica é observar um "cenário" onde o posicionamento está sendo aplicado e "tirar conclusões" sobre a posição de um alvo a partir destas observações. Em sua versão óptica [10] (também conhecida como *análise de cenário* [Barleze 2003]), imagens visuais de um cenário são geradas pela câmera de um observador e comparadas umas com as outras. A posição do alvo pode, então, ser obtida pela comparação das imagens geradas a partir de diferentes posições e diferentes perspectivas, sendo que o alvo pode ser o próprio observador em questão. O reconhecimento de padrões necessário é, em muitos casos, realizado sob uma abordagem assistida por terminal [11], na qual o terminal efetua as medidas (captura de imagens) e transfere os resultados para a rede calcular a posição. A vantagem desta abordagem reside no fato de que, em muitos sistemas, pouca ou nenhuma modificação precisa ser feita nos terminais.

Basicamente, é possível implementar qualquer combinação dos métodos de posicionamento apresentados, considerando diversos e diferentes parâmetros para medição com relação a uma posição fixa. A maior motivação para fazer isso é incrementar a precisão e, consequentemente, o resultado a ser retornado durante uma consulta a um SBL, por exemplo. A propósito, uma baixa precisão pode resultar de erros nos parâmetros medidos, cujas origens repousam em diversas fontes, tais como: falta de sincronia de relógios, refrações ionosféricas e troposféricas, ausência de visada direta, propagação por caminhos múltiplos (distorção de sinais), dificuldades de acesso ao meio de transmissão, coordenadas de estação-base imprecisas e má distribuição geométrica das estações-base. Geralmente, estas fontes de erro contribuem em diferentes níveis e magnitudes nos diferentes métodos e infraestruturas de posicionamento.

A Tabela 2 [Kütter 2005] apresenta uma visão geral dos diferentes métodos de posicionamento implementados em alguns sistemas e compara os seus modos de operação (at = assitido por terminal, bt = baseado em terminal, br = baseado em rede), o tipo de sinais

[10] Na versão não-óptica, o observador coleta e relaciona a potência dos sinais recebidos (*received signal strength - RSS*), formando um vetor de tais valores chamado *fingerprint*.

[11] Os metodos de posicionamento podem ser implementados sob três modos de operação distindos: assistido por terminal, baseado em terminal e baseado em rede.

utilizados, os parâmetros observados e o tipo de rede na qual o método é utilizado.

Tabela 2: Visão geral de posicionamento: Satélite, Celular e Ambiente Interno

Nome	**Método básico**	**Modo**			**Tipo de sinal**	**Medida**	**Tipo de rede**
		at	**bt**	**br**			
Posicionamento em Satélites							
GPS	Lat. Circular		x		Rádio	Tempo	
D-GPS	Lat. Circular		x		Rádio	Tempo	
Galileo	Lat. Circular		x		Rádio	Tempo	
Posicionamento em Rede Celular							
Cell-Id	Sens. Proxim.			x	Rádio	Cell-Id (+RTT)	GSM
E-OTD	Lat. Hiperb.	x	x		Rádio	tempo	GSM
U-TDoA	Lat. Hiperb.			x	Rádio	Tempo	GSM
Cell-Id	Sens. Proxim. (+Angulação)			x	Rádio	Cell-Id (+RTT+AoA)	UMTS
OTDoA	Lat. Hiperb.	x	x		Rádio	tempo	UMTS
E-FLT	Lat. Hiperb.			x	Rádio	Tempo	CDMA*
A-FLT	Lat. Hiperb.	x	x		Rádio	tempo	CDMA*
A-GPS	Lat. Circular	x	x		Rádio	tempo	todas
Posicionamento em Ambiente Interno							
RADAR	Corr. Padrão			x	Rádio	RSS	WLAN
Ekahau	Corr. Padrão	x			Rádio	RSS	WLAN
Indoor GPS	Lat. Circular		x		Rádio	Tempo	
RFID	Sens. Proxim.	x	x	x	Rádio	ID	
ActiveBadge	Sens. Proxim.			x	Infra-verm.	ID	
WIPS	Sens. Proxim.	x			Infra-verm.	ID	WLAN
ActiveBat	Lat. Circular			x	Ultra-som	Tempo	418 MHz (rádio)
Cricket	Sens. Proxim.		x		Ultra-som	ID+Tempo	

* Neste contexto, CDMA refere-se à tecnologia *cdmaOne/2000*.

4 Redes Par-a-Par e Mobilidade

"*A rede é o computador.*"
Sun Microsystems

O termo "par-a-par" (*peer-to-peer - P2P*) refere-se a um classe de sistemas e aplicações que emprega recursos distribuídos (processamento, armazenamento, informações e largura de banda, dentre outros) para desempenhar uma função de forma descentralizada.

A "tecnologia" P2P ganhou visibilidade a partir de 1999 com o emergente compartilhamento de músicas através do programa Napster [Carlsson e Gustavsson 2001], quando passou a ser empregada também como uma importante técnica em diversas áreas, tais como a computação distribuída e colaborativa, incluindo seu uso em redes *ad hoc*, e até mesmo em sistemas de arquivo distribuídos [Dabek et al. 2001].

Este capítulo apresenta brevemente os principais tópicos relacionados aos Sistemas P2P, fornecendo subsídios para a compreensão de seu funcionamento, bem como apresentando detalhes de seus protocolos de comunicação.

Adicionalmente, uma abordagem do ponto de vista da mobilidade permeia os tópicos discutidos, realçando pontos relevantes para usuários móveis, que constituem uma parcela considerável do público alvo dos SBLs.

4.1 Definições, Benefícios e Desvantagens

Semelhantemente ao que acontece com o termo SBL, apresentado no capítulo 3, não há uma definição universalmente aceita para Sistemas P2P. Entretanto, pode-se considerar consenso a capacidade/finalidade de "compartilhamento de recursos". Algumas características comuns a todos os sistemas P2P podem ser apontadas:

- Um 'par' é um computador que pode atuar tanto como servidor quanto como cliente;
- Um sistema P2P é constituído de pelo menos dois pares;
- Os pares devem ser capazes de "trocarem recursos" entre si *diretamente*;
- Servidores dedicados (isto é, que não exercem o papel de clientes) podem estar presentes no sistema, porém apenas para auxiliar os clientes a descobrirem uns aos outros;
- Os pares podem entrar no sistema e/ou deixá-lo livremente;
- Os pares podem pertencer a diferentes donos.

Apesar das controvérsias, [Lv et al. 2002] apresenta a seguinte classificação para as arquiteturas P2P:

Centralizada. Um diretório de informações é constantemente atualizado em uma "localização central". Os nós da rede P2P consultam o servidor desse diretório central para descobrir em quais nós encontram-se as informações desejadas. Essa abordagem centralizada não é escalável e o diretório central apresenta-se como um ponto de falha.

Descentraliza, porém Estruturada. Esses sistemas não possuem um servidor de diretório central (arquitetura descentralizada), entretanto, sua estrutura bem definida permite um eficiente controle tanto da topologia da rede P2P quanto da localização dos recursos distribuídos por esta rede. Apesar de predominarem na literatura pesquisada, tais redes são praticamente inexistentes em termos de implementação efetiva. Assim, é difícil predizer como tais estruturas trabalhariam na presença de uma população de nós extremamente transitórios, comportamento que parece caracterizar os usuários de redes P2P.

Descentraliza e Desestruturada. Existem sistemas que não dependem de um diretório centralizado, nem tão pouco possuem qualquer controle preciso sobre a topologia

da rede e a localização dos recursos. A topologia obtida possui certas características, contudo a localização de recursos não se baseia no conhecimento da topologia. Para encontrar um recurso, um par consulta seus vizinhos, comumente por meio de inundação (*flooding*). Esse projeto não-estruturado é extremamente resistente à falta de constância dos usuários. Todavia, os mecanismos de pesquisa não são bem escaláveis, gerando altas cargas de processamento de troca de mensagens entre os participantes da rede [1].

Independentemente da arquitetura utilizada, os Sistemas P2P possuem vantagens e desvantagens. Alguns benefícios podem ser citados:

- A carga de trabalho é distribuída pelos pares;
- A inexistência do principal gargalo das redes, o servidor, permite o encurtamento do tempo para a conclusão de tarefas;
- As arquiteturas P2P conseguem aproveitar o potencial "ocioso" dos pares da rede;
- Controle e gerenciamento centralizados são desnecessários;
- A rede P2P é altamente escalável, dada a facilidade de inserção de pares em sua estrutura;
- Os usuários podem manter controle de seus recursos, entrando no sistema e deixando-o a qualquer instante.

Por outro lado, certas desvantagens são inerentes à filosofia de atuação distribuída em questão:

- Os pares são mais vulneráveis a ataques;
- É difícil estabelecer padrões para os Sistemas P2P;
- Para certas tarefas específicas não há a possibilidade de distribuição da carga de trabalho entre os pares;
- Uma rede P2P não consegue garantir que determinado recurso esteja sempre disponível, já que os usuários podem encerrar sua participação no sistema a qualquer instante;

[1] O trabalho de [Lv et al. 2002] apresenta alternativas mais escaláveis para os algoritmos de redes Descentralizas e Desestruturadas, focando nos aspectos de *pesquisa* e *replicação*.

- É difícil replicar informações de todos os participantes do sistema;
- É difícil evitar o tráfego ilegal de informações protegidas por direitos autorais;
- Sistemas P2P populares podem gerar grandes quatidades de tráfego de rede.

Alheios a essas desvantagens, os usuários da Internet continuam dedicando considerável parte da largura de banda das redes para o tráfego P2P, conforme estudo apresentado em [Sandvine 2009], que aponta uma média de tráfego da Internet distribuída, principalmente, entre Navegação Web (31,7%), Entretenimento em Tempo-Real (24,6 %) e Compartilhamento P2P (23,4%).

[Sandvine 2010] apresenta um resultado interessante a respeito do tráfego na Internet produzido por usuários móveis, cujo sumário encontra-se na Tabela 3. Este mesmo estudo constata que 30% do tráfego de pico nas Américas Central e do Sul são devidos a fluxo P2P. As colunas "Médias de Usuários mais Ativos" considera o tempo de conexão.

Tabela 3: Visão Geral do Tráfego Móvel de Internet em 2009

Américas Central e do Sul			
Média Global		Média de Usuários mais Ativos	
Navegação Web	26,2%	Navegação Web	39,8%
Entretenimento em Tempo-Real	26,0%	Compartilhamento P2P	27,2%
Compartilhamento P2P	21,0%	Serviços de Armazenamento	12,9%
Redes Sociais	8,9%	Entretenimento em Tempo-Real	11,5%
Serviços de Armazenamento	7,4%	Fluxo Encapsulado	4,7%
Outros	10,5%	Outros	3,7%
Europa			
Média Global		Média de Usuários mais Ativos	
Navegação Web	32,6%	Compartilhamento P2P	38,0%
Entretenimento em Tempo-Real	31,1%	Navegação Web	27,5%
Compartilhamento P2P	13,5%	Entretenimento em Tempo-Real	20,3%
Atualizações de Software	5,2%	Redes Sociais	3,5%
Redes Sociais	3,9%	Fluxo Encapsulado	2,4%
Outros	13,8%	Outros	8,3%
Américas do Norte			
Média Global		Média de Usuários mais Ativos	
Navegação Web	36,3%	Navegação Web	40,0%
Entretenimento em Tempo-Real	26,5%	Compartilhamento P2P	29,6%
Compartilhamento P2P	17,5%	Serviços de Armazenamento	11,1%
Redes Sociais	6,2%	Entretenimento em Tempo-Real	10,1%
Encapsulamento Seguro	3,8%	Fluxo Encapsulado	3,2%
Outros	9,7%	Outros	5,9%

4.2 Princípios do Paradigma P2P

Um sistema P2P é uma coleção distribuída de computadores autônomos (pares) que formam um conjunto de interconexões para compartilhar recursos, de tal forma que os pares desempenham um papel simétrico tanto no roteamento de mensagens quanto no compartilhamento dos recursos. Desta forma, para atingir tais objetivos a contento, esses sistemas devem possuir algumas características específicas [Buford, Heather e Lua 2009] : auto-organização, simetria de papéis, escalabilidade, autonomia dos pares, compartilhamento de recursos e flexibilidade.

Auto-organização. No desenvolvimento de um sistema P2P auto-organizável, o projeto não deve utilizar uma topologia em estrela ou baseada em protocolos de inundação (*broadcast*). A topologia deve ser descentralizada de tal forma que a interconectividade de qualquer par (grau de conectividade) não seja dominante. Em outras palavras, auto-organização significa que todos os pares cooperam na formação e manutenção do sistema, com cada par utilizando informações de estado parciais sobre a rede.

Simetria de Papéis. Diferentemente dos sistemas cliente/servidor, os pares devem assumir papéis simétricos. Qualquer par pode armazenar objetos de interesse de outros pares, responder a requisições (bem como fazê-las) e rotear mensagens. Na prática, esta propriedade ideal é afetada pelo tempo de vida dos pares, pela variação dos *hardwares* dos equipamentos (inclusive sua capacidade de rede) e por questões de redes diversas.

Escalabilidade. Muitas aplicações P2P operam atualmente com milhões de pares, o que exige um eficiente comportamento escalável. Quantitativamente, escalabilidade significa que os recursos de rede e de processamento de cada par possuem uma taxa de crescimento em função do tamanho do sistema "menor" do que linear. Dito de outra forma, a escalabilidade é a habilidade do sistema em lidar com o crescimento no número de nós e de recursos compartilhados.

Autonomia dos Pares. A autonomia dos pares traduz-se de várias formas. Por exemplo, cada par determina suas capacidades baseado em seus próprios recursos. Além diso, cada par determina quando juntar-se ao sistema, que tipo de requisição fazer e quando desconectar-se. Assim, os sistema P2P lidam com a não-previsibilidade na oferta de serviços, e os pares devem estar cientes desse fato. Algumas técnicas para contornar a não-previsibilidade dos sistemas P2P incluem redundância de informação e incentivo de uso da rede.

Compartilhamento de Recursos. O compartilhamento de recursos é o cerne das redes P2P. Tais recursos incluem ciclos de CPU, armazenamento em disco e largura de banda. Deve haver um nível mínimo de contribuição de recursos para um par participar do sistema P2P. Esse recursos são utilizados para proporcionar a operação do sistema e fornecer serviços aos outros pares. Entretanto, a contribuição de cada par deve ser *justa*, o que significa que deve haver limites mínimo e máximo de cooperação para cada integrante do sistema.

Flexibilidade. Sistemas P2P devem ser flexíveis em face do comportamento dinâmico de seus membros. Como os pares possuem uma visão incompleta da topologia da rede e dos integrantes da mesma, o funcionamento do sistema depende da cooperação de pares intermediários para entregar mensagens às mais diversas regiões. Quando os pares entram na rede ou saem do sistema, o processo de roteamento é afetado. A estrutura topológica deve proporcionar múltiplos caminhos entre todos os pares para não ser afetada por esse dinamismo no comportamento da rede.

A essa lista de características (que não pretende ser definitiva), [Harrell et al. 2003] acrescenta as seguintes: Desempenho, Manutenção, Confiabilidade e Usabilidade.

Todas essas são características gerais, aplicáveis à maioria dos sistemas P2P. Entretanto, sistemas específicos possuem características peculiares, tais como os sistemas P2P móveis. "Enquanto atualmente os nós móveis representam uma pequena percentagem dos pares em um sistema P2P, no futuro, à medida em que a capacidade e a largura de banda aumentam e a população de equipamentos cresce, esta situação poderá reverter-se" [Buford e Yu 2010]. Assim, o impacto da mobilidade no desempenho de redes P2P é uma questão importante.

Equipamentos móveis têm quatro propriedades que afetam sua interação com as redes P2P diferentemente dos computadores convencionais (*desktops*): *roaming* [2], heterogeneidade de nós, limitações de energia e variados tipos de interface.

Roaming. Quando um par efetua *roaming*, seu endereço de rede muda. Ao ocorrer a troca, o terminal ainda pode se conectar à Internet, mas, dependendo de como ocorra a transição, as aplicações em curso poderão ser interrompidas. Os sistemas P2P são um tipo dessas aplicações. Adicionalmente, todos os pares vizinhos do terminal em *roaming* perderão a referência (endereço) para se comunicar com o par errante.

[2] *Roaming* é um termo empregado em telefonia móvel, mas também aplicável a outras tecnologias de rede sem-fio, e designa a habilidade de um terminal de rede em obter conectividade em áreas onde seja visitante.

Heterogeneidade. Esta propriedade refere-se à variação na capacidade de rede e nos recursos computacionais disponíveis ao longo do nós da rede. Para lidar com essa situação, o sistema pode adotar a política de distinguir entre nós de maior e de menor capacidade, adaptando o algoritmo de manutenção dos serviços (de acordo com a largura de banda, por exemplo).

Limitações de Energia. O gerenciamento de energia envolve a combinação de uma série de técnicas, incluindo interfaces de rede que podem disparar o modo de economia de energia do equipamento enquanto desabilitam certas atividades de rede e certos protocolos de rede para reduzir o consumo de energia. Na visão atual, se um par entra no modo de economia de energia, ele é tratado como se tivesse deixado o sistema.

Múltiplas Interfaces de Rede. Um equipamento em *roaming* não deve apenas modificar seu endereço de rede, mas pode também precisar trocar a tecnologia de acessso (802.11, WiMax, dentre outras). Essas tecnologia diferem em alcance e largura de banda. A diferença de largura de banda afeta a habilidade do nó repassar mensagens.

Uma última palavra sobre as características dos sistemas P2P cabe ao assunto segurança. Arquiteturas de distribuição de conteúdo P2P apresentam um desafio particular para o provimento de níveis de *disponibilidade*, *privacidade*, *confidencialidade*, *integridade* e *autenticidade* geralmente requeridos devido à sua natureza aberta e autônoma. "Os nós da rede devem ser considerados partes não-confiáveis, e nenhuma suposição pode ser feita em relação ao seu comportamento" [Androutsellis-Theotokis e Spinellis 2004]. Além das questões de segurança convencionais, os sistemas P2P possuem alguns riscos de segurança específicos, incluindo ataques impessoais em larga escala, seu uso como plataforma para DDoS (*Distributed Denial of Service*) e "poluição" de arquivos.

No tocante à privacidade e à confidencialidade, a *anonimidade* é um dos principais focus de muitas infra-estruturas baseadas em P2P e sistemas de distribuição de conteúdo. Em tais sistemas, a anonimidade refere-se:

- ao autor do conteúdo;
- à identidade de um par que armazena o conteúdo;
- aos detalhes do conteúdo em si; e
- aos detalhes da consulta relacionada ao conteúdo.

Algumas abordagens foram desenvolvidas para garantir a anonimidade, como a *Dis-*

sociação da Fonte do Conteúdo e do Consultante [3], as *Camadas de Conexão Anônima* [4] e a *Pesquisa Resistente à Censura* [5].

4.3 Protocolos de Pesquisa e Roteamento

A seção 4.1 apresentou uma classificação para os sistemas P2P baseada em suas arquiteturas (*Centratlizada*, *Descentraliza/Estruturada* e *Não-Estruturada*). A arquitetura do sistema P2P influencia e, em última análise, determina os esquemas de roteamento e de pesquisa por recurso adotados. Os protocolos de funcionamento definidos para essas duas modalidades de atividades, vitais para a continuidade do serviço da rede sobreposta (rede *overlay*), são decisivos para a obtenção de um desempenho satisfatório.

As arquiteturas centralizadas não exigem esquemas elaborados de pesquisa, já que a informação indexada de localização encontra-se estritamente armazenada nos servidores centrais do sistema. Não exigem, tão pouco, técnicas de roteamento complexas, pois a comunicação entre os pares "finais" não é intermediada por outros pares. Nestas arquiteturas, há um ponto único de falhas (o servidor central) que torna o sistema inerentemente não-escalável e vulnerável a censuras, falhas técnicas e ataques maliciosos. O sistema *Napster* [6] é um exemplo de arquitetura centralizada: empregando um simples esquema cliente-servidor, os índices de recursos (arquivos disponíveis nos clientes) são armazenados no servidor central, que responde às consultas por localização fornecendo o endereço do cliente mantenedor do arquivo e, após esse processo, a comunição entre os clientes é feita sem intermediários. Resta, assim, a análise dos protocolos empregados nas estruturas descentralizadas.

4.3.1 Arquiteturas Não-Estruturadas

Em arquiteturas não-estruturadas, os depósitos de recursos estão completamente desvinculados da topologia do sistema. Obviamente, tais recursos precisam ser localizados. Os mecanismos de pesquisa empregados variam desde métodos de força bruta (como inundação da rede com dados das pesquisas) até estratégias mais sofisticadas que preservem os recursos do sistema, incluindo o uso de "caminhos aleatórios" (*random walks*) e

[3] Implementada no sistema *Freenet* [Clarke et al. 2000].

[4] Implementada nos sistemas *Onion Routing* [Goldschlag, Reed e Syverson 1999] e *Mix Networks* [Chaum 1981] [Berthold, Federrath e Kopsell 2001].

[5] Implementada no sistema *Achord* [Hazel e Wiley 2002].

[6] Em [Lv et al. 2002], a esturutura Napster é classificada como *Centralizada*, apesar de ser classificada como *Não-Estruturada* em [Androutsellis-Theotokis e Spinellis 2004].

índices de roteamento. Observa-se, porém, que estes mecanismos de pesquisa possuem implicações relacionadas a disponibilidade e escalabilidade. Populações de nós altamente transitórios (usuários móveis, por exemplo) apresentam desafios para os quais o emprego de sistemas não-estruturados é mais indicado, dada as dificuldades inerentes da falta de estrutura topológica. Alguns representantes de sistemas não-estruturados são Gnutella [Ripeanu 2001] [Ripeanu e Foster 2002], KaZaA [Shin, Jung e Balakrishnan 2006], Edutella [Nejdl et al. 2002] e FreeHaven [Dingledine, Freedman e Molnar 2001].

4.3.2 Arquiteturas Estruturadas

Os diversos sistemas P2P estruturados usam diferentes mecanismos de roteamento para localizar nós e recuperar itens de dados desses nós. Alguns mecanismos incluem *similaridade* para estimar a localização de recursos, *tabelas de roteamento distribuídas* contendo a localização exata dos recursos, *coordenadas cartesianas multidimensionais* para mapear o espaço de distribuição das tabelas de roteamento e *estruturas de dados em malha* para incrementar a tolerância a falhas. Como representantes destes mecanismos, podem ser citados os sistemas *Freenet* [Clarke et al. 2000], *Chord*, *CAN* e *Tapestry* [Zhao, Kubiatowicz e Joseph 2001]. As arquiteturas Chord e CAN são brevemente analisadas nas seções seguintes.

Chord

Um problema fundamental enfrentado pelas aplicações P2P é localizar com eficiência os nós que armazenam os recursos desejados. *Chord* [Stoica et al. 2003] é um protocolo distribuído de pesquisas que se propõe a resolver essa problemática. Dada uma chave, o protocolo consegue mapeá-la para um nó da rede. A localização de dados pode ser facilmente implementada no topo da arquitetura pela associação de uma chave a cada item de dados e pelo armazenamento do par chave/dado no nó para o qual a chave aponta. A arquitetura Chord adapta-se eficientemente à medida que nós juntam-se ao sistema ou deixam-no, podendo responder a consultas mesmo que o sistema altere-se constantemente. A escalabilidade é alcançada pela escala logarítmica (em função da quantidade de nós) com que cada par precisa manter informações sobre o estado da rede e pelo custo de comunicações que obedece a mesma função de crescimento logarítmico.

De acordo com seus idealizadores, Chord simplifica o projeto de sistemas e aplicações P2P procurando resolver as seguintes dificuldades:

Balanceamento de Carga. O protocolo age como uma função de Hash distribuída (*Distributed Hash Table - DHT*), espalhando as chaves uniformemente entre os nós.

Descentralização. Nenhum nó é mais importante do que outro, o que aumenta a robustez da rede e torna o protocolo apropriado a sistemas P2P fracamente organizados.

Escalabilidade. O crescimento logarítmico do custo para pesquisas possibilita a utilização do protocolo inclusive em sistemas populosos.

Disponibilidade. Existe um ajuste automático das tabelas internas do sistema à medida que novos nós entram na rede ou dela saem, procurando evitar com que falhas impeçam a localização de recursos.

Nomeação Flexível. Não existe restrição na estrutura de pesquisa de chaves, permitindo-se com que as aplicações mapeiem de diversas formas seus próprios nomes para as chaves Chord.

Para a correta identificação de nós, o endereço IP de cada par é mapeado por meio de uma função de Hash para uma chave n que serve como seu identificador (*Id*) na rede Chord. Esses nós passam, então, a integrar o anel Chord no qual os pares se organizam estruturalmente.

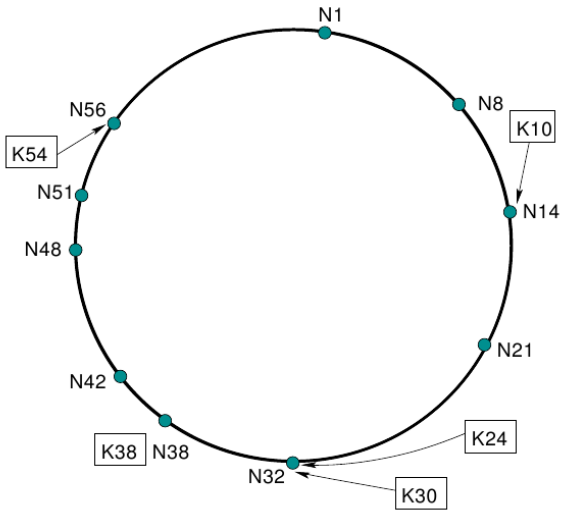


Figura 7: Anel Chord

Cada item de dado, adequadamente identificado por uma chave numérica k, é associado ao nó de mesmo valor n, caso este exista. Não havendo tal nó na rede, o próximo valor no anel é escolhido para a associação do dado. Esse nó é chamado de *sucessor(k)* (*successor(k)*) e armazenará efetivamente o item de dado.

A Figura 7 ilustra um anel Chord de módulo 2^m, onde $m = 6$, com 10 nós integrantes e

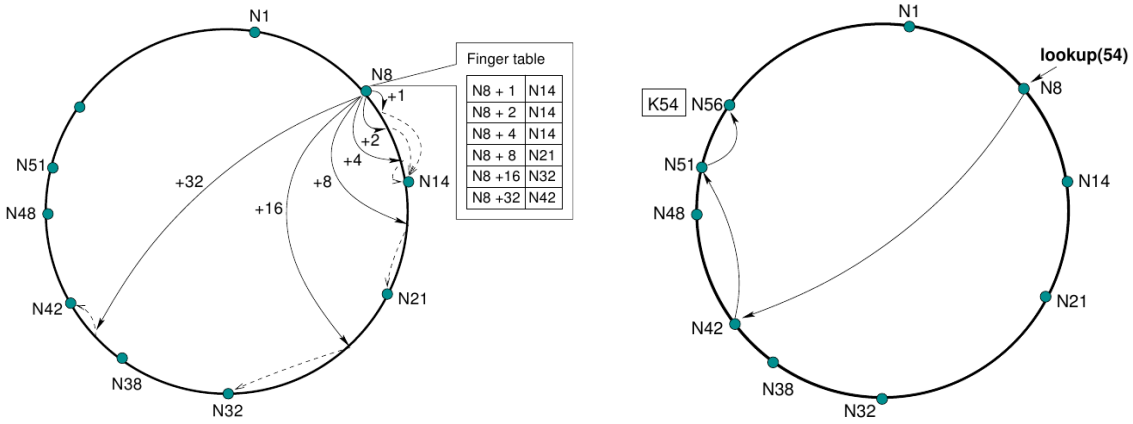
suas respectivas chaves n. O anel armazena 5 itens de dados com chaves k. Observa-se que a chave *k10* está armazenada no nó *n14*, pois esse é o primeiro valor após 10 (inexistente no anel). Similarmente, as chaves *k24* e *k30* estão armazenadas no nó *n32*, a chave *k38* no nó *n38* e a chave *k54* no nó *n56*.

Para uma eficiente localização de itens de dados no anel Chord, cada nó n mantém uma tabela de roteamento (chamada *finger table*) contendo m entradas ($O(\log n)$) que apontam para os nós sucessores de n no anel. Evidentemente, a tabela também inclui o endereço IP dos nós sucessores para que a comunicação efetivamente se estabeleça.

Quando um nó deseja localizar um determinado item de dado com chave k, consulta a sua *finger table* à procura do *sucessor(k)*. Caso o valor k seja encontrado, a busca termina. Se o valor desejado não estiver presente, a consulta é encaminhada para o nó sucessor correspondente, que efetuará o mesmo processo até que o nó destino seja encontrado.

A Figura 8(a) apresenta um exemplo de *finger table* para o nó $n8$. Os cáculos para o preenchimento da tabela são efetuados com aritmética modular, sendo que cada entrada i é calculada por

$$finger[i] = (n + 2^{k-1}) \ mod \ 2^m, 1 \leq i \leq m. \tag{4.1}$$



(a) Entradas na *finger table* do nó $n8$

(b) Pesquisa pela chave $k54$ iniciando no nó $n8$

Figura 8: O mecanismo de busca Chord

A Figura 8(b) ilustra uma pesquisa pela chave $k54$ (armazenada no nó $n56$) efetuada pelo nó $n8$. Este esquema de roteamento exige poucas consultas globais, proporcionando robustez e escalabilidade ao sistema.

Quando um novo nó entra na rede, algumas chaves previamente associadas a alguns

nós precisam ser atualizadas. Similarmente, quando um nó deixa a rede, todas as chaves associadas a este nó precisam ser reassocidas ao nó sucessor. Entradas e saídas, normalmente, têm um custo de $O(\log n)$ com alta probabilidade, podendo chegar a $O(n)$ no pior caso.

Content-Addressable Network – CAN

Assim como Chord, a arquitetura CAN [Ratnasamy et al. 2001] monta uma tabela baseada numa função de Hash para mapear valores em chaves a serem distribuídas pela rede, possuindo três operações básicas: inserção, consulta e exclusão de pares (chave k, valor v). Cada nó armazena uma *zona* da tabela de Hash e algumas informações sobre as zonas adjacentes à sua. Cada requisição (inserção, pesquisa ou exclusão) por uma chave em particular é roteada pelos nós intermediários até o nó cuja zona contém a chave. Inteiramente implementável no nível de aplicação, o projeto CAN é completamente distribuído (não requer qualquer tipo de controle, coordenação ou configuração centralizados), é escalável (os nós mantêm apenas uma pequena quantidade de estados de controle, independente do número de nós no sistema) e tolerante a falhas (nós podem rotear "contornando" os pontos de falha).

CAN usa um espaço virtual d-dimensional de coordenadas Cartesianas para armazenar os pares (k, v). A zona da tabela de Hash pela qual um nó é responsável corresponde a um segmento deste espaço de coordenadas. Consequentemente, qualquer chave k é deterministicamente mapeada em um ponto p do espaço de coordenadas. Assim, o par (k, v) é armazenado no nó responsável pela zona na qual encontra-se o ponto p.

A Figura 9 ilustra um exemplo de mapeamento em um espaço bidimensional. As regiões de A a E representam as zonas de valores da tabela de Hash mapeados no sistema. Quaisquer valores de chaves 2-D cujas coordenadas (x, y) estejam entre os limites ($0-0,5$; $0-0,5$) serão armazenados pelo nó responsável pela zona "A". Semelhantemente, valores de chaves entre os limites ($0,5-0,75$; $0,5-1$) serão armazenados pelo nó responsável pela região "D". Para recuperar uma entrada correspondente a determinada chave k, qualquer nó pode aplicar a mesma função de Hash determinística para efetuar o mapeamento para o ponto p e obter o valor desejado a partir do nó responsável pela zona em que este ponto se encontra. Desta forma, as consultas, em geral, precisam ser encaminhadas entre os diversos nós do sistema CAN, o que exige um eficiente mecanismo de roteamento.

Os nós da arquitetura CAN são auto-organizáveis em uma rede *overlay* que representa o espaço virtual de coordenadas para o qual as chaves são mapeadas. Um nó aprende e

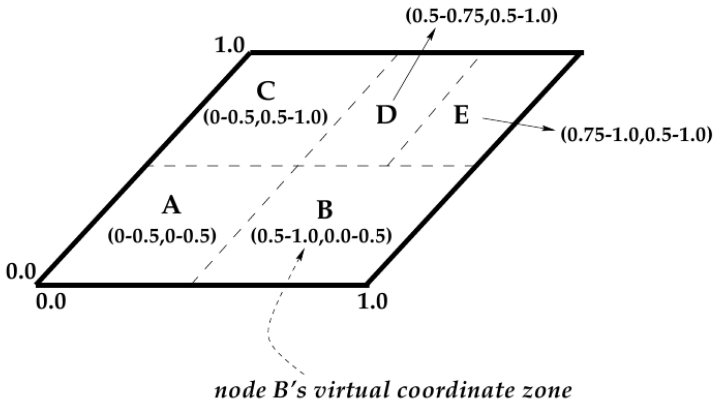


Figura 9: Mapeamento CAN

mantém os endereços IP dos nós adjacentes à sua zona. Este conjunto de nós vizinhos no espaço de coordenadas atua como uma tabela de roteamento que permite o repasse de mensagens para qualquer ponto do espaço d-dimensional utilizado.

Os três aspectos mais importantes desta arquitetura são: roteamento, construção do espaço virtual de coordenadas e manunteção deste espaço virtual.

Roteamento. Um nó CAN matém uma tabela de roteamento que contém o endereço IP e a zona de cada um dos seus vizinhos imediatos no espaço virtual. Este estado local dos vizinhos é suficiente para viabilizar o roteamento entre dois pontos arbitrários no espaço d-dimensional. Uma mensagem CAN inclui o endereço IP das coordenadas de destino. Assim, utilizando os conjuntos de coordenadas de sua vizinhança, um nó encaminha a mensagem para o seu destino através de uma técnica gulosa escolhendo o vizinho cuja coordenada mais aproxime-se da coordenada de destino. Para um espaço d-dimensional particionado em n zonas de mesmo tamanho, o comprimento médio do caminho de roteamento é de $(d/4)(n^{1/d})$ saltos e nós individuais mantêm informações de estado sobre $2d$ nós vizinhos. Isto significa que pode-se aumentar a quantidade de nós (e, portanto, de zonas) sem aumentar a quantidade de informações de estado por nó, ao passo que o caminho médio de roteamento cresce em $O(n^{1/d})$.

Construção. O espaço CAN é dividido entre os nós correntemente no sistema. Quando um novo nó entra no sistema, uma porção do espaço de coordenadas é a ele alocada. Para tanto, é necessário dividir pela metade a zona sob responsabilidade de algum nó interno, ficando uma metade com este nó e a outra com o nó entrante. O processo segue três etapas:

- Primeiramente, o novo nó deve contactar um nó interno;

- Em seguida, usando o mecanismo de roteamento, deve ser encontrado o nó cuja zona será dividida;
- Finalmente, os nós vizinhos à zona dividida devem ser notificados para a devida atualização em suas tabelas de roteamento.

A Figura 10 ilustra a inserção do nó 7 no sistema CAN bidimensional contendo alguns nós. O nó entrante comunica-se com o nó responsável pela zona que contém as coordenadas (x, y) manifestando a intenção de compor o sistema. A mensagem é roteada até o nó de destino (nó 1, no exemplo) cuja zona é dividida ao meio e distribuída entre os nós 1 e 7. Por fim, os vizinhos são orientados a atualizarem adequadamente suas tabelas de roteamento. Observa-se que a adição de um novo nó afeta apenas um pequeno número de nós existentes em uma restrita localidade do espaço de coordenadas. O número de nós vizinhos cujos estados um nó mantém depende somente da dimensionalidade do espaço de coordenadas, sendo independente do número total de nós no sistema. Assim, a inserção de um nó no sistema CAN d-dimensional afeta tão somente $O(d)$ nós existentes, o que é uma característica importante para sistemas com uma grande quantidade de nós.

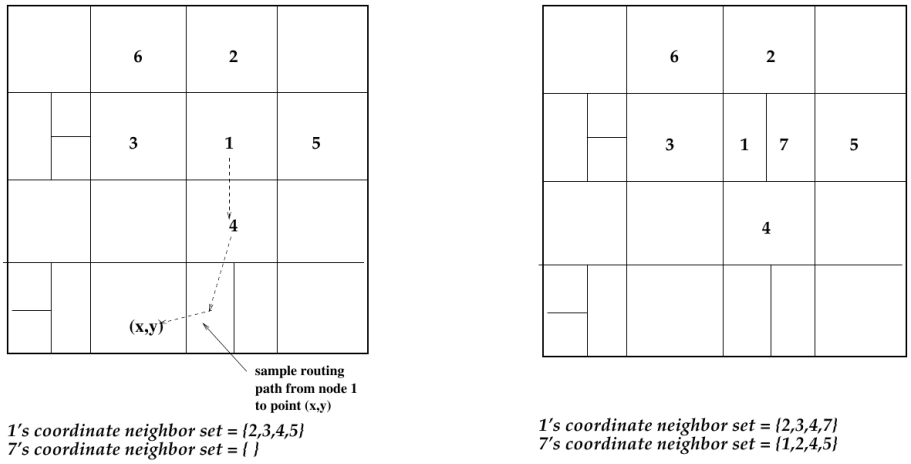


Figura 10: Inserção do nó 7 no sistema CAN

Manutenção. Quando um nó deixa o sistema, sua respectiva zona deve ser abarcada adequadamente por algum de seus vizinhos, o que é feito por um processo simples que busca (re)criar uma nova zona válida. Uma questão mais delicada refere-se à reação do sistema na presença de falhas, quando um ou mais nós tornam-se simplesmente inalcançáveis. Periodicamente, os nós vizinhos trocam informações entre si, informando os limites de suas zonas e os dados de seus vizinhos. Quando essas mensagens de atualização deixam de chegar, uma falha é identificada e os nós vizinhos iniciam um processo de "tomada de

posse" da zona remanescente. Baseados em alguma característica previamente estabelecida (como o volume de zona ou a qualidade conexão, por exemplo), os nós disparam um temporizador interno e, após as primeiras expirações, definem o novo responsável pela área desocupada e atualizam suas tabelas de roteamento.

Uma série de melhorias de desempenho para o projeto CAN foram propostas, tais como a sobrecarga de zonas de coordendas, que consiste na atribuição de responsabilidade por uma zona a mais de um nó, sem efetuar a divisão comumente empregada. Este procedimento, aumenta a robustez do sistema, diminuindo a probabilidade de falhas de comunicação com zonas adjacentes, aumentando o número de possibilidades de roteamento e incrementando a disponibilidade de informações. Outras melhorias incluem: espaços de coordenadas múltiplos (realidades), melhores métricas de roteamento, funções de Hash múltiplas, particionamento mais uniforme e *caching* e replicação de dados.

Uma última questão relevante a ser observada na arquitetura CAN refere-se ao seu custo operacional de roteamento, $O(dn^{1/d})$. Muitas abordagens existentes propõem um roteamento em $O(\log\ n)$ saltos com cada nó mantendo o estado de $O(\log\ n)$ vizinhos. Nota-se, contudo, que ao selecionar um número de dimensões $d \geq (\log\ n)/2$, atingem-se as mesmas propriedades. A escolha por manter um d fixo independente de n reside na intenção dos autores em permitir a implementação de suas ideias a sistemas bastante numerosos e com mudanças frequentes de topologia, pois em tais sistemas é importante manter o número de vizinhos independente do tamanho do sistema.

4.3.3 Comparação entre as Abordagens

[Lua et al. 2005] apresenta um quadro comparativo que resume as principais características das arquiteturas P2P aqui brevemente apresentadas. Parte deste quadro é reproduzida de forma adaptada na Tabela 4.

Uma comparação entre as arquiteturas que utilizam DHT como base de sua técnica de armazenamento de informações e de roteamento de mensagens é apresentada em [Xu 2002] num gráfico aqui reproduzido adaptadamente na Figura 11. O gráfico apresenta uma curva simbólica (coordenadas assintóticas em vez de valores reais) comparando o *tamanho da tabela de roteamento* armazenada nos nós do sistema com o *diâmetro da rede*, em função da quantidade n de pares. Deve-se entender por diâmetro da rede o número de saltos necessários para uma requisição alcançar seu destino no pior caso.

As pesquisas existentes, bem como os produtos desenvolvidos, demonstram que as

Tabela 4: Comparação entre os vários esquemas de Redes P2P

Algoritmo	**Napster**	**Gnutella**	**Freenet**	**Chord**	**CAN**	**Tapestry**
Descentralização	Centralizado	Topologia plana	Funcionalidade DHT fraca	Funcionalidade DHT em escala		
Arquitetura	Centralização de Ids no servidor	Rede Ad-Hoc de servents	Palavras-chave e texto descritivo como Id	Espaço circular e unidirecional para Id de nós	Espaço de coordenadas Id multidimensional	Malha global estilo Plaxton
Protocolo de Busca	Comunicação direta entre pares após obtenção de Id	Inundação de requisições (*flooding*)	Chaves textuais pesquisadas par-a-par	Coincidência entre chave e Id de nó	par *chave*,*valor* mapeado em um ponto p do espaço virtual	Coincidência entre sufixo e Id de nó
Parâmetros do Sistema	Nenhum	Nenhum	Nenhum	n pares na rede [a]	n pares distribuídos em um espaço de d-dimensões	n pares [a], base B de identificadores
Desempenho de Roteamento	Conteúdo disponível enquanto par na rede	Sem garantia de localização; bom desempenho com conteúdo popular	Garantia de localização até o alcance do limite HTL	$O(\log n)$	$O(d.n^{1/d})$	$O(\log_B n)$
Informações de Estado	–	Constante	Constante	$\log n$	$2d$	$\log_B n$
Entrada/Saída de Pares	–	Constante	Constante	$(\log n)^2$	$2d$	$\log_B n$
Confiabilidade/ Tolerância a Falhas	–	Pares recebem múltiplas cópias como resposta, o que degrada o desempenho	Sem hierarquia ou ponto central sujeito a falhas	Dados replicados permitem recuperação na presença de falhas	Múltiplos pares responsáveis pelo mesmo dado evitam falhas na rede	Dados replicados e múltiplos caminhos evitam falhas na rede

[a] [Harrell et al. 2003] indica que, para Chord e Tapestry, n é o tamanho total do espaço virtual (e não o número de pares no sistema)

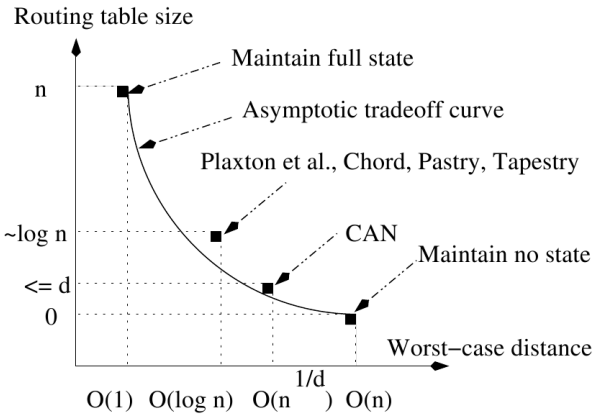


Figura 11: Comparação entre algumas arquiteturas DHT

redes P2P estruturadas são uma importante tecnologia de valor prático. Essa tecnologia ajuda a superar os problemas de escalabilidade e de desempenho enfrentados pelas tecnologias P2P não-estruturadas.

[Dhara et al. 2010] apresenta algumas considerações sobre as redes P2P, das quais destacam-se as seguintes:

- Em geral, as arquiteturas *overlay* P2P podem incrementar desempenho e escalabilidade. Entretanto, para muitas aplicações de tempo-real, existem requisitos de desempenho que precisam ser atingidos, tais como nas aplicações de Voz sobre IP (VoIP) e IPTV. Contudo, como garantir tais requisitos mantendo a escalabilidade e a distribuição de carga do sistema como um todo é uma importante questão a ser tratada pelos algoritmos de redes P2P estruturadas;

- Confiança e reputação são importantes para proporcionar comunicações seguras e confiáveis entre pares de uma rede P2P. Existem muitos tópicos a serem pesquisados nessa área para as redes P2P, tais como anonimidade, ataques de negação de serviço, nós com comportamento malicioso, reputação e incentivo;

- Em alguns cenários, os pares pertencentes a diferentes redes P2P podem precisar conversar entre si. Isto requer protocolos/interfaces bem definidas e um estudo delicado sobre interoperabilidade entre nós P2P;

- No mundo real, diversos aspectos podem afetar o desempenho das redes P2P, tais como disponibilidade de rede e sua largura de banda, latência nas comunicações, poder computacional e de armazenamento dos pares, etc. Assim, o suporte à heterogeneidade é uma questão importante do ponto de vista prático;

- No contexto de aplicações P2P móveis e de redes sem-fio *ad-hoc*, as abordagens devem permitir que os pares tenham controle de fluxo otimizado, mecanismos de balanceamento de carga e "consciência" de proximidade.

Sistemas P2P *overlay* baseados em DHT são suceptíveis a brechas de segurança relativas a ataques de pares maliciosos (falhas bizantinas). Um ataque simples em tais sistemas é o fornecimento de objetos de dados inválidos como reposta a pesquisas por recursos. A autenticidade dos objetos de dados pode ser obtida utilizando-se técnicas de criptografia, seja por meio de chaves públicas seja pelo uso de conteúdo *hasheado*, com consequente aumento no custo de processamento/armazenamento. Tais técnicas, entretanto, não evitam a presença na rede de objetos de dados indesejáveis e nem proporcionam proteção contra ataques de negação de serviço (*DoS – Denial of Service*). Pares maliciosos podem ainda ser capazes de corromper objetos de dados, negar acesso a esses dados ou responder com réplicas deturpadas a pesquisas por recursos, além de serem capazes de se fazerem passar por outros pares.

[Buford, Heather e Lua 2009] sumariza considerações interessantes a respeito das tendências e do futuro das redes P2P, destacando o fato da simplicidade de disponibilização de aplicações distribuídas proporcionada pelas aplicações P2P devido à sua limitada dependência de novas infraestruturas. Mas, é provável que, à medida que as aplicações P2P tornem-se um fenômeno, as expectativas dos usuários por serviços de qualidade cresça. Os métodos de gerência de aplicações e de redes convencionais podem ser considerados para gerenciar os futuros sistemas P2P *overlays*. Contudo, algumas diferenças nítidas requerem adaptação, incluindo a característica auto-organização dos pares. Nas próprias palavras de [Buford, Heather e Lua 2009]:

> "As atuais aplicações P2P são do tipo *melhor esforço*, mas, no futuro, elas poderão oferecer diferentes classes de serviço com múltiplos níveis de tarifação. O suporte a tais modelos de serviço requer monitoramento em tempo-real e a habilidade de proporcionar novos recursos para incrementar o desempenho".

5 Curvas de Preechimento

"*Quando então um grande cabalista revelar-lhe algo, o que ele disser não será frívolo, vulgar, comum, mas, ao contrário, um mistério, um oráculo...*"
Tommaso Garzoni, *Il Theatro de'vari, e diversi cervelli mondani*, Veneza, 1583, discorso XXXVI

As *Curvas de Preenchimento Espacial* [1] são também chamadas de *Curvas de Peano* em homenagem ao matemático Giuseppe Peano que descobriu, em 1890, a primeira curva pertencente a esta categoria de curvas planas [Peano 1890]. Apesar deste fato, a popularização da existência de tais curvas deu-se pelo matemático David Hilbert, que lançou luz sobre esse fascinante assunto em 1891 com seu trabalho sobre "mapeamento contínuo de uma linha em uma figura plana" (*Üeber die stetige Abbildung einer Linie auf ein Flächenstück*) [Hilbert 1891]. Nele, é apresentado o primeiro procedimento generalizado de geração geométrica para as curvas em questão, o que possibilitou a construção de toda uma classe de curvas de preenchimento espacial.

As curvas de Peano são utilizadas em uma série de aplicações, tais como a redução de um problema multidimensional em um problema unidimensional [Lawder e King 2001], a manipulação e análise de imagens digitais [Agranov e Gotsman 1995], a definição de regras para distribuição de componentes VLSI [O'Sullivan 1995], a definição de estruturas de dados [Gotsman e Lindenbaum 1996], criptologia [Matias e Shamir 1988] e compressão de dados [Salomon 2004], para citar algumas. [Ghinita, Kalnis e Skiadopoulos 2007] utilizou, recentemente, uma curva de preenchimento espacial para ordenar coordenadas bidimensionais de pontos no espaço Cartesiano, estabelecendo uma regra de proximidade entre pontos baseada nessa ordenação. Esta aplicação, em particular, é que motiva o estudo das curvas de preenchimento espacial e sua inclusão no presente trabalho.

Este capítulo apresenta conceitos básicos, sem grande aprofundamento matemático, relacionados às curvas de preenchimento espacial, enfatizando a clássica *Curva de Hilbert*, talvez a mais popular dentre as curvas desta categoria [Alber e Niedermeier 1998].

[1] A noção de *Curvas de Preenchimento Espacial* tem sua origem no desenvolvimento do conceito de *conjunto de Cantor* [Cantor 1883].

5.1 Noções Preliminares

Nas palavras de [Sagan 1994], "O assunto *curvas de preenchimento espacial* fascina os matemáticos há cerca de um século e tem intrigado muitas gerações de estudantes."

Mas, para a correta compreensão dos conceitos envolvidos, necessário se faz uma explanação preliminar sobre curvas planas. Segundo [Alencar e Santos 2003], a noção intuitiva de curva plana é a de uma figura que pode ser "desenhada" com um único traço, "sem tirar o lápis do papel". Tornando mais precisa a definição, uma curva é uma deformação contínua de um intervalo, ou ainda, a trajetória do deslocamento de uma partícula no plano.

Apoiando-se na Geometria Analítica, poder-se-ia considerar uma curva em R^2 como o conjunto de pontos $(x, y) \in R^2$ que satisfazem a uma equação do tipo

$$F(x, y) = 0. \tag{5.1}$$

Alguns exemplos são ilustrados na Figura 12.

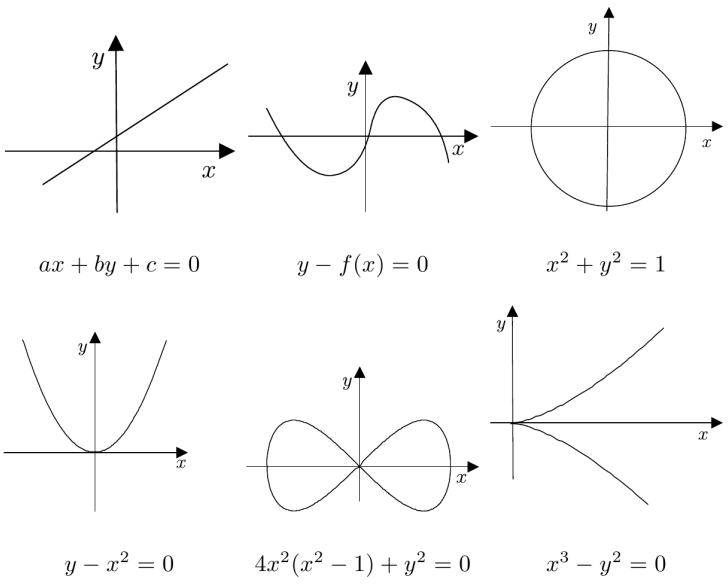


Figura 12: Exemplos de curvas planas

Contudo, tal definição é restritiva, englobando apenas funções muito bem comportadas. Em vez disso, considera-se-á a seguinte definição:

Definição 5.1 *Uma curva contínua no plano R^2 é uma aplicação contínua $\alpha : I \to R^2$, definida num intervalo $I \subset R$. A aplicação α, dada por $\alpha(t) = (x(t), y(t))$, é contínua, se*

cada função coordenada $x, y : I \to R$ é uma função contínua.

O conjunto imagem C da aplicação α, dado por

$$C = \alpha(t) = (x(t), y(t)), t \in I$$

é chamado de *traço* de α.

Esta forma parametrizada de expressar o traço C, através de α e de seu parâmetro t, proporciona maior flexibilidade na seleção das curvas planas em questão. O intervalo $I = [a, b]$ determina os pontos inicial e final de α, quais sejam $\alpha(a)$ e $\alpha(b)$. Para o caso particular em que $\alpha(a) = \alpha(b)$, diz-se que a curva é *fechada*.

Como exemplo, considere-se a função contínua p(t) dada pelas seguintes expressões

$$p(t) = \left\{ \begin{array}{ccc} 0 & para & 0 \leq t \leq \frac{1}{3} \\ 3t-1 & para & \frac{1}{3} \leq t \leq \frac{2}{3} \\ 1 & para & \frac{2}{3} \leq t \leq 1 \end{array} \right\}, \; p(-t) = p(t), \; p(t+2) = p(t) \tag{5.2}$$

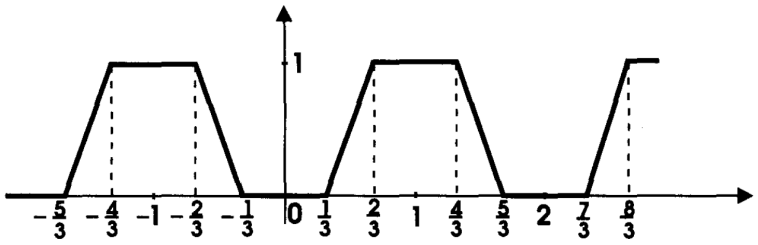


Figura 13: Gráfico da Função Geradora $p(t)$

O gráfico apresentado na Figura 13 mostra que $p(t)$ é contínua. Isaac J. Schoenberg [Schoenberg 1938] (apud [Sagan 1994]) propôs uma curva de preenchimento espacial periódica definida pelas componentes φ e ψ, parametrizada em t como a seguir

$$\varphi(t) = \frac{1}{2} \sum_{k=0}^{\infty} p(3^{2k}t)/2^k, \; \psi(t) = \frac{1}{2} \sum_{k=0}^{\infty} p(3^{2k+1}t)/2^k \tag{5.3}$$

[Sagan 1994] apresenta uma prova, baseada em séries geométricas, da continuidade e da sobrejetividade de 5.3. Neste mesmo trabalho são apresentados gráficos gerados pelas primeira, segunda, terceira e quarta aproximação de polígonos, cuja reprodução está ilustrada na Figura 14

No limite de $k \to \infty$, a curva descrita pelas duas funções apresentadas em 5.2 e 5.3 preenche todo o espaço considerado.

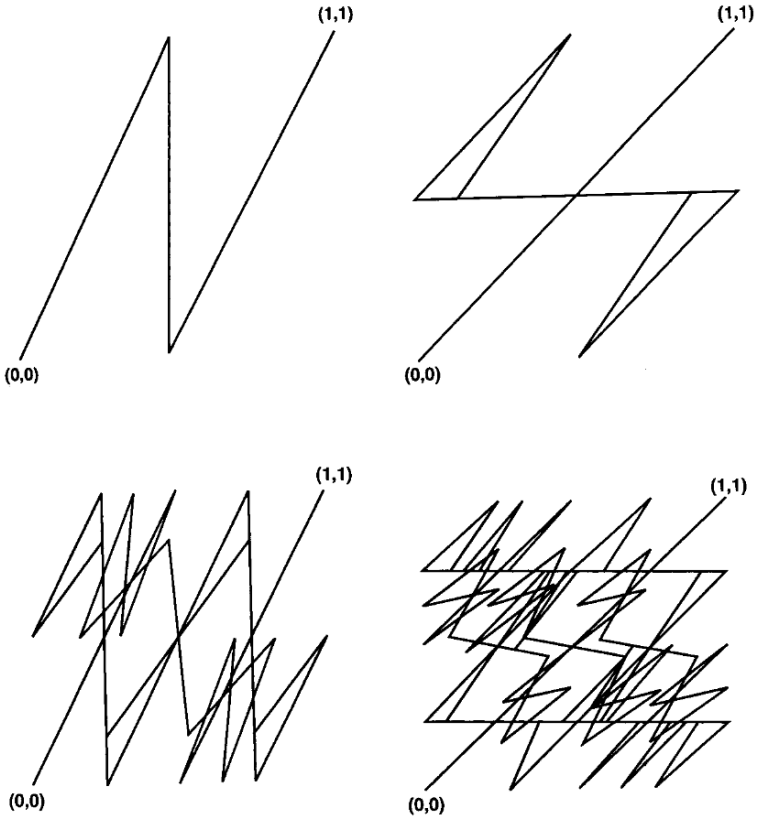


Figura 14: Polígonos de Aproximação de $\varphi(t)$ e $\psi(t)$, para $1 \leq k \leq 4$

A descrição das curvas de preenchimento espacial por meio de funções paramétricas, por vezes, não é trivial, tornando-se, um muitos casos, uma tarefa extremamente complexa. Contudo, outros métodos existem para expressar tais curvas. Desde que as curvas de preenchimento espacial de Hilbert passaram a ser extensivamente utilizadas em várias aplicações, diferentes métodos de geração têm sido sugeridos para reduzir a complexidade de tempo computacional e espaço de armazenamento.

Em [Butz 1967] (apud [Chen et al. 2010]) é utilizado um algoritmo iterativo para computar o mapeamento da função com uma técnica orientada a byte, com OU-exclusivo, deslocamento de bits, *etc.* [Sagan 1994] apresenta um método aritmético para a geração, produzindo uma aproximação por polígonos da curva de Hilbert. Uma geração baseada em fractais é apresentada em [Peitgen, Jürgens e Saupe 2004]. Uma obtenção recursiva para a curva de Hilbert encontra-se em [Breinholt e Schierz 1998]. [Jin e Mellor-Crummey 2005] indica uma eficiente abordagem para enumeração de pontos em uma curva de preenchimento utilizando tabelas de especificação de posições.

5.2 A Curva de Hilbert

David Hilbert (1862–1943) foi um proeminente matemático alemão do final do século XIX e início do século XX, cujas grandes contribuições dizem respeito a diversas áreas da matemática: formas algébricas, teoria algébrica dos números, fundamentos de geometria, análise, equações integrais, dentre outras [Sagan 1994].

Em seu brilhante trabalho [Hilbert 1891], é apresentado o primeiro procedimento generalizado de geração geométrica para as curvas de preenchimento espacial. O princípio apresentado é, basicamente, o que segue:

> Se o intervalo I pode ser mapeado continuamente em um quadrado Q, então, após particionar I em quatro sub-intervalos congruentes e Q em quatro subquadrados também congruentes, cada sub-intervalo pode ser mapeado continuamente para cada um dos sub-quadrados. A seguir, cada sub-intervalo (e cada subquadrado) pode, por sua vez, ser particionado em outros quatro sub-intervalos congruentes, e o argumento se repete. Se o processo for levado ao infinito, I e Q serão particionados em 2^{2k} réplicas congruentes para $k = 1, 2, 3 ...$

Hilbert demonstrou que os subquadrados podem ser arranjados de tal forma que os sub-intervalos adjacentes correspondam a subquadrados com um lado em comum, o que preserva as relações de inclusão, isto é, se um quadrado correponde a um intervalo, então seus subquadrados correpondem aos sub-intervalos deste intervalo. A Figura 15 ilustra os três primeiros passos do processo descrito. A linha poligonal destacada indica a ordem em que os subquadrados são arranjados para satisfazer os requisitos apresentados.

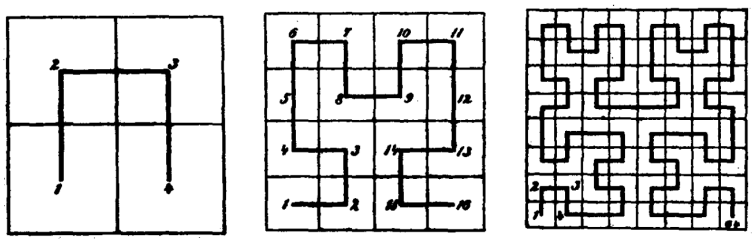


Figura 15: Geração de curvas de Hilbert

Eliakim Hastings Moore, matemático americano, propôs uma versão para as curvas de Hilbert na qual seus pontos iniciais e finais são adjacentes no espaço mapeado, conforme ilustra a Figura 16.

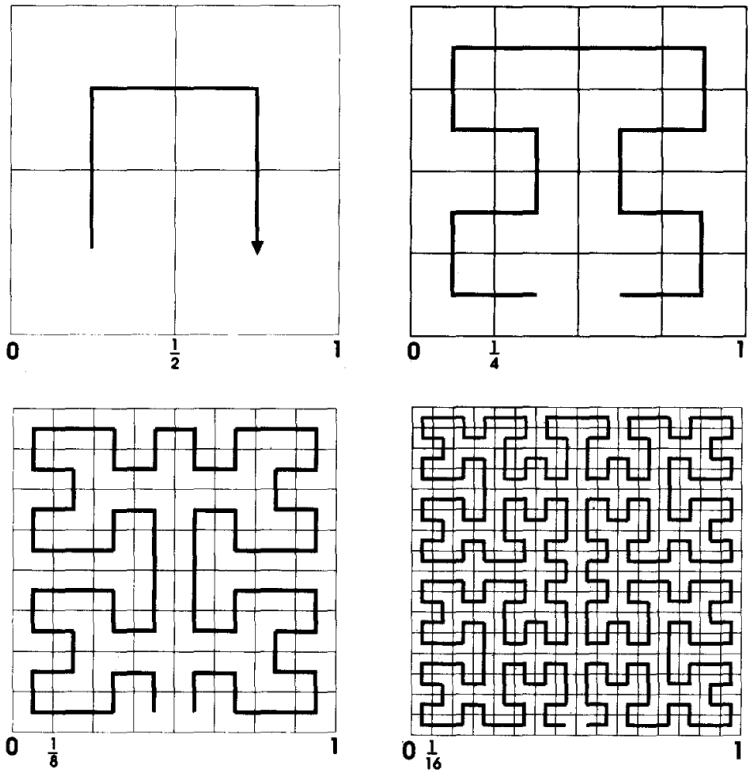


Figura 16: Curva de Hilbert: pontos inicial e final adjacentes

Essas ideias podem ser aplicadas a espaço n-dimensionais. A Figura 17 ilustra o segundo passo para a geração de um curva de Hilbert tridimensional.

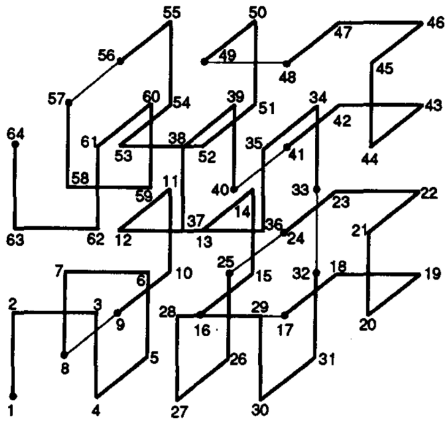


Figura 17: Curva de Hilbert tridimensional

5.2.1 Características da Curva de Hilbert

As aplicações que utilizam a curva de Hilbert mutliplicam-se constantemente, o que se dá devido a algumas características peculiares dessa classe de curvas planas. Uma característica de especial interesse para o presente trabalho diz respeito à *localidade* dos

dados n-dimensionais mapeados sobre a curva. O termo localidade refere-se à comparação da distância entre dois pontos sobre a curva com a distância deste pontos no espaço original. Para o sistema *AnoniMobi*, é desejável que pontos próximos no espaço n-dimensional estejam próximos no resultado do mapeamento.

[Voorhies 1994] apresenta uma medida empírica de localidade para as curvas de preenchimento de Hilbert e de Peano no contexto de renderização de imagens. A grandeza *coerência* [2] é avaliada segundo uma metodologia pragmática estreitamente vinculada à localidade, cujo resultado está ilustrado na Figura 18. A grandeza *Radius* expressa na abcissa tem relação direta com a metodologia implementada.

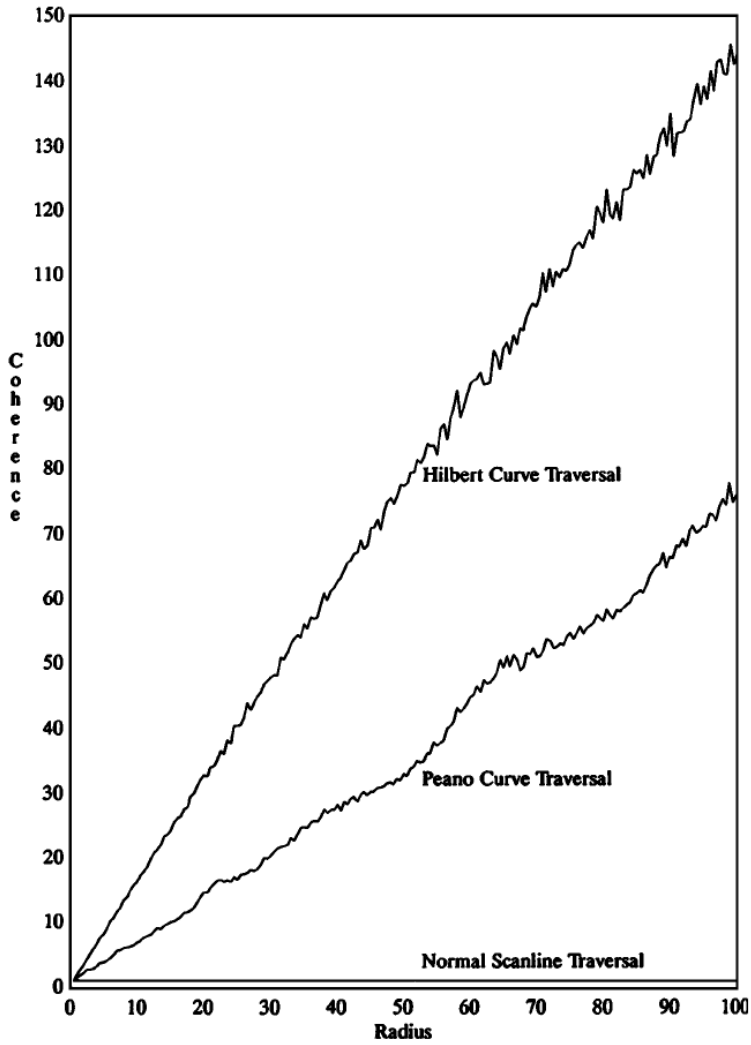


Figura 18: Coerência média das curvas de Peano e de Hilbert

Observa-se que a curva de Hilbert oferece uma alta coerência comparada à curva de Peano. Este fato pode ser traduzido como uma medida de localidade. Num trabalho sobre

[2] O conceito de coerência está ligado às propriedades ópticas das ondas eletromagnéticas.

as propriedades métricas das curvas de preenchimento, [Gotsman e Lindenbaum 1996] avalia analiticamente a localidade de curvas de preenchimento e conclui que "a curva de Hilbert é um excelente exemplo de curva de preenchimento espacial que preserva a localidade" dos dados originais. Um resultado semelhante é obtido de forma empírica por [Faloutsos e Roseman 1989] e [Moon et al. 2001].

O emprego de curvas de Hilbert bidimensionais para definir a proximidade geográfica entre os usuários do sistema Privé em [Ghinita, Kalnis e Skiadopoulos 2007] aproveita-se da boa preservação de localidade destas curvas, garantindo que pontos próximos no mapa sejam mapeados para pontos próximos na curva.

Um outro aspecto importante a ser considerado sobre as curvas de Hilbert diz respeito a sua análise de ordem de complexidade, tanto para o tempo de geração da curva quanto para o espaço de armazenamento requerido. Gerar a curva de Hilbert envolve a reprodução de um padrão básico, rotacionado-o adequadamente em 90 ou 180 a cada nova iteração do procedimento. A Figura 19 ilustra o procedimento.

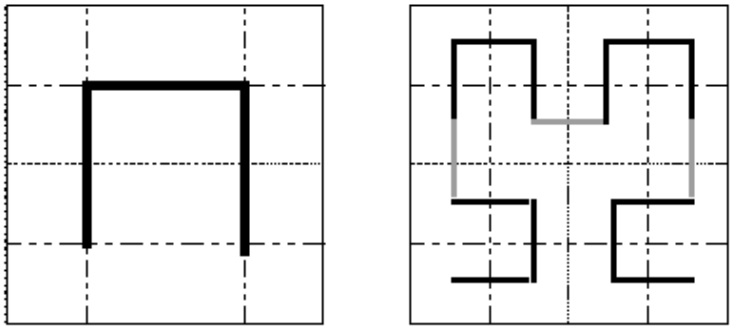


Figura 19: Construção de uma Curva de Hilbert

Tomando-se um espaço bidimensional de 2^{2k} de área (isto é, 2^k x 2^k pontos no plano), o procedimento de geração da curva deverá se repetir até atingir o nível k de granularidade (k-ésimo passo de subdivisão do espaço n-dimensional). Denotar-se-á por H_k^n a curva de Hilbert que preenche todo o espaço n-dimensional na passo k. Para o caso bidimensional, tem-se H_k^2.

Considerando-se como operação básica a construção do nível $k = 1$ (ou seja, o preenchimento do espaço 2 x 2), o preenchimento de todo o espaço considerado irá requerer o percorrimento de 2^k x 2^k pontos e diversas operações de movimentação. [Chen et al. 2010] indica uma ordem de complexidade $O(k4^k)$ para esse procedimento e um espaço de armazenamento da ordem de $O(4^k)$. Definindo-se a quantidade de pontos $m = 2^k$, essas ordens de grandeza assumem os valores $O(m \log m)$ e $O(\log m)$, respectivamente.

6 Algoritmos de Agrupamento

"*Each memorable verse of a true poet has two or three times the written content.*"
Alfred de Musset

A Análise de Agrupamento (*Cluster Analysis*) é um processo não-supervisionado que divide um conjunto de objetos em grupos homogêneos. Existem diversos *algoritmos de agrupamento* (ou *algoritmos de aglomeração*, ou *algoritmos de clusterização*) disseminados nas mais diversificadas áreas.

Devido ao grande número de genes e à complexidade das redes biológicas, a análise de agrupamento é uma técnica exploratória útil para analisar dados de expressão genética, conforme salienta [Yeung e Ruzzo 2000]. No desenvolvimento de sistemas de atenção e cuidado à saúde, a análise de agrupamento é utilizada para identificar grupos de pessoas que podem ser beneficiadas por serviços específicos [Hodges e Womtring 2000]. [Comaniciu e Meer 1999] utiliza a análise de agrupamento para detectar as bordas dos objetos no processo de decomposição de imagens coloridas ou em tons de cinza (segmentação de imagens). A tarefa de agrupamento de dados geralmente é uma etapa preliminar à atividade de mineração de dados [Berry e Linoff 2000], um processo de exploração e análise de grandes quantidades de dados com o intuito de descobrir informações úteis.

Este capítulo introduz alguns conceitos básicos relacionados ao universo dos problemas de agrupamento, focando atenção no problema das k-médias (*k-means problem*) e em seu algoritmo padrão de resolução, objeto de estudo e implementação no presente trabalho.

6.1 Definição e Conceito

Análise de Agrupamento é um nome genérico para uma variedade de métodos matemáticos que podem ser utilizados para definir quais objetos são similares em um conjunto fornecido [Romesburg 2004]. É uma ferramenta para a exploração da estrutura de dados que não requer as comuns assunções da maioria dos métodos estatísticos [Jain e Dubes 1988] e é rotulada como "aprendizado não-supervisionado" na literatura de reconhecimento de padrões e de inteligência artificial. [Kaufman e Rousseeuw 2005] indicam outros termos pelos quais a Análise de Agrupamento é conhecida, tais como *Taxonomia Numérica*, *Classificação Automática*, *botriologia* (*botryology*) e *análise tipológica*.

Com o passar do tempo, uma série de algoritmos têm sido desenvolvidos para realizar a Análise de Agrupamento. Uma das razões para tal reside no fato de que não existe uma definição geral para um agrupamento (*cluster*), existindo diversos deles. Além disso, aplicações diferentes fazem uso de tipos de dados diferentes, tais como variáveis contínuas, variáveis discretas, similaridades e dissimilaridades. Consequentemente, é necessário adaptar métodos de agrupamentos diferentes a certos tipos de aplicação e aos tipos de dados envolvidos.

A Figura 20 ilustra uma análise de agrupamento onde cada uma das três cores indica um grupo (ou classe) diferente para os quadrados fornecidos.

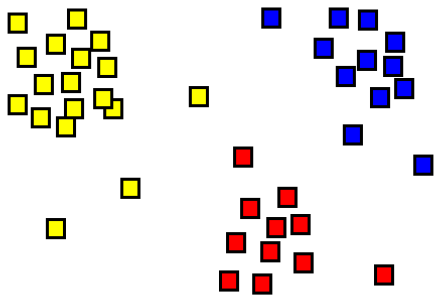


Figura 20: Resultado de uma análise de agrupamento em três *classes*

6.2 Taxonomia

[Gan 2007] apresenta uma divisão básica para os algoritmos de agrupamento, classificando-os em *Agrupamento Hard* e *Agrupamento Fuzzy*.

No Agrupamento Hard, os algoritmos associam um rótulo de classe $l_i \in \{1, 2, ..., k\}$ a cada objeto x_i para identificar seu grupo (classe), onde k é o número de grupos. Em outras palavras, assume-se que cada objeto pertence a apenas um grupo. Matematicamente, o resultado de um algoritmo de agrupamento desta categoria pode ser representado por uma matriz $k \times n$

$$U = \begin{pmatrix} u_{11} & u_{2l} & \cdots & u_{1n} \\ u_{22} & u_{22} & \cdots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ u_{k1} & u_{k2} & \cdots & u_{kn} \end{pmatrix}$$

onde n denota o número de registros no conjunto de dados, k representa a quantidade agrupamentos e u_{ij} satisfaz as seguintes restrições

$$u_{ij} \in \{0, 1\}, 1 \leq j \leq k, 1 \leq i \leq n, \tag{6.1}$$

$$\sum_{j=1}^{k} u_{ji} = 1, \ 1 \leq i \leq n, \tag{6.2}$$

$$\sum_{i=1}^{n} u_{ji} > 0, \ 1 \leq j \leq k. \tag{6.3}$$

A primeira restrição (6.1) implica em cada objeto pertencer ou não a um agrupamento. A restrição 6.2 força cada objeto a pertencer a apenas um grupo. Já a restrição 6.3 diz que cada grupo contém pelo menos um objeto, isto é, grupos vazios não são permitidos. $U = (u_{ji})$ é chamado de partição-k *hard* do conjunto de dados.

No Agrupamento Fuzzy, a assunção é relaxada e cada objeto pode pertencer a um ou mais grupos com certa probabilidade. O resultado matemático de um algoritmo desta categoria pode ser representado pela mesma matriz $U_{k \times n}$ com a primeira retrição relaxada

$$u_{ij} \in [0, 1], 1 \leq j \leq k, 1 \leq i \leq n, \tag{6.4}$$

Os algoritmos de Aglomeração Hard podem ser subdividios em *Algoritmos Hierárquicos* e *Algoritmos de Partição*. Existem dois tipos de Algoritmos Hierárquicos: Algoritmos Hierárquicos de Divisão e Algoritmos Hierárquicos de Aglomeração. Um *Algoritmo Hierárquico de Divisão* processa do topo para a base, isto é, o algoritmo inicia com um grupo

grande contendo todos os dados do conjunto e prossegue definindo subgrupos. Já um *Algoritmo Hierárquico de Aglomeração* processa da base para o topo, ou seja, inicia com vários grupos contendo um elemento do conjunto de dados e prossegue mesclando-os. Diferentemente dos algoritmos hierárquicos, os *Algoritmos de Partição* criam um nível único de não-sobreposição dos dados: eles geram uma única partição dos dados na tentativa de recuperar os grupos naturais presentes nos dados.

6.3 O Problema das *k*-médias

Como o problema de agrupamento (*clustering problem*) aparece em muitas e variadas aplicações, diversos métodos de solução têm sido desenvolvidos em diferentes disciplinas. Na literatura das comunicações (ou teoria da informação), um primeiro método de agrupamento foi sugerido por Stuart Lloyd para efetuar a quantização escalar de vetores [1] em 1957 [Lloyd 1982] (embora tenha sido publicado apenas em 1982), conhecido como algoritmo de Lloyd.

Um algoritmo similar foi proposto por [Macqueen 1967] para a área de reconhecimento de padrões e, neste trabalho, foi utilizado pela primeira vez o termo "*k-means*" (k-médias). Assim, o problema de agrupamento é, corriqueiramente, referenciado como "Problema das k-médias".

A exata definição do problema de agrupamento difere sutilmente de uma área para outra, mas, em todas elas, sua solução é uma robusta ferramenta para a análise ou o processamento de dados sem um conhecimento prévio da distribuição considerada. A enunciação do problema geralmente é feita de forma matematicamente precisa pela definição de um critério de custo a ser minimizado.

Seja D um conjunto de dados com n instâncias e seja C_1, C_2, ..., C_k os k grupos desconexos de D. Define-se a função de erro como

$$E = \sum_{i=1}^{k} \sum_{x \in C_i} d(x, \mu(C_i)), \tag{6.5}$$

onde $\mu(C_i)$ é o centróide do grupo C_i. O valor $d(x, \mu(C_i))$ denota a distância entre x e $\mu(C_i)$, e pode ser uma das diversas medidas de distância existentes (distância Euclideana,

[1] A quantização escalar de vetores é o processo de substituir um grupo de vetores parecidos por um vetor característico deste grupo, produzindo uma compressão de dados com perda de informação [Gersho e Gray 1991].

distância de Mahalanobis, distância de Manhattan, etc) [2].

Desta forma, o problema (de otimização) passa a ser a definição dos k grupos de forma a minimizar a função 6.5. [Dasgupta 2007] e [Aloise et al. 2009] indicam que, mesmo para $k \geq 2$, este problema é NP-difícil. [Mahajan, Nimbhorkar e Varadarajan 2009] prova esta característica inclusive para instâncias planares do problema (dimensão $d = 2$).

Um problema semelhante ao problema das k-médias é o "Problema das k-medianas" (*k-Medians Problem*), no qual os k centróides são escolhidos dentre os elementos que compõem o conjunto D de dados. Este problema é uma espécie de especialização do "Problema de Localização de Facilidades" (*Facility-Location Problem*), num formato não-capacitado. Nessa mesma linha, há ainda o "Problema das p-medianas" (*p-Medians Problem*), em que o grupo de p centróides é escolhido em um conjunto pré-fixado diferente de D. Por fim, o Problema dos k-centros (*k-Center Problem*) busca encontrar os k centróides que minimizam a máxima distância de qualquer elemento ao seu respectivo centróide [3].

Diversos algoritmos e muitas variantes do algoritmo de Lloyd foram propostos para solucionar o problema das k-médias. [Inaba, Katoh e Imai 1994] propõem uma conjugação de diagramas ponderados de Voronoi e aleatoriedade para a solução aproximada do problema. O primeiro algoritmo $(1 + \epsilon)$-aproximado de tempo linear para o problema das k-médias é apresentado em [Kumar, Sabharwal e Sen 2004]. Uma técnica de definição aleatória de semente é aplicada por [Arthur e Vassilvitskii 2007] para obter um algoritmo $\Theta(\log k)$-competitivo com a solução ótima, chamado de algoritmo *k-means++*. [Kanungo et al. 2002] apresenta um algoritmo simples e eficiente implementação, chamada de algoritmo de filtragem. A desigualdade de triângulos é utilizada em [Elkan 2003] para acelerar a busca pela solução de mesma qualidade do algoritmo padrão.

O algoritmo de Lloyd é tão citado na literatura para solucionar o problema apresentado que também é conhecido por *Algoritmo k-Means* [4]. Formalmente, pode ser apresentado como a seguir:

> Seja $D = \{x_1, x_2, ..., x_n\}$ um conjunto com n pontos sobre o plano Cartesiano. Dado um conjunto inicial de k médias $m_1^{(1)}, ..., m_k^{(1)}$, especificadas aleatoriamente (ou segundo alguma heurística), o algoritmo prossegue alternando entre dois passos:

[2] Alguns autores indicam especificamente a "média quadrática das distâncias" como a medida *formal* para o problema das k-médias.

[3] A literatura pesquisada nem sempre foi consensual nas terminologias. Optou-se, aqui, por descrições genéricas a título informativo.

[4] Uma descrição formal é apresentada por [MacKay 2005].

Passo de Associação. Associe cada ponto x_j de D ao agrupamento $S_i^{(t)}$ cujo centróide m_i é o mais próximo (distância Euclideana).

$$S_i^{(t)} = \{x_j : \parallel x_j - m_i^{(t)} \parallel \leq \parallel x_j - m_{i^*}^{(t)} \parallel,\ \forall\ i^* = 1, ..., k\} \tag{6.6}$$

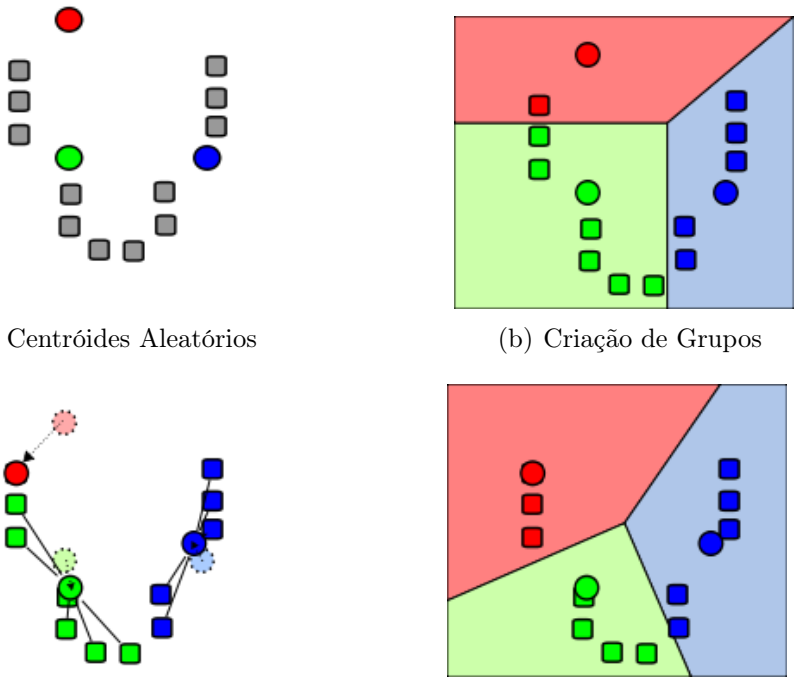
Passo de Atualização. Calcule o novo centróide para cada agrupamento.

$$m_i^{(t+1)} = \frac{1}{|S_i^{(t)}|} \sum_{x_j \in S_i^{(t)}} x_j \tag{6.7}$$

A convergência é atingida quando nenhuma associação é atualizada novamente.

Por ser um algoritmo heurístico, não há garantia de convergência para um ótimo global, sendo que os resultados obtidos dependem dos agrupamentos incialmente definidos. [Arthur e Vassilvitskii 2006] demonstra que para um certo agrupamento inicial, o tempo de convergência pode tornar-se *superpolinomial*: $2^{\Omega(\sqrt{n})}$. Uma prática corriqueira é a adoção de um critério de parada após algumas alternâncias entre os passos 6.6 e 6.7.

A Figura 21 ilustra uma iteração do algoritmo para $k = 3$ e um conjunto D com $n = 12$ pontos.



(a) Centróides Aleatórios

(b) Criação de Grupos

(c) Atualização de Centróides

(d) Redefinição de Grupos

Figura 21: Algoritmo k-Médias: exemplo de iteração

7 *AnoniMobi*, a Abordagem Proposta

"*All human knowledge begins with intuitions, thence passes to concepts and ends with ideas.*"

Kant, *Kritik der reinen Vernunft Elementariehre*, Part 2, Sec. 2.

O capítulo apresenta a abordagem proposta para a obtenção de informações de SBLs públicos de forma privada, isto é, sem revelar identidade/localização do usuário consultante.

A região de anonimização (ASR) proposta traz uma inovação ao substituir o uso de retângulos (ou quaisquer figuras planas) por distribuições aleatórias de k pontos médios. Essa medida propicia a obtenção de informaões adicionais acerca dos pontos de interesse, possibilitando a obtenção de rotas até tais pontos.

Adicionalmente, uma arquitetura de rede P2P, baseada na arquitetura CAN, é estruturada visando otimizar o tempo de processamento e de transmissão de informações entre os diversos nós.

7.1 Visão Geral

A composição da estrutura proposta para proporcionar a um usuário consultante efetuar pesquisas anônimas a SBLs públicos envolve:

- Um servidor de autenticação para ingresso dos usuários móveis no sistema;
- Uma estratégia de anonimização da localização geográfica do usuário consultante.
- Uma infraestrutura P2P para efetiva comunicação entre os nós (usuários);
- Um conjunto de operações básicas de troca de mensagens entre os nós;

A seção 7.2 tece comentários referentes ao *Servidor de Auntenticação* do sistema, responsável pelo ingresso de usuários na rede P2P. Na 7.3 são apresentados os conceitos da estratégia adotada para a construção da região espacial de anonimização de nível K. Detalhes sobre a métrica de proximidade adotada são fornecidos na seção 7.4. A arquitetura P2P elaborada é abordada nas seções 7.5 e 7.6. A *consulta-K*, operação objetivo dos usuários do sistema, é detalhada na seção 7.7, e os problemas de privacidade envolvendo esta operação são tratados na seção 7.8. As diversas operações de comunicação necessárias para a manutenação do sistema são discutidas na seção 7.9. Por fim, são feitas algumas últimas considerações na seção 7.10.

A Figura 22 ilustra o cenário caracteristico. A "rede *AnoniMobi*" é formada por nós móveis e estáticos que se comunicam através de um protocolo pré-definido (vide seção 7.9). Para ingressar no sistema, os usuários precisam primeiramente autenticar-se no "Servidor de Autenticação". Uma vez integrados à rede, os usuários podem efetuar consultas anônimas a SBLs públicos.

Ressalta-se que o foco adotado neste estudo concentra-se no aspecto de "*privacidade* das comunicações entre um SBL público e seus usuários". Soluções para outras problemáticas de segurança não são cobertas pela presente proposta.

7.2 Servidor de Autenticação

Para fazer parte da rede de anonimização, um usuário precisa primeiramente ser autenticado pelo sistema. Um Servidor de Autenticação dedicado faz esse papel, exigindo *login* e *senha* para acesso à rede de anonimização.

Figura 22: Cenário *AnoniMobi*

Esse servidor não fica sobrecarregado, pois, diferentemente das propostas que se utilizam de um servidor anonimizador, o servidor de autenticação não acumula a tarefa de intermediar a comunicação do usuário consultante com o SBL.

Com essa postura, não existe o risco de as consultas anônimas serem interceptadas caso o Servidor de Autenticação seja comprometido. Além disso, uma falha nesse servidor não impede o sistema de continuar funcionando, impossibilitando apenas a entrada de novos usuários por falta de ponto de acesso.

Isto posto, a sistemática de acesso ao sistema é a que se segue. Um usuário autentica-se no Servidor de Autenticação, fornecendo *login* e *senha*. Devidamente autenticado, o usuário indica ao servidor sua posição geográfica (vide seção 7.4), obtendo os endereços IP dos prováveis líderes de grupos aos quais o usuário ingressante deverá pertencer (vide seção 7.5).

O usuário, então, tenta contactar os referidos líderes, indicando sua intenção em ingressar no sistema. Ao se comunicar com um deles, recebe a mensagem ou de estar vinculado a um grupo ou de precisar contactar outro líder (cujo endereço IP é fornecido).

Após vincular-se a um grupo, o usuário torna-se apto a efetuar pesquisas anônimas a SBLs públicos.

Para fornecer os endereços IPs dos prováveis líderes de grupos aos quais o usuário ingressante deverá pertencer, o Servidor de Autenticação necessita da informação de posição geográfica dos líderes, bem como de seus respectivos endereços IP. Desta forma, o Servidor de Autenticação acumula a tarefa de receber atualizações periódicas das posições dos líderes de grupo. Essas "requisições" de atualização têm prioridade inferior ao atendimento a requisições de autenticação, tarefa primordial do Servidor de Autenticação.

À medida que o número de usuários aumenta, a carga de atualizações efetuadas pelos líderes pode crescer a ponto de o Servidor de Autenticação não conseguir atendê-las a contento. A escalabilidade do sistema pode ser alcançada com a utilização de múltiplos Servidores de Autenticação comunicando-se entre si e dividindo a tarefa de atualizações por regiões geográficas.

7.3 Estratégia de Anonimização

Na bibliografia disponível, é comum encontrar abordagens de anonimização baseadas em uma K-ASR (vide capítulo 2), nas quais o retângulo definido precisa, ao menos, englobar K usuários (incluindo o usuário consultante). Apesar de efetivo no tocante à consulta a SBLs públicos com níveis satisfatórios de anonimidade, este método deixa a desejar em situações nas quais o usuário necessita de orientações adicionais para alcançar o alvo desejado, isto é, rotas bem-definidas indicando como chegar ao destino indicado [Vieira, Martinello e Marcondes 2009]. A abordagem proposta neste trabalho desvincula-se da figura plana do "retângulo", substituindo-a por um conjunto de pontos.

A ideia-chave continua baseando-se na coleta das coordenadas dos $K-1$ usuários mais próximos (segundo alguma métrica de proximidade) do usuário consultante, com a diferença de que a K-ASR será representada por k pontos médios (agrupamentos ou *clusters*) definidos, inicialmente, de forma aleatória [1]. As coordenadas dessas k-médias (que compõem a k-ASR) são enviadas ao SBL, que retorna rotas destes k pontos para as coordenadas dos pontos de interesse da pesquisa mais próximos. Estas rotas obtidas podem ser filtradas e apenas a rota mais próxima do usuário consultante é efetivamente

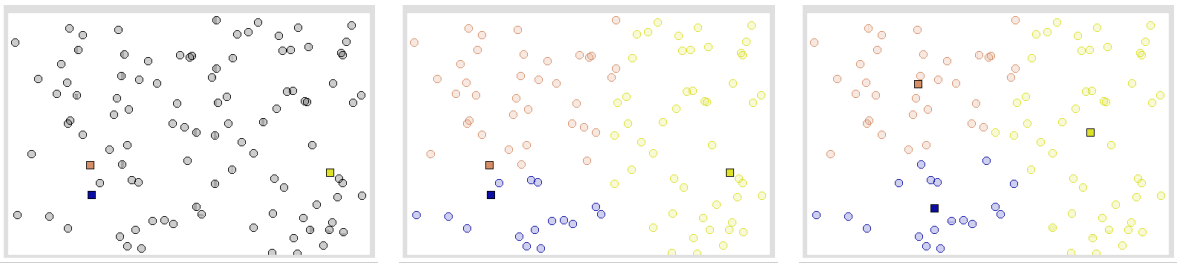
[1] Mesmo correndo o risco de gerar uma confusão inicial, optou-se por utilizar a mesma letra tanto para a reprentação da quantidade de pontos da ASR (K maiúsculo) quanto para a indicação da quantidade de pontos médios (k minúsculo), devido à quase unanimidade no uso desta letra para ambas as situações na literatura pesquisada. Desta forma, comparações futuras com outros trabalhos não sofrerão com a incômoda troca de nomenclatura.

exibida para orientação espacial.

O método indicado, isto é, o uso de k pontos médios como região espacial de anonimização, traz a vantagem adicional de dificultar a inferência, por parte de um atacante, dos usuários participantes do processo de anonimização. Na K-ASR tradicional, tais usuários encontram-se inseridos em um retângulo, facilitando o trabalho do atacante que, no pior caso, conhece as coordenadas de todos os usuários do sistema. Com o uso de pontos médios, não há qualquer indicação explícita neste sentido.

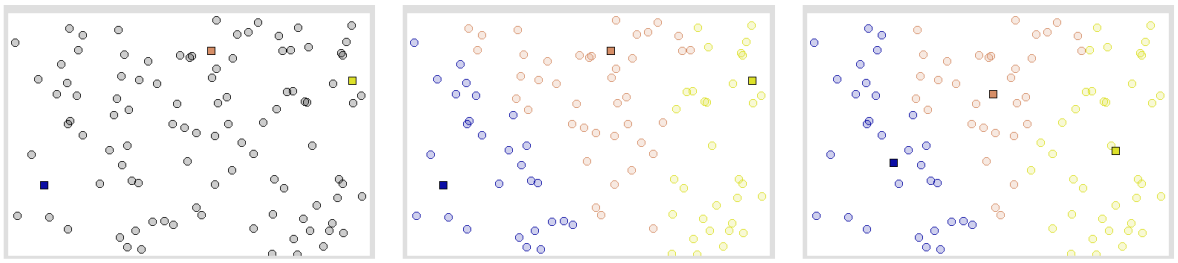
Entretanto, ainda assim, dependendo do nível K de anonimidade escolhido, da quantidade de usuários no sistema e do poder de processamento disponível, teoricamente é possível obter em tempo hábil o conjunto de pontos que gerou a ASR fornecida ao SBL.

Pode-se contornar o problema com o uso do algoritmo k-médias (algoritmo de Lloyd) para a definição dos pontos da k-ASR [Vieira et al. 2010]. Conforme explanado na seção 6.3, este algoritmo não garante o fornecimento de um ótimo global. Desta forma, execuções distintas do algoritmo tendem a fornecer resultados distintos, devido à aleatoriedade na escolha dos pontos iniciais. Esta situação dificulta a inferência dos pontos geradores dos centróides de cada agrupamento e, consequentemente, fornece um nível adicional de anonimidade ao processo de pesquisa. As Figuras 23 e 24 ilustram duas simulações com resultados distintos para o mesmo conjunto de 100 pontos, admitindo-se $k = 3$.



(a) Posição inicial dos centróides (b) Agrupamento inicial (c) Posição final dos centróides

Figura 23: 1ª Simulação do algoritmo k-médias



(a) Posição inicial dos centróides (b) Agrupamento inicial (c) Posição final dos centróides

Figura 24: 2ª Simulação do algoritmo k-médias

Como o retorno fornecido pelo SBL considera como ponto de partida os k centróides, não há garantia de que alguma das rotas passe próximo ao usuário consultante. Na verdade, os centróides funcionam como um ponto único de atendimento a todos os usuários pertencentes ao respectivo agrupamento. Evidentemente, quanto maior o valor de k, maior a probabilidade de o usuário consultante ser atendido a contento.

É possível, ainda, solicitar ao SBL público rotas que interliguem os centróides, atitude que possibilita uma cobertura mais eficiente da área em que se encontram os usuários participantes do processo de anonimização. Contudo, o aumento na quantidade de rotas aumenta, proporcionalmente, o esforço computacional no cliente para a definição da rota mais próxima. Consequentemente, o tempo de resposta também se eleva, situação que exige um equilíbrio entre a *qualidade da resposta* obtida e o *tempo de espera* pela mesma. A Figura 25 ilustra um possível retorno do SBL conforme a proposta apresentada (usuário em verde, centróides em azul).

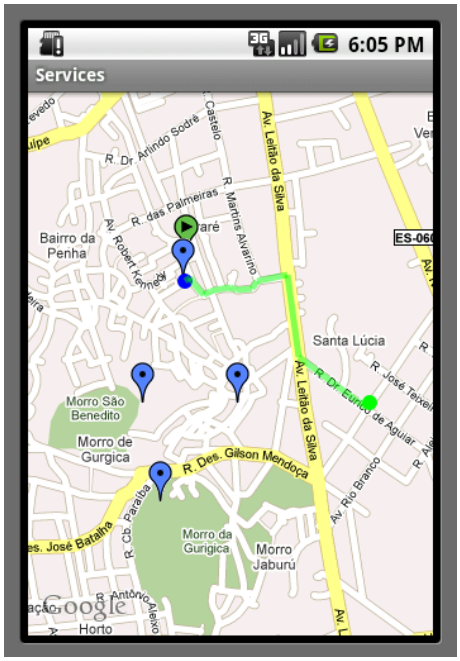


Figura 25: Retorno do SBL

7.4 Curva de Preenchimento: métrica de proximidade

Uma métrica de proximidade precisa ser adotada para a definição dos $K-1$ usuários cujas coordenadas ajudarão a compor a k-ASR enviada ao SBL pelo usuário consultante. Mais do que uma simples métrica de proximidade, uma estratégia eficiente de ordenação precisa ser estabelecida para que as consultas por usuários mais próximos sejam realizadas

com rapidez e de forma distribuída.

Conforme explanado no Capítulo 5, as curvas de preenchimento operaram um mapeamento 2-D para 1-D ao atribuir valores únicos para coordenadas no plano. Mais especificamente, a Curva de Hilbert executa tal mapeamento com um bom índice de preservação da localidade dos dados originais. Assim, essa curva foi escolhida para compor a estrutura da proposta apresentada neste trabalho.

Efetuado o mapeamento, é possível ordenar os pontos mapeados sob a lógica de Hilbert, de tal forma que o grau de proximidade é avaliado por comparações simples de valores sobre uma reta, conforme ilustra a Figura 4(a) (reproduzida, aqui, na Figura 26). Observa-se que para $p8$ (57), por exemplo, os dois pontos mais próximos são $p7$ (36) e $p9$ (62) (apesar de não o serem sob a óptica Cartesiana). Desta forma, fica fácil perceber que ao entrar no sistema, o usuário u (9) irá situar-se entre os pontos $p1$ (6) e $p2$ (15).

A proposta deste trabalho considera que os usuários móveis irão se auto-organizar em grupos segundo o grau de proximidade e eleger um nó como líder para a comunicação com outros grupos. No exemplo da Figura 26, qualquer um dos 10 usuários poderia ser escolhido como líder, ficando este responsável por intermediar requisições a outros grupos auto-organizados de forma idêntica.

Se algum usuário deste grupo solicitar ao líder as coordenadas de 49 outros usuários para compor sua k-ASR e efetuar uma pesquisa anônima com um nível de anonimidade $K = 50$, serão necessárias requisições aos líderes de grupos vizinhos para completar a quantidade de coordenadas demandada.

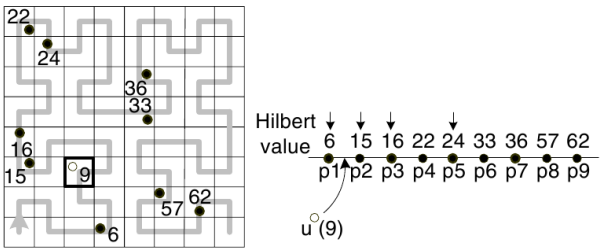


Figura 26: Métrica de proximidade: Curva de Hilbert

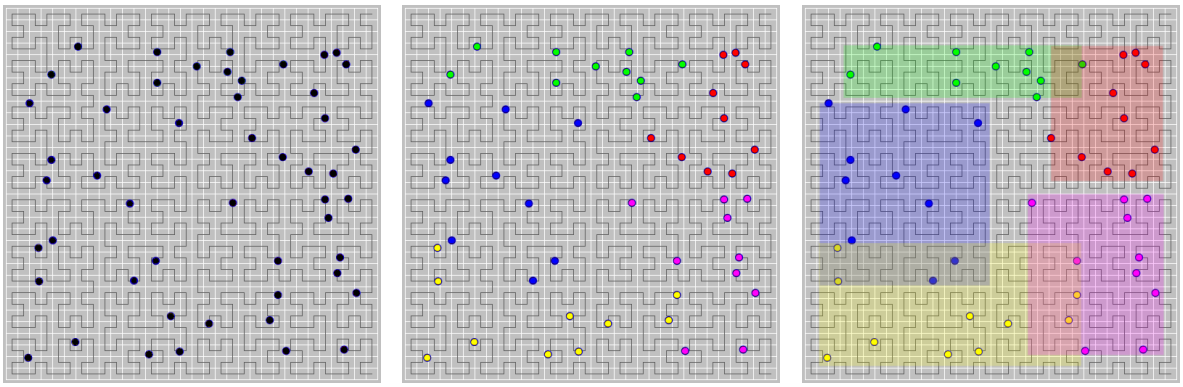
7.5 Infraestrutura P2P

Para auto-organizarem-se em grupos, os usuários móveis precisam de uma infraestrutura P2P que lhes indique claramente as regras de agrupamento a serem seguidas. A

proposta deste trabalho baseia-se na infraestrutura conhecida como *Content-Addressable Network* (CAN), apresentada brevemente no Capítulo 4.

A arquitetura CAN vem ao encontro das necessidade de auto-organização requeridas nesta proposta na medida em que prioriza as comunicações entre nós (ou grupos) vizinhos. Se os grupos vizinhos organizam-se sob a lógica de proximidade da curva de Hilbert, tem-se, então, que requisições por coordenadas de usuários mais próximos devem ser feitas aos próprios líderes de grupos vizinhos.

A Figura 27 exemplifica a ideia para 50 usuários distribuídos aleatoriamente sobre um plano mapeado em 1024 (32x32) quadrantes de Hilbert. Em 27(b), os 50 usuários são agrupados de 10 em 10, conforme o grau de proximidade indicado pela curva de Hilbert. Tais agrupamentos geram regiões na camada CAN, delimitadas pelas coordenadas extremas de seus usuários componentes.



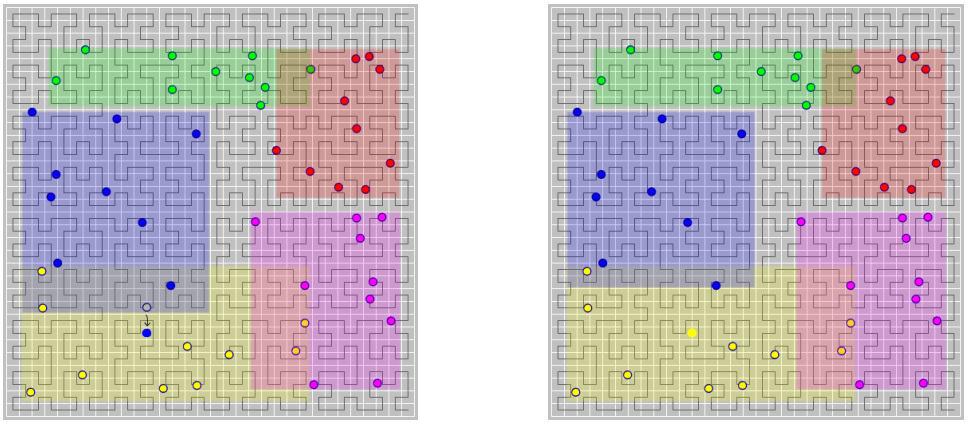
(a) 50 usuários sobre o plano (b) Grupos de 10 usuários (c) Regiões CAN geradas

Figura 27: Formação de grupos: Hilbert + CAN

Como pode ser observado, o resultado final são regiões parcialmente sobrepostas (retângulos coloridos) sob a coordenação de um usuário líder. Caso algum usuário requisite uma quantidade de coordenadas superior ao número de componentes do grupo, seu líder contactará o líder do grupo vizinho, solicitando uma quantidade complementar de coordenadas. No exemplo, dada a curva de Hilbert, a região vermelha é considerada vizinha da região verde, assim como essa última também é vizinha da região azul. Por sua vez, a região amarela é vizinha das regiões azul e rosa. Contudo, apesar de encontrarem-se geograficamente lado a lado, as regiões vermelha e rosa não são consideradas vizinhas por estarem distantes na curva de Hilbert. São consideradas regiões "colaterais não-vizinhas".

No contexto de usuários móveis, as dimensões das regiões CAN definidas para cada grupo alteram-se constantemente. Se os usuários dirigirem-se para o centro da região, esta irá "encolher" gradativamente, ao passo que movimentos em direções opostas tendem a

expandir o retângulo. É possível, inclusive, que os usuários migrem de região à medida que se deslocam no plano. A Figura 28 ilustra o caso de um usuário da região azul migrando para a região amarela.



(a) Migração para a região amarela (b) Redução da região azul

Figura 28: Migração de região CAN

A migração ilustrada na Figura 28 processar-se-ia sem qualquer dificuldade: o usuário migrante comunica ao líder de seu grupo a alteração de sua coordenada; o líder percebe que a nova posição encontra-se fora dos limites sob sua responsabilidade e informa ao usuário migrante a necessidade de se comunicar com o líder de outro grupo; o usuário se comunica com esse líder e recebe a informação de nova pertinência; o líder do antigo grupo atualiza as coordenadas da região sob sua responsabilidade, informando aos líderes vizinhos sobre a atualização efetuada.

Duas questões pertinentes surgem com relação à migração de usuários: quantidades mínima e máxima de usuários em cada grupo; migração de usuários entre regiões "colaterais não-vizinhas".

7.5.1 Parâmetros α e β

Para manter a gerência de um grupo sob controle, sem sobrecarga para os líderes, é necessário estabelecer um limite superior para a quantidade de usuários em cada grupo. Ao mesmo tempo, é preciso definir um limite inferior para essa quantidade com o intuito de amenizar a frequência de troca de mensagens entre líderes durante requisições de K coordenadas. Denotar-se-á por α o limite inferior e por β o limite superior.

No exemplo da Figura 28, tanto α quanto β precisam ser diferentes de 10 para que as migrações ocorram sem quaisquer outras modificações. Uma possibilidade seria $\alpha = 5$

e $\beta = 15$.

Caso o líder de algum grupo perceba o decréscimo na quantidade de componentes a um valor inferior ao de α, uma operação de junção a um dos grupos vizinhos deve ser iniciada. Por outro lado, ao perceber que a quantidade de usuários componentes de seu grupo ultrapassou o valor de β, o líder deve iniciar o processo de subdivisão do grupo em duas partes.

7.5.2 Migração entre regiões "colaterais não-vizinhas"

Líderes de grupos vizinhos comunicam-se diretamente, conforme a filosofia da arquitetura CAN. Diferentemente, líderes de grupos não-vizinhos precisam rotear suas comunicações entre líderes intermediários. No exemplo da Figura 28, as mensagens trocadas entre os líderes das regiões vermelha e rosa precisariam ser encaminhadas através dos líderes das regiões verde, azul e amarela.

Extrapolando o raciocínio, percebe-se que pode ocorrer de líderes colaterais necessitarem rotear suas mensagens por diversos outros líderes para efetivarem suas comunicações. O incoveniente dessa situação revela-se no contexto das constantes migrações inerentes à realidade de usuários móveis. O tempo de comunicação entre líderes colaterais precisa ser minimizado para garantir que usuários migrantes obtenham informações acerca do novo líder em tempo hábil.

Para contornar a problemática, cada líder deve armazenar o endereço IP dos líderes das regiões colaterais (vizinhas ou não), bem como as coordenadas das regiões sob suas responsabilidades. Assim, num primeiro momento, líderes de regiões colaterais não-vizinhas precisarão rotear mensagens de "descoberta de vizinhança" através dos líderes de regiões intermediárias. Após a conclusão das descobertas, os líderes colaterais passam a se comunicar diretamente, atualizando periodicamente suas informações.

Com essa medida, um usuário que migrasse da região vermelha para a região rosa, por exemplo, não sofreria com atrasos de comunicação para conhecer seu novo líder.

7.6 Reciprocidade

No Capítulo 2, foi apresentado o conceito de *Reciprocidade*, utilizado para desenvolver as estratégias de anonimização da arquitetura Privé [Ghinita, Kalnis e Skiadopoulos 2007]. Nesta arquitetura, boa parte do esforço de troca de mensagens entre os líderes dos grupos

está direcionado para o cálculo do $rank_u$, que é a posição do usuário u na estrutura global de valores ordenados de Hilbert, correspondente a todos os usuários do sistema. Esta grandeza possibilita a definição da K-ASR a ser utilizada durante uma pesquisa anônima (operação *consulta-K*).

Em Privé, o custo para tal operação é da ordem de $O(N + K)$ e torna-se impraticável para $N > 10.000$ [Ghinita, Skiadopoulos e Kalnis 2007]. A arquitetura MobiHide [Ghinita, Skiadopoulos e Kalnis 2007] estrutura-se sobre um anel Chord e consegue efetuar a operação *consulta-K* em $O(K)$. Não há qualquer preocupação com a manutenção da característica de reciprocidade das K-ASR definidas, pois a organização em anel permite a utilização de um parâmetro l para a garantia do nível K de anonimidade requerido.

Na proposta deste trabalho, a reciprocidade também é irrelevante. Como a região de anonimização definida baseia-se em pontos, não há qualquer figura plana (retângulos) envolvendo os usuários participantes do processo de anonimização. Desta forma, um atacante que conheça a posição de todos os usuários do sistema não possui qualquer subsídio para inferir quais destes usuários definiram os k centróides cujas coordenadas são enviadas ao SBL público. Para encontrar tal grupo de usuários sem conhecimento do valor de K escolhido pelo usuário u, o atacante precisaria testar inúmeras combinações diferentes de usuários, atividade que se configura inviável em tempo hábil.

No contexto dos grupos de usuários distribuídos em regiões CAN, a quantidade de troca de mensagens na rede P2P para efetuar a operação *consulta-K* dependerá dos parâmetros K, α e β. No pior caso (cada grupo está com a quantidade mínima de usuários permitida), tem-se $\frac{K}{\alpha}$ requisições para a construção da região de anonimização. Na média, são necessárias $\frac{2K}{\alpha+\beta}$ requisições, o que resulta em $O(\frac{K}{\alpha+\beta})$.

7.7 Consultas-K

Todos os usuários do sistema têm a liberdade de comunicarem-se uns com os outros. Entretanto, qualquer comunicação para efetuar a operação *consulta-K* é realizada entre o integrante de um grupo e seu líder, ou entre líderes de grupos vizinhos. A Figura 29 ilustra o processo. Quando um usuário, no passo 1, solicita ao líder de seu grupo as K coordenadas dos usuários mais próximos (incluindo a sua própria coordenada), o líder verifica se a quantidade de usuários atualmente no grupo é suficiente para atender à requisição. Não sendo, este líder comunica-se com líderes vizinhos para complementar as informações (passos 2 e 3) e retorna-as para o usuário solicitante (passo 4). A quantidade

de comunicações será determinada pelo valor de K e pelas quantidades de usuários em cada grupo (fortemente influenciadas pelos parâmetros α e β).

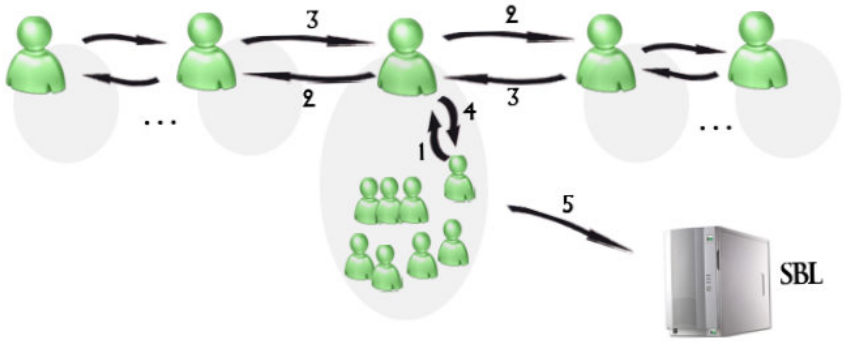


Figura 29: Cenário de uma consulta-K: passos de 1 a 5

A Figura 30 ilustra a situação. Os 15 retângulos azuis indicados pelas letras de A a O representam grupos de usuários (regiões CAN) cujos líderes são denotados por p_a, p_b, p_c, ..., p_o, respectivamente. Supondo-se os valores hipotéticos $\alpha = 20$ e $\beta = 40$, a Tabela 5 indica a quantidade de usuários em cada grupo. Tais grupos estão ordenados considerando suas posições sobre a curva de Hilbert, ou seja, A é vizinho de B, que é vizinho de C e assim sucessivamente.

Tabela 5: Quantidade de usuários por grupo (região CAN)

Grupo	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Nº Usuários	23	32	31	20	27	40	40	38	35	29	21	32	21	34	20

Supondo que o usuário u_g integrante do grupo G requisite a seu líder $K = 30$ coordenadas para compor uma região de anonimização, o líder p_g poderá fornecer as informação solicitada por si mesmo, já que o grupo G contém 40 usuários. Todavia, se u_g requisitar $K = 200$ coordenadas, p_g precisará solicitar uma complementação de informação aos líderes de grupos vizinhos.

Como o grupo G possui 40 usuários, p_g precisa de 160 outras coordenadas para atender à requisição de u_g. Assim, ele requisitará 80 coordenadas a p_f e 80 coordenadas a p_h. De posse de apenas 40 coordenadas, p_f solicitará a p_e outras 40 coordenadas. Para atender a esta requisição, p_e (27) buscará obter com p_d 13 coordenadas e, então, retornará a informação solicitada por p_f. Por fim, p_f entregará a p_g as 160 coordenadas solicitadas. Analogamente, p_h obterá as outras 80 coordenadas e entrega-las-á a p_g. O líder do

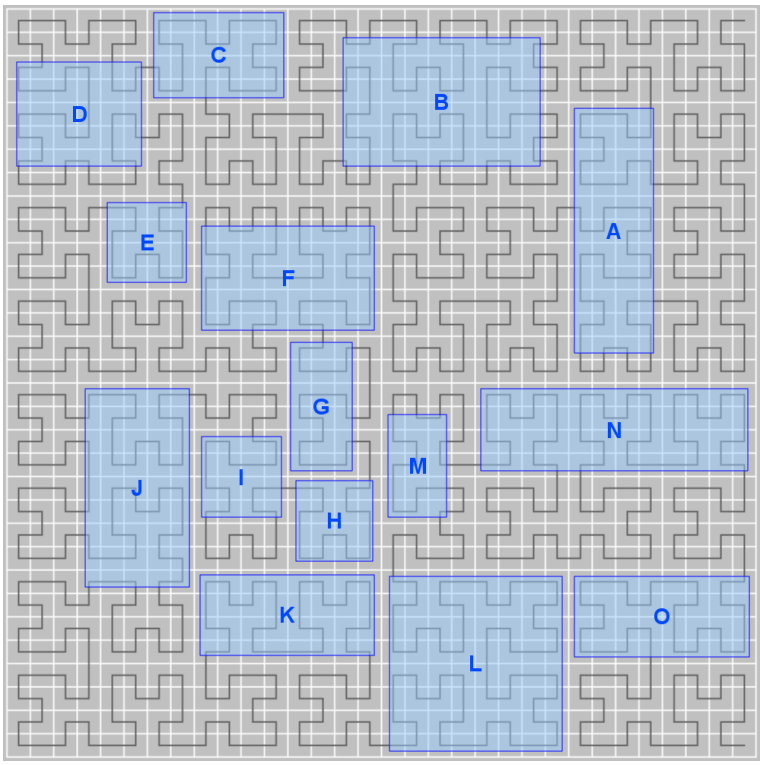


Figura 30: Grupos de usuários (regiões CAN)

grupo G, então, atenderá à requisição de u_g, que poderá calcular os k centróides cujas coordenadas serão enviadas ao SBL público.

7.8 Ataques Internos

Nos trabalhos pesquisados, a privacidade das consultas anônimas a SBLs públicos é tratada sob o aspecto de um atacante externo ao sistema que, no pior cenário, conhece a localização de todos os usuários do sistema. Conforme explanado na seção anterior, o uso de centróides como região espacial de anonimização (substituindo a clássica figura do retângulo) neutraliza a ação do atacante externo, pois, sem conhecer o parâmetro K utilizado e não possuindo qualquer subsídio que o oriente nesse sentido (como é o caso da confinação de usuários em uma área), a tarefa de descobrir quais usuários participaram do processo de anonimização torna-se impraticável.

Assim, o trabalho do atacante externo resume-se a descobrir o valor K utilizado e, a partir desta informação, tentar inferir o sub-grupo de usuários que gerou os centróides utilizados. Evidentemente, resta a situação em que o atacante encontra-se autenticado no sistema e é, portanto, parte do mesmo.

A *consulta-K*, operação sensível em termos de privacidade, é realizada através de comunicações do usuário consultante (u) com o líder de seu grupo (p) e, eventualmente, deste com outros líderes. Os líderes vizinhos não conseguem, por si sós, inferir a valor de K original, pois não têm como definir se a requisição feita pelo líder vizinho foi originada naquele grupo ou se está sendo repassada para complementar informações requisitadas. Desta forma, o elo fraco desta corrente de segurança é o líder p.

Além de p conhecer o valor do parâmetro K da consulta, u irá requisitar diretamente a ele as coordenadas dos usuários mais próximos, revelando-se como o autor de uma futura consulta ao SBL. Como os usuários de cada grupo trocam de líder periodicamente para efeito de balanceamento de carga, haverá um momento em que o atacante interno tornar-se-á líder de seu grupo, intermediando requisições do tipo *consulta-K*.

Para tentar neutralizar a ação deste atacante interno, o presente trabalho propõe que as operações *consulta-K* originais (isto é, de u para p) sejam feitas através de alguma infraestrutura que implemente camadas de conexão anônima, como *Onion Routing* [Goldschlag, Reed e Syverson 1999] e *Crowds* [Reiter e Rubin 1998]. Esta medida evita que o atacante interno identifique diretamente pela requisição qual o usuário consultante, sem contudo evitar que o nível K de anonimidade seja reduzido para a quantidade de usuários do grupo. No pior caso, ter-se-á $K = \alpha$ para as situações em que o atacante esteja exercendo o papel de líder de um grupo.

7.9 Comunicações entre usuários

Tanto as comunicações entre os usuários de um grupo e seu líder quanto aquelas entre líderes de grupos colaterais podem ocorrer via conexões TCP ou através de segmentos UDP (comunicações sem conexão). As primeiras, mais onerosas, estão reservadas para situações em que a entrega de informações seja imprescindível, ao passo que as últimas são utilizadas em situações menos exigentes. A seguir, algumas operações que exigem comunicações entre pares na arquitetura proposta são apresentadas como sugestão indicativa para futura implementação.

Ingresso em um Grupo

Esta é a primeira operação de um usuário após ser autenticado e aceito no sistema. Ao receber do *Servidor de Autenticação* (vide seção 7.2) alguns endereços IP de líderes de possíveis grupos aos quais possa vir a pertencer, o usuário entrante precisa se comunicar para tentar ingressar efetivamente na rede. A operação *ingressarGrupo*, executada através

de uma conexão TCP, é efetuada por iniciativa do usuário e tem como destino um dos líderes indicados. Caso o líder contactado seja o responsável pelo grupo de usuários que se encontram na região CAN que engloba a coordenada do usuário, este é aceito no grupo.

Como a dinâmica de mobilidade é intensa na rede P2P em questão, pode ocorrer de as coordenadas da região CAN terem sido alteradas e ainda não atualizadas na base de dados do Servidor de Autenticação. Neste caso, se a coordenada do usuário entrante não mais pertencer ao conjunto de coordenadas da região CAN contactada, o líder perceberá a falha e informará ao usuário a necessidade de continuar sua busca pelo grupo adequado, indicando o endereço IP ou do líder responsável por tal grupo ou do líder que poderá fornecer uma informação mais precisa (já que cada líder conhece apenas as regiões colaterais).

Atualização de Coordenadas

Para que o líder de um grupo possa gerenciar satisfatoriamente a região CAN sob sua responsabilidade, é preciso que os usuários o mantenham informados sobre suas localizações (coordenadas). A operação *atualizarCoodenada* permite que cada usuário envie periodicamente, ao líder de seu grupo, segmentos UDP informando a sua coordenada atual. Se uma dessas mensagens é perdida, o líder assume que o usuário não alterou sua posição desde a última comunicação. Se o usuário deixa de enviar atualizações por um período de tempo pré-determinado, assume-se que ele abandonou o sistema.

Migração de Grupo

Se um usuário afasta-se do centro da região CAN a qual pertence, a ponto de invadir os domínios geográficos de alguma região CAN vizinha, o líder de seu grupo deve avisá-lo da necessidade de migrar para outro grupo. Isto é feito através da operação *migrarGrupo* que faz uso de segmentos UDP para alcançar o usuário migrante. Caso a mensagem se perca, o usuário não será informado da nova situação e permacerá tentando se comunicar com o antigo líder, até ser devidamente orientado a entrar em contato com outro líder (atraves da operação *migrarGrupo*).

Eleição de líder

Os líderes de grupo concentram as tarefas de gerência, consumindo recursos importantíssimos do *hardware* subjacente, como largura de banda, processamento e, principalmente, bateria. Para minimizar tal sobrecarga, existe um rodízio periódico de liderança entre os componentes de cada grupo. Assim, a necessidade de eleição de um novo líder é anunciada pelo líder atual através da operação *elegerLider*. Segmentos UDP são enviados

a cada usuário, incitando-os a gerarem um número aleatório que será informado ao líder atual por meio da operação *votarLider* (também através de segmentos UDP). Este líder irá avaliar os valores coletados e, obedecendo a alguma regra pré-estabelecida, definirá o novo líder que será anunciado aos demais integrantes do grupo pela operação *anunciarLider* (segmentos UDP). As informações acerca do grupo de usuários da região CAN serão transmitidas pelo ex-líder para o novo líder através da operação *transferirLideranca*, executada sobre uma conexão TCP.

A Operação *consulta-K*

A finalidade da organização dos usuários do sistema proposto em uma arquitetura P2P é o fornecimento de informações (coordenadas) necessárias para a construção de uma região espacial de anonimização de nível K (K-ASR). Para tanto, um usuário faz uso da operação *consulta-K*, enviada para o líder de seu grupo, através de alguma infraestrutura organizada em camadas de conexão anônima (vide seção 7.8). Com esta operação executando sobre uma conexão TCP, requisita as coordenadas dos K usuários mais próximos de sua localização.

Como sua coordenada não é fornecida para preservar sua identidade frente a um possível falso líder (atacante interno), a resposta obtida sempre contém as coordenadas de todos os usuários do grupo ao qual pertence, complementadas por coordenadas de usuários de grupos vizinhos caso necessário. De posse da resposta obtida, o usuário efetua uma eventual filtragem dos dados e costrói a k-ASR adequada (conforme descrito na seção 7.7).

A complementação das K coordenadas a serem entregues ao usuário consultante é efetuada através da replicação da *consulta-K* (sem uso de camadas de conexão anônima), pelo líder, para os líderes das regiões vizinhas, ajustado-se adequadamente o parâmetro K.

Comunicação entre líderes

Além da *consulta-K*, outras comunicações entre os líderes de grupos são necessárias para a manutenção do correto funcionamento do sistema. Tais comunicações são efetuadas, não apenas entre líderes vizinhos, mas também entre líderes colaterais.

Periodicamente, o líder de grupo envia aos líderes colaterais informações sobre sua região CAN através da operação *informarRegiaoCAN*. Uma consulta explícita por tais informações pode ser efetuada com o uso da operação *consultarRegiaoCAN*. Ambas as operações executam utilizando segmentos UDP.

Alterações substanticiais nos grupos, como troca de liderança ou modificação radical na geografia da região CAN (junção ou subdivisão, seção 7.5.1), devem ser informadas imediatamente aos grupos vizinhos através da operação *atualizarRegiaoCAN* (conexão TCP) para não comprometer o funcionamento do sistema.

Comunicações com o *Servidor de Autenticação*

O Servidor de Autenticação também é informado, pelos líderes, sobre a situação de liderança de cada grupo por meio da operação *informarLideranca* (segmentos UDP). Essa mensagem periódica indica o atual líder e as coordenadas da região CAN em questão.

Caso o administrador do sistema modifique dinamicamente os valores dos parâmetros α e β com o intuito de alterar o desempenho, o Servidor de Autenticação repassa aos líderes de cada grupo a nova informação através da operação *atualizaParametros* (segmentos UDP). A mesma operação é utilizada para informar o endereço IP de outros servidores de autenticação eventualmente utilizados para efetuar balanceamento de carga (vide seção 7.2).

Obs.: como a liderança é alternada entre os usuários do grupo periodicamente, pode ocorrer de mensagens direcionadas a um líder serem entregues acidentalmente a um usuário ex-líder. Quando tal situação ocorrer, este ex-líder deverá esclarecer ao remetente a situação e informar-lhe o endereço IP do atual líder.

7.10 Considerações Finais

A rotatividade de nós nas redes P2P é intensa e configura-se como um dos grandes desafios no desenvolvimento de aplicações e compartilhamento de recursos. Um par, que em dado momento pode ser portador de determinada informação, por exemplo, no instante seguinte pode não mais ser integrante do sistema.

Para a abordagem de anonimização proposta, baseada na arquitetura P2P "Content-Addressable Network", a rotatividade torna-se um problema considerando-se o nó que assume o papel de líder de um grupo. Se algum desses nós responsáveis por gerenciar as regiões CAN, coordenar os grupos e intermediar as operações *consulta-K* torna-se inalcançável, a efetividade do sistema fica abalada, comprometendo a função de anonimização almejada. Já a falha de comunicação com outros nós, apesar de incoveniente, não chega a interromper o funcionamento do sistema e nem tão pouco impede o atendimento de requisições feitas a líderes de quaisquer grupos.

Para minimizar a ocorrência de falhas de comunicação com os líderes e dotar o sistema de maior robustez, existe o papel de *vice-líder*, nó integrante de um grupo que compartilha as informações de gerência com o líder deste grupo. Havendo alguma falha de comunicação com este líder, seja por problemas de conexão ou por desconexão proposital com o sistema, o vice-lider deve assumir a função, comunicando os demais integrantes do grupo (e os líderes de regiões CAN colaterais) sobre a nova situação.

8 Análise de Sensibilidade de Resultados

"*Não pense que o mundo acaba ali aonde a vista alcança. Quem não ouve a melodia acha maluco quem dança.*"

Oswaldo Montenegro, *Mudar Dói, Não Mudar Dói Muito.*

Este capítulo expõe os resultados de experimentos efetuados com o intuito de validar a abordagem proposta. A definição do valor do parâmetro k (quantidade de centróides) é avaliada como um compromisso entre a *qualidade* da resposta obtida do SBL e o tempo desta resposta.

Uma análise dos parâmetros α e β também é apresentada, com o objetivo de avaliar o impacto de seus valores no tempo necessário para execução da operação *consulta-K*.

Por fim, o desempenho da arquitetura P2P proposta é avaliado analiticamente e comparado com outras arquiteturas de mesmo propósito.

8.1 O parâmetro *k*

Na abordagem proposta, a quantidade "k" de centróides tem relação direta com a *qualidade* do retorno obtido do SBL, haja vista que um aumento no valor desta grandeza implica no aumento do número de rotas fornecidas e, consequentemente, na probabilidade de uma destas rotas ser útil para a correta orientação do usuário consultante.

Pode-se considerar uma rota como útil caso esteja a uma distância razoável do usuário, a ponto de ser possível obter uma informação de orientação geográfica suficientemente precisa para se alcançar o destino almejado.

Tendo-se em mente que, no contexto do problema de agrupamento, os pontos ótimos atendem igualitariamente e de forma "justa" (sem privilégios) a todos os componentes de cada grupo formado, uma avaliação do impacto que o número k de centróides exerce sobre a qualidade das respostas do SBL pode ser extraída por meio de execuções de algum algoritmo que solucione instâncias do problema de agrupamento.

Conforme explanado na seção 6.3, o algoritmo de agrupamento é extremamente custoso (NP-Difícil) e, à medida que a quantidade de usuários cresce, o tempo para a obtenção da solução ótima aumenta drasticamente. Assim, para efeito de breve análise, optou-se por conjuntos de apenas 10 usuários, buscando-se encontrar os k centróides ótimos para cada grupo, com $1 \leq k \leq 10$. A Figura 31 apresenta os resultados obtidos.

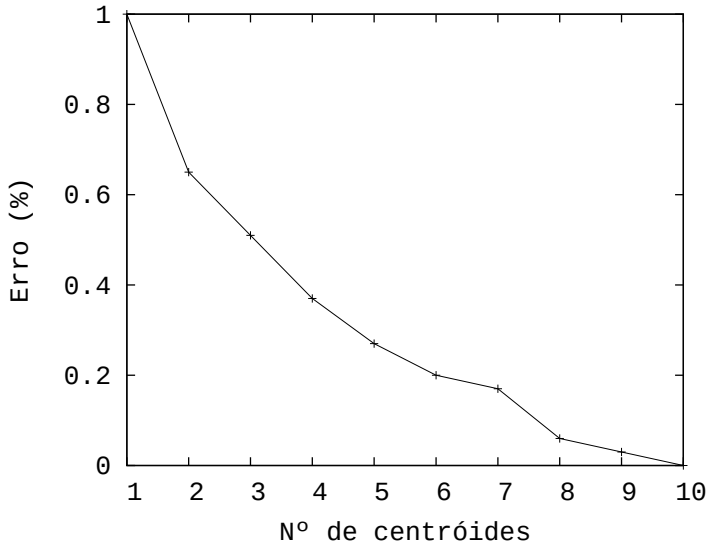


Figura 31: Impacto da quantidade pontos no percentual de erro

A grandeza *erro* refere-se ao somatório das distâncias de cada ponto ao seu respectivo centróide. Para apenas um centróide (centro de massa do sistema), atribuiu-se o erro

máximo (máximo somatório de distâncias). O erro mínimo ocorreu quando a quantidade de centróides igualou-se ao número de pontos ($k = 10$), resultando em um somatório de distâncias nulo.

O decaimento exponencial da curva obtida sugere o uso do maior número de centróides possível. Salienta-se, entretanto, que um número k de centróides igual a quantidade de usuários iria revelar quais destes usuários participaram do processo de anonimização, mantendo, contudo, o nível K de anonimidade escolhido.

Todavia, sendo extremamente custoso um algoritmo que retorne os pontos ótimos para o problema das k-médias, necessário se faz utilizar um algoritmo aproximado, como o algoritmo de Lloyd. Uma bateria de experimentos foi executada para determinar o comportamento do algoritmo face a um conjunto de 100 usuários localizados aleatoriamente sobre o plano. A distribuição dos usuários obedeceu a 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].

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 vínvulo até o alcance da condição de parada. Tornam-se, portanto, inutilizáveis. Classificam-se tais centróides sem vínculos de "não-associados". À medida que a quantidade k de centróides aumenta, o índice de não-associação altera-se, diferentemente, para cada distribuição utilizada. A Figura 32 resume os resultados.

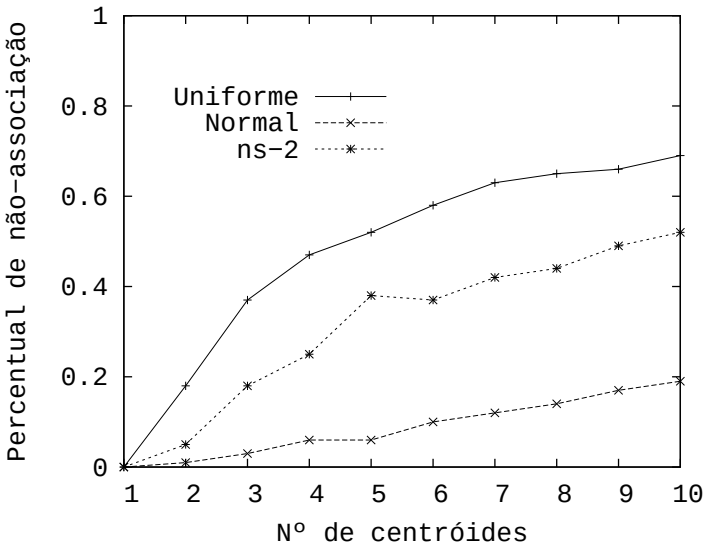


Figura 32: Percentual de centróides não-associados a qualquer usuário

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 (ou seja, para $k = 10$, têm-se 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 $k = 10$).

Os resultados demonstram que o aumento indiscriminado no valor de k não garante uma melhoria na qualidade da resposta do SBL obtida, haja vista que nem todos os centróides contribuem com rotas úteis aos usuários participantes do processo de anonimização.

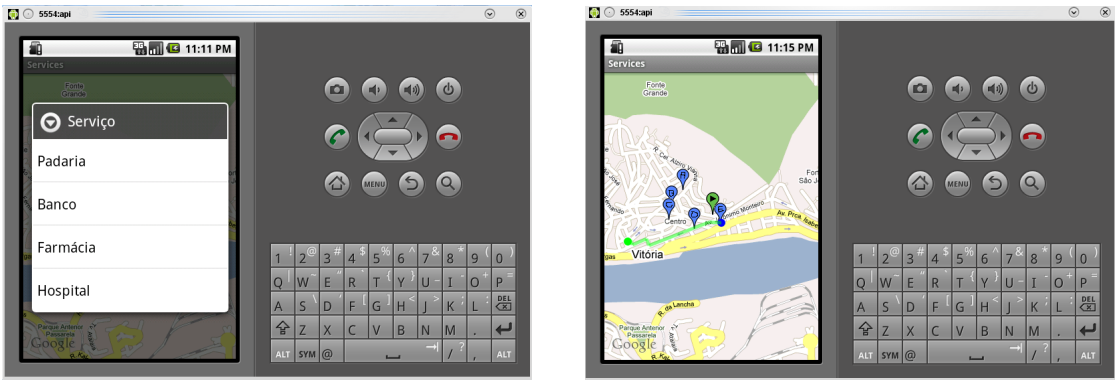
Além do índice de não-associação, outro fator que relaciona-se diretamente com a quantidade k de centróides é o tempo de execução de uma operação *consulta-K* e posterior requisição do usuário ao SBL por k rotas.

O tempo total de uma consulta anônima é afetado pelos seguintes componentes:

- t_{P2P}: tempo para a *consulta-K* na rede P2P;
- t_k: tempo de cálculo dos k pontos (centróides) de anonimização;
- t_t: tempo de transmissão dos dados do usuário consultante para o SBL;
- t_{SBL}: tempo de processamento, no SBL, para obtenção de rotas;
- t_r: tempo de recepção dos dados do SBL pelo usuário;
- t_u: tempo de processamento, no usuário, para filtragem de resultados (exibição apenas da rota mais próxima);

Com o intuito de avaliar o tempo de consulta, uma aplicação foi desenvolvida na plataforma Android [Google Inc. 2010], que serviu como ambiente de experimentação. A partir de um aparelho G1 conectado à Internet através de uma rede 3G, foram efetuadas pesquisas ao *GoogleMaps*. Este experimento avaliou o tempo de requisição $t_{requisicao}$ ($= t_t + t_{SBL} + t_r + t_u$) ao SBL para obter k (centróides) rotas, com $1 \leq k \leq 10$.

Na Figura 33(a) é apresentada a tela inicial da aplicação que solicita o tipo de estabelecimento a ser requisitado ao SBL. A resposta a uma consulta para $k = 5$ rotas é apresentada na Figura 33(b). Apenas a rota mais próxima do usuário consultante (bandeira verde) é exibida.



(a) Menu da aplicação Android (b) Retorno do SBL para $k = 5$

Figura 33: Ambiente de simulação da aplicação Android

O resultado do experimento é apresentado na Figura 34. Em média, leva-se cerca de 1s para obter cada rota.

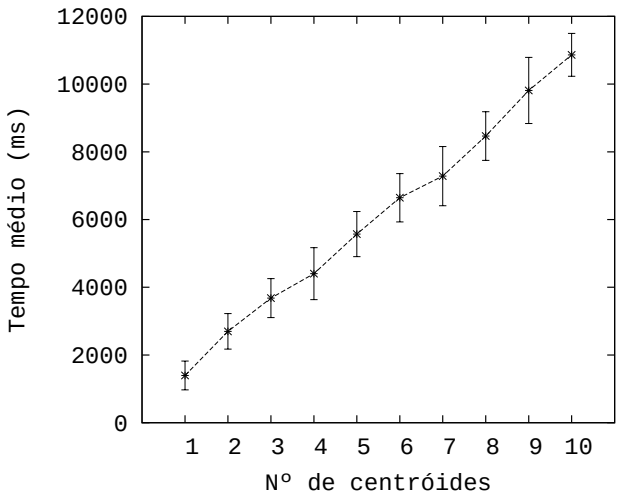
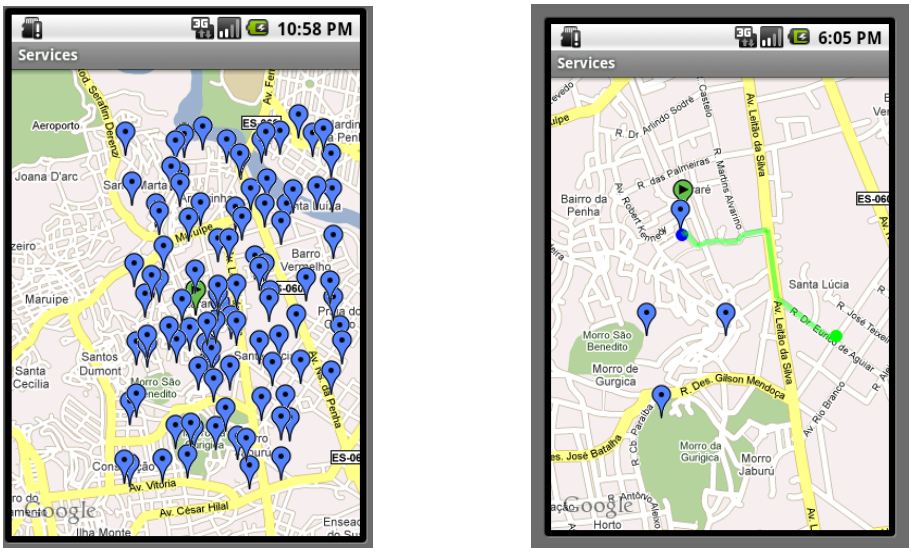


Figura 34: Tempo médio de requisição ao SBL

Outro teste utilizando o mesmo ambiente de experimentação (Android) buscou avaliar o tempo t_k necessário para execução do algoritmo de Lloyd no aparelho G1. A Figura 35 ilustra o processo implementado. Para uma quantidade de $K = 100$ pontos, executou-se o algoritmo variando a quantidade k de centróides de 1 a 10 e de 10 a 100, em passos de 10.

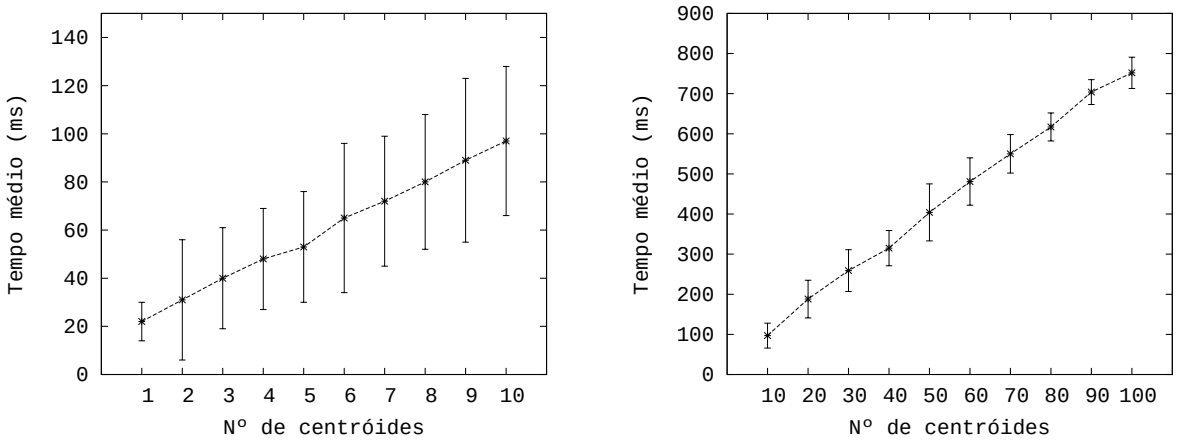
Os resultados para os dois conjuntos de dados são apresentados na Figura 36. A apesar do crescimento linear esboçado, o tempo de execução, para todas as instâncias de usuários, manteve-se abaixo de 1s, o que evidencia não ser o algoritmo de Lloyd um gargalo para implementações práticas.

Um último, mas não menos importante, elemento a ser considerado para a definição



(a) $K = 100$ usuários (b) $k = 4$ centróides

Figura 35: Implementação do algoritmo de Lloyd no ambiente Android



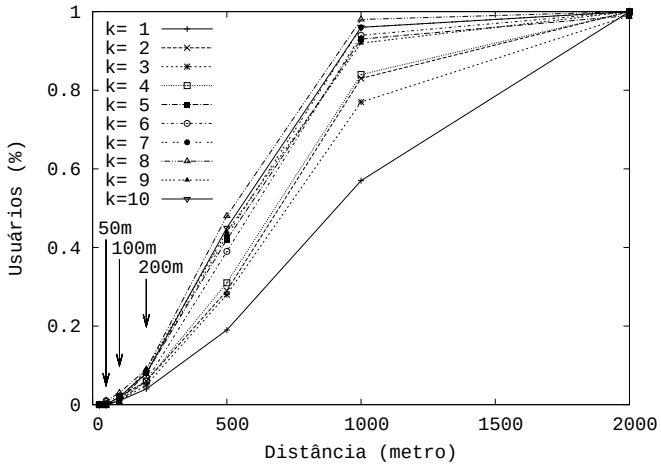
(a) 1 a 10 centróides (b) 10 a 100 centróides

Figura 36: Tempo médio de execução do algoritmo de Lloyd

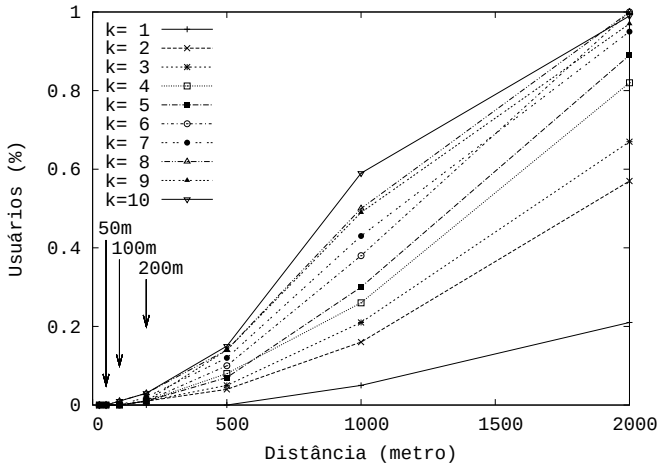
da quantidade k de centróides é a *qualidade* da resposta do SBL obtida. Considerando-se que as rotas fornecidas pelo SBL são dos centróides até os estabelecimentos de interesse, a medida dessa qualidade considera a distância D (variável aleatória) que cada usuário encontra-se de seu respectivo centróide.

Utilizando as 3 simulações de distribuição de usuários anteriormente citadas (Uniforme, Normal, Traços de Mobilidade), mediu-se estas distâncias. A Figura 37 apresenta a distribuição cumulativa de probabilidade $P[D \leq d]$ em função das distâncias d (eixo das abcissas). A Tabela 6 resume alguns dados relevantes para os valores $k = 4$ e $k = 5$.

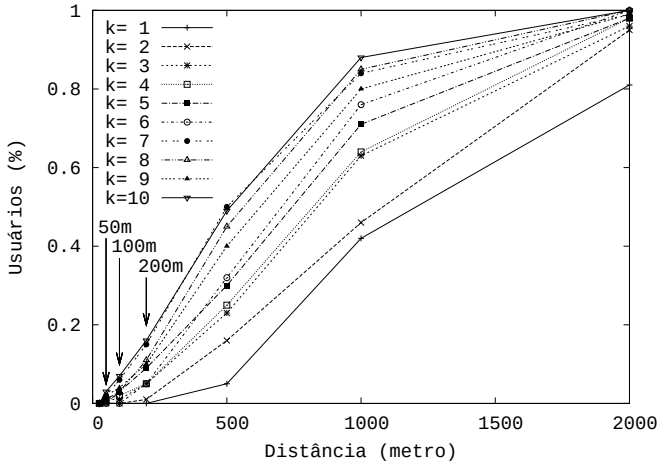
Observa-se que, para $k = 4$, tem-se de 54% a 92% dos usuários atendidos pelos



(a) Distribuição Uniforme



(b) Distribuição Normal



(c) Simulação via ns-2

Figura 37: Porcentagem de usuários e distâncias dos centróides

respectivos centróides para distâncias inferiores a 1500 metros, conforme a distribuição utilizada. Para $k = 5$, esses valores vão de 60% a 96%.

Tabela 6: Porcentagem de usuários para $k = 4$ e $k = 5$

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	$k = 4$	$0,25$	$0,64$	$0,81$	$0,98$
	$k = 5$	$0,30$	$0,71$	$0,85$	$0,98$

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 têm a mesma probabilidade de ocorrer, enquanto as distâncias médias na distribuição normal ocorrem com maior probabilidade, o que explica as diferenças obvervadas no resultado.

Analisando conjuntamente os resultados obtidos quando tem-se $k = 4$ e $k = 5$, um compromisso interessante pode ser observado entre a qualidade da resposta obtida (distância média dos usuários até seus respectivos centróides), o tempo de espera do usuário ($t_{requisicao} + t_k$) e o percentual de não-associação dos centróides calculados. A Tabela 7 resume alguns dos valores obtidos.

Tabela 7: Quadro comparativo para $k = 4$ e $k = 5$

Parâmetro avaliado	***N° de centróides***	
	$k = 4$	$k = 5$
Índice de não-associação (%)	0,06 a 0,47	0,06 a 0,52
Tempo de requisição $t_{requisicao}$ (ms) [a]	4403,1 [766,9]	5571,0 [665,3]
Tempo do k-médias t_k (ms) [a]	48,10 [21,52]	53,20 [23,18]
Distâncias dos centróides: $d < 1500m$ (%) [b]	0,54 a 0,92	0,60 a 0,96

[a] O valor entre "[ ]" indica o desvio-padrão calculado

[b] 1500 m equivalem, aproximadamente, a 11 quadras

Apesar de o tempo de execução da operação *consulta-K* não ter sido avaliado nestes experimentos (vide seção 8.2), os resultados sugerem a definição do parâmetro k em 4 ou 5, sem comprometimento para a qualidade da resposta do SBL e com um emprego moderado de recursos de *hardware*.

Salienta-se que o processo de comunicação com o *GoogleMaps* (SBL público) durante os testes evidenciou o fato de que a filtragem de resultados (exibição apenas da rota mais próxima do usuário) não envolve qualquer nova consulta ao serviço. Esta situação permite a adequada implementação da aplicação anônima, com a garantia de que a coordenada (identidade) do usuário consultante permaneça oculta.

8.2 O parâmetro α

Conforme a abordagem proposta no presente trabalho, o tempo total para a execução de uma consulta anônima em um SBL público inclui o tempo de pesquisa na rede P2P subjacente (t_P2P) para a obtenção das K coordenadas mais próximas sobre a curva de Hilbert.

O tempo t_P2P está intimamente relacionado com o parâmetro α, pois a menor quantidade de usuários de cada agrupamento define o limite superior para o número de mensagens a serem trocadas entre os líderes para o atendimento da requisição de uma operação *consulta-K*. Quanto menos usuários em cada região CAN, mais líderes precisam ser consultados e, consequentemente, mais mensagens são trocadas entre eles.

Denotando-se por N a quantidade de líderes consultados durante uma operação *consulta-K*, a quantidade máxima de nós (em função do parâmetro α), é fornecida por

$$N_{max} = 1 + 2 \cdot \left\lceil \frac{K - \alpha}{2 \cdot \alpha} \right\rceil \tag{8.1}$$

Esta equação considera que o valor K de coordenadas será requisitado simetricamente para cada lado da curva de Hilbert, isto é, $\lceil K/2 \cdot \alpha \rceil$ para cada sentido. Como o líder do grupo em que se encontra o usuário consultante conhece α coordenadas, apenas a quantidade complementar $K - \alpha$ é requisitada ao sistema P2P.

Para cada líder consultado, um par de mensagens é trocado: uma mensagem para a requisição e outra para a resposta. Assim, a quantidade máxima de mensagens q_{max} a serem trocadas na rede P2P é dada por

$$q_{max} = 2 + 2 \cdot N_{max} \tag{8.2}$$

Ou seja, 2 mensagens entre o usuário consultante e seu líder e $2 \cdot N_{max}$ mensagens complementares entre líderes. Escrevendo em função de α, tem-se

$$q_{max} = 2 + 4 \cdot \left\lceil \frac{K - \alpha}{2 \cdot \alpha} \right\rceil \tag{8.3}$$

Observa-se, portanto, que o tempo de execução da operação *consulta-K* independe da quantidade de usuários na rede P2P, característica extremamente desejável para a manutenção da escalabilidade do sistema de anonimização. A quantidade de usuários irá limitar, contudo, o valor máximo do nível de anonimização K suportado.

8.3 A arquitetura P2P

Além da escalabilidade referente à operação *consulta-K*, o sistema de anonimização proposto (baseado na arquitetura CAN) também responde bem à dinamicidade de usuários na rede P2P.

Conforme explanado na seção 4.3.2, a quantidade d de dimensões da arquitetura CAN define que cada líder deve armazenar informações sobre no mínimo $2 \cdot d$ outros nós, independemente da quantidade de usuários no sistema. Essa quantidade varia dinamicamente de acordo com as regiões que vão se formando, sendo a mesma para a abordagem proposta.

Em estruturas convencionais baseadas na arquitetura CAN, o comprimento médio do caminho de roteamento é $(d/4)(n^{1/d})$, para n usuários no sistema. Todavia, na arquitetura de anonimização proposta, mensagens são trocadas apenas entre líderes colaterais (inclusive os líderes vizinhos), pois as informações de coordenadas mais próximas (objeto da *consulta-K*) encontram-se estrategicamente localizadas nas regiões CAN vizinhas (sobre a curva de Hilbert) e a manutenção de informações de outras regiões CAN envolve apenas líderes colaterais. Desta forma, o comprimento máximo de roteamento l_{max}, em cada um dos dois sentidos de tráfego de mensagens do tipo *consulta-K*, é dado por $N_{max}/2$, isto é

$$l_{max} = \left\lceil \frac{K - \alpha}{2 \cdot \alpha} \right\rceil \tag{8.4}$$

Observa-se que l_{max} reflete a latência do sistema, haja vista que a operação *consulta-K* é encaminhada simultaneamente aos líderes vizinhos em ambos os sentidos. Admitindo-se uma simplificação que considera o tempo de resposta equivalente nesses dois sentidos, tem-se que o tempo total de consulta será equivalente ao tempo de consulta de apenas $1 + N_{max}/2$ nós líderes.

Os resultados do capítulo expressam o nível de escalabilidade do sistema *AnoniMobi*.

9 Conclusões e Trabalhos Futuros

"Viver em um mundo mental de erros e acertos absolutos pode levar a imaginar que todas as teorias são erradas. O que acontece na verdade é que uma vez que os cientistas tomam um bom coneito, eles o refinam gradualmente e o estendem com sutileza crescente à medida que seus instrumentos se aprimoram. As teorias não são erradas, são incompletas. Mesmo quando uma nova teoria parece representar uma revolução, ela geralmente surge de pequenos refinamentos."

Isaac Asimov, *Skeptival Inquirer*, v. 14, n. 1., 1989, p. 35–34.

9.1 Contribuições

A principal contribuição do presente trabalho é o desenvolvimento de uma proposta, apoiada em algoritmo de aglomeração, para o fornecimento de informações adicionais de orientação geográfica (rotas) durante consultas anônimas a SBLs públicos, sem prejuízo para o nível K de privacidade solicitado pelo usuário do sistema.

Os resultados obtidos através da avaliação conduzida, baseados em simulações, experimentos e análises, demonstram o potencial das ideias desenvolvidas no contexto de anonimização de consultas. Alguns pontos podem ser destacados:

Nível de privacidade. O nível K de privacidade das consultas efetuados, segundo a demanda do usuário, é garantido através da estratégia de anonimização adotada pelo sistema *AnoniMobi*. Independentemente da quantidade k de pontos apresentados ao SBL para a obtenção de um número equivalente de rotas, garante-se que o usuário consultante não será identificado dentre outros $K - 1$ usuários (com probabilidade de $1/K$).

Viabilidade de processamento em aparelhos portáteis. Os cálculos necessários para a construção da ASR baseada em pontos são extremamente simples (algoritmo de Lloyd), fato que não inviabiliza a implementação da estratégia de anonimização em aparelhos portáteis, cujas restrições de armazenamento, processamento e consumo de bateria devem ser cuidadosamente consideradas.

Tempo de consulta. O tempo para efetuar consultas por rotas utilizando o sistema *AnoniMobi* é uma função do nível de precisão [1] desejado. Quanto mais rotas solicitadas ao SBL, maior a probabilidade de diminuição da distância da rota mais próxima do usuário.

Escalabilidade. O nível de escalibilidade do sistema *AnoniMobi*, avaliado analiticamente, sugere que a arquitetura P2P proposta reage apropriadamente à dinamicidade característica de usuários móveis. Essa característica é imprescindível para um funcionamento satisfatório de qualquer arquitetura P2P móvel.

Semântica de consultas. A semântica de consultas a SBLs públicos comumente encontrada baseia-se em pontos (em vez de áreas). Como a abordagem *AnoniMobi* apoia-se nesta semântica básica, tal caracteristica permite a obtenção de rotas como retorno sem a necessidade de quaisquer adaptações ou conversões prévias no lado dos servidores.

Subsídios a ataques externos. Diferentemente das abordagens que utilizam re-

[1] A precisão da resposta obtida do SBL utilizando o *AnoniMobi* pode ser avaliada de acordo com a distância média do usuário até a rota mais próxima de sua localização dentre as rotas fornecidas.

tângulos como ASR (indicando os limites da região geográfica em que se encontram os usuários participantes do processo de anonimização), *AnoniMobi* baseia-se em pontos (centróides). Desta forma, inferências diretas sobre os possíveis usuários consultantes não são possíveis.

Discussão sobre ataques internos. Adicionalmente, o presente trabalho contribui para o estudo de anonimização de consultas em ambientes não-confiáveis ao expor a necessidade de tratamento dos ataques internos, propondo uma solução primária (baseada em camadas de navegação anônima) para evitar que o nível de anonimização possa ser totalmente anulado durante situações específicas, porém factíveis.

9.2 Trabalhos Futuros

A proposta da infraestrutura P2P subjacente que dá suporte ao sistema *AnoniMobi* foi concebida sobre os pilares teóricos da arquitetura CAN. Não tendo ainda sido devidamente simulada, toda avaliação desta infraestrutura baseou-se tão somente em desenvolvimentos analíticos. A simulação da arquitetura proposta é uma necessidade premente para uma validação prática das ideias apresentadas.

Para tanto, é necessária a aquisição de traços de mobilidade reais que sirvam como dados de entrada para uma simulação de efetiva utilização do sistema *AnoniMobi*. Com isso, não só o protocolo de comunicação (operações/mensagens entre usuários) poderá ser validado, como índices diversos poderão ser medidos (como a latência das comunicações, os valores ideais de α e β, a quantidade k de centróides mais adequada, etc).

Além dos traços de mobilidade, uma adequada plataforma de simulação deve ser desenvolvida, tendo em mente principalemnte as peculiaridades relativas à adaptação da arquitetura CAN, ao sistema sobreposto de georeferenciamento baseado na curva de Hilbert e ao protocolo de manutenção dos grupos de usuários.

Para a definição do nível K de anonimidade, a opção corriqueiramente adotada (e utilizada como exemplo neste trabalho) é a de permitir que o usuário faça sua escolha segundo algum critério pessoal, subjetivo. Tal decisão, em determinadas circunstânicias, pode não refletir as reais necessidades de privacidade durante uma consulta anônima. Em outras situações, o nível de privacidade escolhido pode ser demasiadamente exagerado, onerando o consumo de recursos para a execução das operações necessárias.

Um aspecto a ser considerado em trabalhos futuros é a inserção da possibilidade

de definição automática do nível K de anonimidade adequado para uma determinada consulta considerando-se o contexto em que a mesma esteja inserida, seja avaliando o tipo de consulta a ser efetuada, seja levando em conta a localização geográfica do usuário consultante, ou mesmo o seu perfil, por exemplo.

Como resultado final, pretende-se obter o desenvolvimento de um cliente *AnoniMobi* implementado na plataforma Android e executando em aparelhos compatíveis.

Referências

[Agranov e Gotsman 1995]Agranov, G.; Gotsman, C. Algorithms for Rendering Realistic Terrain Image Sequences and Their Parallel Implementation. *The Visual Computer*, v. 11, p. 455–464, 1995. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.24.9411&rep=rep1&type=pdf>.

[Alber e Niedermeier 1998]Alber, J.; Niedermeier, R. On Multi-dimensional Hilbert Indexings. In: *COCOON '98: Proceedings of the 4th Annual International Conference on Computing and Combinatorics*. London, UK: Springer-Verlag, 1998. p. 329–338. ISBN 3-540-64824-0. Disponível em: <http://theinf1.informatik.uni-jena.de/publications/cocoon98.pdf>.

[Alencar e Santos 2003]Alencar, H.; Santos, W. *Geometria Diferencial das Curvas Planas*. [S.l.]: Instituto de Matemática Pura e Aplicada, 2003.

[Aloise et al. 2009]Aloise, D. et al. Np-hardness of euclidean sum-of-squares clustering. *Mach. Learn.*, Kluwer Academic Publishers, Hingham, MA, USA, v. 75, n. 2, p. 245–248, 2009. ISSN 0885-6125.

[Anderson 2007]Anderson, P. *What is Web 2.0 Ideas, technologies and implications for education*. University of Bristol, 2007. Disponível em: <http://www.jisc.ac.uk/media/documents/techwatch/tsw0701b.pdf>.

[Androutsellis-Theotokis e Spinellis 2004]Androutsellis-Theotokis, S.; Spinellis, D. A Survey of Peer-to-Peer Content Distribution Technologies. *ACM Comput. Surv.*, ACM, New York, NY, USA, v. 36, n. 4, p. 335–371, 2004. ISSN 0360-0300. Disponível em: <http://www.dmst.aueb.gr/dds/pubs/jrnl/2004-ACMCS-p2p/html/AS04.pdf>.

[Arthur e Vassilvitskii 2006]Arthur, D.; Vassilvitskii, S. How Slow is the k-Means Method? In: *SCG '06: Proceedings of the twenty-second annual symposium on Computational geometry*. New York, NY, USA: ACM, 2006. p. 144–153. ISBN 1-59593-340-9. Disponível em: <http://www.cs.duke.edu/courses/spring07/cps296.2/papers/kMeans-socg.pdf>.

[Arthur e Vassilvitskii 2007]Arthur, D.; Vassilvitskii, S. k-means++: The Advantages of Careful Seeding. In: *SODA '07: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2007. p. 1027–1035. ISBN 978-0-898716-24-5.

[Barleze 2003]Barleze, A. *Fusão de Dados em Esquemas Híbridos Envolvendo AGPS para Localização de Posicionamento*. Dissertação (Mestrado) — Pontífica Universidade Católica do Paraná, 2003. Disponível em: <http://www.ppgia.pucpr.br/lib/exe/fetch.php?id=teses&cache=cache&media=dissertacoes:2005:alessandro_barleze-2003l.pdf>.

[Berry e Linoff 2000]Berry, M. J. A.; Linoff, G. *Mastering Data Mining: The Art of Science of Customer Relationship Management*. [S.l.]: John Wiley & Sons, Ltd, 2000.

[Berthold, Federrath e Kopsell 2001]Berthold, O.; Federrath, H.; Kopsell, S. Web MIXes: A System for Anonymous and Unobservable Internet Access. In: *International Workshop on Designing Privacy Enhancing Technologies*. New York, NY, USA: Springer-Verlag New York, Inc., 2001. p. 115–129. ISBN 3-540-41724-9. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.25.5586&rep=rep1&type=pdf>.

[Breinholt e Schierz 1998]Breinholt, G.; Schierz, C. Algorithm 781: Generating Hilbert's Space-Filling Curve by Recursion. *ACM Trans. Math. Softw.*, ACM, New York, NY, USA, v. 24, n. 2, p. 184–189, 1998. ISSN 0098-3500.

[Buford e Yu 2010]Buford, J. F.; Yu, H. Peer-to-Peer Networkins and Applications: Synopsis and Research Directions. In: Shen, X. et al. (Ed.). *Handbook of Peer-to-Peer Networking*. Springer, 2010. p. 3–46. Disponível em: <http://www.springer.com/engineering/signals/book/978-0-387-09750-3>.

[Buford, Heather e Lua 2009]Buford John F.; Heather Yu; Lua Eng K. *P2P: Networking and Applications*. [S.l.]: Morgan Kaufmann Publishers, 2009.

[Butz 1967]Butz, A. R. *Space Filling Curves and Mathematical Programming*. [S.l.], 1967. Disponível em: <http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=AD0663170>.

[Cantor 1883]Cantor, D. Üeber unedeliche, lineare punktmannigfaltigkeiten v. In: Klein Felix; Dyck Walther; Mayer Adolph (Ed.). *Mathematische Annalen*. Springer, 1883. v. 21, p. 545–591. Disponível em: <http://ia350605.us.archive.org/3/items/mathematischean01behngoog/mathematischean01behngoog.pdf>.

[Carlsson e Gustavsson 2001]Carlsson, B.; Gustavsson, R. The Rise and Fall of Napster - an Evolutionary Approach. In: *AMT '01: Proceedings of the 6th International Computer Science Conference on Active Media Technology*. London, UK: Springer-Verlag, 2001. p. 347–354. ISBN 3-540-43035-0. Disponível em: <http://www.bth.se/fou/forskinfo.nsf/6753b78eb2944e0ac1256608004f0535/ce24da4fd82b44a9c1256f790032f1f5/$file/The%20Rise%20and%20Fall%20of%20Napster.pdf>.

[Chaum 1981]Chaum, D. L. Untraceable Eletronic Mail, Return Address and Digital Pseudonyms. *Commun. ACM*, ACM, New York, NY, USA, v. 24, n. 2, p. 84–90, 1981. ISSN 0001-0782. Disponível em: <http://www.cs.utexas.edu/ shmat/courses/cs395t_fall04/chaum81.pdf>.

[Chen et al. 2010]Chen, C.-S. et al. A Closed-Form Algorithm for Converting Hilbert Space-Filling Curve Indices. *IAENG International Journal of Computaer Science*, v. 2, n. 1, 2010. Disponível em: <http://www.iaeng.org/IJCS/issues_v37/issue_1/IJCS_37_1_02.pdf>.

[Clarke et al. 2000]Clarke, I. et al. Freenet: A Distributed Anomymous Information Storage and Retrieval System. In: *Proceedings of the Workshop on Design Issues in Anonymity and Unobservability*. Berkeley, CA, USA: [s.n.], 2000. p. 46–66. Disponível em: <http://www.comp.nus.edu.sg/ cs6203/guidelines/topic6/clarke00freenet.pdf>.

[Comaniciu e Meer 1999]Comaniciu, D.; Meer, P. Mean Shift Analysis and Applications. In: *ICCV '99: Proceedings of the International Conference on Computer Vision-Volume 2*. Washington, DC, USA: IEEE Computer Society, 1999. p. 1197. ISBN 0-7695-0164-8. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=0BF8D5334462AC9006945EBFC960501D?doi=10.1.1.27.8250&rep=rep1&type=pdf>.

[Dabek et al. 2001]Dabek, F. et al. Wide-area cooperative storage with CFS. In: *Proceedings of the 18th ACM Symposium on Operating Systems Principles (SOSP '01)*. Chateau Lake Louise, Banff, Canada: [s.n.], 2001. Disponível em: <http://pdos.csail.mit.edu/papers/cfs:sosp01/cfs_sosp.pdf>.

[Dasgupta 2007]Dasgupta, S. *The Hardness of k-means Clustering*. [S.l.], 2007. Disponível em: <http://cseweb.ucsd.edu/Dienst/Repository/2.0/Body/ncstrl.ucsd_cse/CS2008-0916/postscript>.

[Dhara et al. 2010]Dhara, K. et al. Overview of Structured Peer-to-Peer Overlay Algorithms. In: Shen, X. et al. (Ed.). *Handbook of Peer-to-Peer Networking*. Springer, 2010. p. 233–256. Disponível em: <http://www.springer.com/engineering/signals/book/978-0-387-09750-3>.

[Dingledine, Freedman e Molnar 2001]Dingledine, R.; Freedman, M. J.; Molnar, D. The Free Haven Project: Distributed Anonymous Storage Service. In: *International workshop on Designing privacy enhancing technologies*. New York, NY, USA: Springer-Verlag New York, Inc., 2001. p. 67–95. ISBN 3-540-41724-9. Disponível em: <http://gnunet.org/papers/freehaven10.ps>.

[Elkan 2003]Elkan, C. Using the Triangle Inequality to Accelerate k-Means. In: *ICML '03: Proceedings of the Twentieth International Conference on Machine Learning*. Washington, DC, USA: Association for the Advancement of Artificial Intelligence, 2003. p. 147–153. Disponível em: <http://www.aaai.org/Papers/ICML/2003/ICML03-022.pdf>.

[Faloutsos e Roseman 1989]FALOUTSOS, C.; ROSEMAN, S. Fractals for secondary key retrieval. In: *PODS '89: Proceedings of the eighth ACM SIGACT-SIGMOD-SIGART symposium on Principles of database systems*. New York, NY, USA: ACM, 1989. p. 247–252. ISBN 0-89791-308-6. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.45.9043&rep=rep1&type=pdf>.

[FCC 2010]FCC Federal Communications Commission. *Wireless 911 Services*. 2010. Disponível em: <http://www.fcc.gov/cgb/consumerfacts/wireless911srvc.html>.

[Gan 2007]Gan, G. *Data Clustering, Theory, Algorithms and Applications*. [S.l.]: Society for Industrial and Applied Mathematics & American Statistical Association, 2007.

[Gedik e Liu 2005]Gedik, B.; Liu, L. Location privacy in mobile systems: A personalized anonymization model. In: *ICDCS '05: Proceedings of the 25th IEEE International Conference on Distributed Computing Systems*. Washington, DC, USA: IEEE Computer Society, 2005. p. 620–629. ISBN 0-7695-2331-5. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.126.2697&rep=rep1&type=pdf>.

[Geocaching 2010]Geocaching. *The Official Global GPS Cache Hunt Site*. 2010. Último acesso: 02/04/2010. Disponível em: <http://www.geocaching.com/>.

[Gersho e Gray 1991]GERSHO, A.; GRAY, R. M. *Vector quantization and signal compression*. Norwell, MA, USA: Kluwer Academic Publishers, 1991. ISBN 0-7923-9181-0.

[Ghinita, Kalnis e Skiadopoulos 2007]Ghinita, G.; Kalnis, P.; Skiadopoulos, S. PRIVÉ: Anonymous Location-Based Queries in Distributed Mobile Systems. In: *WWW '07: Proceedings of the 16th international conference on World Wide Web*. New York, NY, USA: ACM, 2007. p. 371–380. ISBN 978-1-59593-654-7. Disponível em: <http://dl.comp.nus.edu.sg/dspace/bitstream/1900.100/2243/1/trb7-06.pdf>.

[Ghinita, Skiadopoulos e Kalnis 2007]Ghinita, G.; Skiadopoulos, S.; Kalnis, P. 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*. Berlin, Heidelberg: Springer-Verlag, 2007. p. 221–238. ISBN 978-3-540-73539-7. Disponível em: <http://www.eng.auburn.edu/ weishinn/Comp7970/Papers/MobiHide.pdf>.

[Goldschlag, Reed e Syverson 1999]Goldschlag, D.; Reed, M.; Syverson, P. Onion Routing for Anonymous and Private Internet Connections. *Commun. ACM*, ACM, New York, NY, USA, v. 42, n. 2, p. 39–41, 1999. ISSN 0001-0782. Disponível em: <http://www.onion-router.net/Publications/CACM-1999.pdf>.

[Google Inc. 2010]Google Inc. *Android*. 2010. Último acesso: 03/08/2010. Disponível em: <http://www.android.com>.

[Gotsman e Lindenbaum 1996]Gotsman, C.; Lindenbaum, M. On the Metric Properties of Discrete Space-Filling Curves. In: *Transactions on Image Processing*. IEEE, 1996. p. 794–797. Disponível em: <http://www.cs.technion.ac.il/ gotsman/AmendedPubl/OnTheMetric/OnTheMetric.p df>.

[Gruteser e Grunwald 2003]GRUTESER, M.; GRUNWALD, D. Anonymous usage of location-based services through spatial and temporal cloaking. In: *MobiSys '03: Proceedings of the 1st international conference on Mobile systems, applications and services*. New York, NY, USA: ACM, 2003. p. 31–42. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.85.3772&rep=rep1&type=pdf>.

[Harrell et al. 2003]Harrell, F. et al. *Survey of Locating & Routing in Peer-to-Peer Systems*. [S.l.], 2003. Disponível em: <http://www.huaxiaspace.net/academic/classes/CSE221project.pdf>.

[Hazel e Wiley 2002]Hazel, S.; Wiley, B. Achord: a Variant of the Chord Lookup Service for Use in Censorship Resistant Peer-to-Peer Publishing Systems. In: *IPTPS '02: Proceedings of the 1st International Workshop on Peert-to-Peer Systems*. [s.n.], 2002. Disponível em: <http://thalassocracy.org/achord/achord-iptps.html>.

[Hilbert 1891]Hilbert, D. Üeber die stetige Abbildung einer Linie auf ein Flächenstück. In: Klein Felix; Dyck Walther; Mayer Adolph (Ed.). *Mathematische Annalen*. Springer, 1891. v. 38, p. 459–460. Disponível em: <http://ia361300.us.archive.org/21/items/mathematischean26behngoog/mathematischean26behngoog.pdf>.

[Hodges e Womtring 2000]Hodges, K.; Womtring, J. Client Typology Based on Functioning Across Domains using the CAFAS: Implications for Service Planning. *The Journal of Behavioral Health Services and Research*, Springer, New York, USA, v. 27, n. 3, p. 257–270, 2000. ISSN 1094-3412 (Print) 1556-3308 (Online). Disponível em: <http://www.springerlink.com/content/v3098m58h5505g68/>.

[Inaba, Katoh e Imai 1994]Inaba, M.; Katoh, N.; Imai, H. Applications of Weighted Voronoi Diagrams and Randomization to Variance-Based k-Clustering. In: *SCG '94: Proceedings of the tenth annual symposium on Computational geometry*. New York, NY, USA: ACM, 1994. p. 332–339. ISBN 0-89791-648-4. Disponível em: <http://www.cs.princeton.edu/courses/archive/spring04/cos598B/bib/InabaKI.pdf>.

[Jain e Dubes 1988]Jain, A. K.; Dubes, R. C. *Algorithms for Clustering Data*. [S.l.]: Prentice-Hall, Inc., 1988.

[Jin e Mellor-Crummey 2005]Jin, G.; Mellor-Crummey, J. Search and replication in unstructured peer-to-peer networks. In: *Proceeding of the Los Alamos Computer Science Institute Sixth Annual Symposium*. LACSI, 2005. Disponível em: <http://www.cs.rice.edu/ jin/papers/lacsi05.pdf>.

[Kalnis et al. 2006]Kalnis, P. et al. *Preserving Anonymity in Location Services*. [S.l.], 2006. Disponível em: <https://dl.comp.nus.edu.sg/dspace/bitstream/1900.100/2215/1/TRB6-06.pdf>.

[Kanungo et al. 2002]Kanungo, T. et al. An Efficient k-Means Clustering Algorithm: Analysis and Implementation. *IEEE Trans. Pattern Anal. Mach. Intell.*, IEEE Computer Society, Washington, DC, USA, v. 24, n. 7, p. 881–892, 2002. ISSN 0162-8828. Disponível em: <http://www.cs.umd.edu/ mount/Papers/pami02.pdf>.

[Kaufman e Rousseeuw 2005]Kaufman, L.; Rousseeuw, P. J. *Finding Groups Data: An Introduction to Cluster Analysis*. [S.l.]: John Wiley & Sons, Ltd, 2005.

[Kolodziej e Hjelm 2006]Kolodziej, K. W.; Hjelm, J. *Local Positioning Systems: LBS Applications and Services*. [S.l.]: Taylor & Francis, LLC, 2006.

[Kumar, Sabharwal e Sen 2004]KUMAR, A.; SABHARWAL, Y.; SEN, S. A Simple Linear Time (1+e)-Approximation Algorithm for k-Means Clustering in Any Dimensions. In: *FOCS '04: Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science*. Washington, DC, USA: IEEE Computer Society, 2004. p. 454–462. ISBN 0-7695-2228-9. Disponível em: <http://www.cse.iitd.ernet.in/ amitk/km.ps>.

[Küpper e Treu 2010]Küpper, A.; Treu, G. Next Generation Location-based Services: Merging Positioning and Web 2.0. In: YANG, L. T. et al. (Ed.). *Mobile Intelligence*. John Wiley & Sons, Ltd, 2010. p. 213–236. Disponível em: <http://www3.interscience.wiley.com/cgi-bin/bookhome/123299904>.

[Kütter 2005]Kütter, A. *Location-based Services: Fundamentals and Operations*. [S.l.]: John Wiley & Sons, Ltd, 2005.

[Lawder e King 2001]Lawder, J. K.; King, P. J. H. Querying Multi-dimensional Data Indexed Using the Hilbert Space-Filling Curve. *SIGMOD Rec.*, ACM, New York, NY, USA, v. 30, n. 1, p. 19–24, 2001. ISSN 0163-5808. Disponível em: <http://www.dcs.bbk.ac.uk/TriStarp/pubs/JL3_00.pdf>.

[Lloyd 1982]Lloyd, S. Least Squares Quantization in PCM. *Information Theory, IEEE Transactions on*, v. 28, n. 2, p. 129–137, 1982. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.131.1338&rep=rep1&type=pdf>.

[Lua et al. 2005]Lua, K. et al. A Survey and Comparison of Peer-to-Peer Overlay Network Schemes. *Communications Surveys & Tutorials, IEEE*, p. 72–93, 2005. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.109.6124&rep=rep1&type=pdf>.

[Lv et al. 2002]LV, Q. et al. Search and replication in unstructured peer-to-peer networks. In: *ICS '02: Proceedings of the 16th international conference on Supercomputing*. New York, NY, USA: ACM, 2002. p. 84–95. ISBN 1-58113-483-5. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.86.2764&rep=rep1&type=pdf>.

[MacKay 2005]MacKay, D. J. C. *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, 2005. Disponível em: <http://www.inference.phy.cam.ac.uk/itprnn/book.pdf>.

[Macqueen 1967]Macqueen, J. B. Some Methods of Classification and Analysis of Multivariate Observations. In: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. [s.n.], 1967. p. 281–297. Disponível em: <http://web.inf.unibz.it/dis/teaching/DWDM/project2009/K-means.pdf>.

[Mahajan, Nimbhorkar e Varadarajan 2009]Mahajan, M.; Nimbhorkar, P.; Varadarajan, K. The Planar k-Means Problem is NP-Hard. In: *WALCOM '09: Proceedings of the 3rd International Workshop on Algorithms and Computation*. Berlin, Heidelberg: Springer-Verlag, 2009. p. 274–285. ISBN 978-3-642-00201-4. Disponível em: <http://www.imsc.res.in/ meena/papers/kmeans.pdf>.

[Matias e Shamir 1988]Matias, Y.; Shamir, A. A Video Scrambling Technique Based On Space Filling Curves. In: *CRYPTO '87: A Conference on the Theory and Applications of Cryptographic Techniques on Advances in Cryptology*. London, UK: Springer-Verlag, 1988. p. 398–417. ISBN 3-540-18796-0. Disponível em: <http://dsns.csie.nctu.edu.tw/research/crypto/HTML/PDF/C87/398.PDF>.

[Mokbel, Chow e Aref 2006]Mokbel, M. F.; Chow, C.-Y.; Aref, W. G. The New Casper: Query Processing for Location Services without Compromising Privacy. In: *VLDB '06: Proceedings of the 32nd international conference on Very large data bases*. VLDB Endowment, 2006. p. 763–774. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=908C94A596BC594D2E13B82A95F4E95A?doi=10.1.1.96.5056&rep=rep1&type=pdf>.

[Moon et al. 2001]Moon, B. et al. Analysis of the Clustering Properties of the Hilbert Space-Filling Curve. *IEEE Trans. on Knowl. and Data Eng.*, IEEE Educational Activities Department, Piscataway, NJ, USA, v. 13, n. 1, p. 124–141, 2001. ISSN 1041-4347. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.129.1888&rep=rep1&type=pdf>.

[Nagel 2010]NAGEL, K. *Multi-Agent Microscopic Traffic Simulator*. 2010. Último acesso: 01/02/2010. Disponível em: <http://www.lst.inf.ethz.ch/research/ad-hoc/>.

[NavStar 2008]NavStar Global Position System. *GPS SPS Performance Standard, 4th ed.* [S.l.], 2008. Disponível em: <http://www.navcen.uscg.gov/gps/geninfo/2008SPSPerformanceStandardFINAL.pdf>.

[Nejdl et al. 2002]Nejdl, W. et al. EDUTELLA: a P2P Networking Infrastructure Based on RDF. In: *WWW '02: Proceedings of the 11th international conference on World Wide Web*. New York, NY, USA: ACM, 2002. p. 604–615. ISBN 1-58113-449-5. Disponível em: <http://cid.nada.kth.se/pdf/CID-205.pdf>.

[O'Sullivan 1995]O'Sullivan, B. *Applying Partial Evaluation to VLSI Design Rule Checking*. [S.l.], 1995. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.31.5141&rep=rep1&type=pdf>.

[Peano 1890]Peano, G. Sur une courbe, qui remplit toute une aire plane. In: Klein Felix; Dyck Walther; Mayer Adolph (Ed.). *Mathematische Annalen*. Springer, 1890. v. 36, p. 157–160. Disponível em: <http://ia341202.us.archive.org/2/items/mathematischean52behngoog/mathematischean52behngoog.pdf>.

[Peitgen, Jürgens e Saupe 2004]Peitgen, H.-O.; Jürgens, H.; Saupe, D. *Chaos and Fractals: New Frontiers of Science*. 2. ed. [S.l.]: Springer, 2004.

[Pfitzmann e Köhntopp 2001]Pfitzmann, A.; Köhntopp, M. Anonymity, Unobservability, and Pseudonymity - A Proposal for Terminology. In: . [s.n.], 2001. p. 1–9. Disponível em: <http://dud.inf.tu-dresden.de/literatur/Anon_Terminology_v0.31.pdf>.

[Ratnasamy et al. 2001]Ratnasamy, S. et al. A Scalable Content-Addressable Network. In: *SIGCOMM '01: Proceedings of the 2001 conference on Applications, technologies, architectures, and protocols for computer communications*. New York, NY, USA: ACM, 2001. p. 161–172. ISBN 1-58113-411-8. Disponível em: <http://lsirwww.epfl.ch/courses/dis/2003ws/papers/p13-ratnasamy.pdf>.

[Reichenbacher 2004]Reichenbacher, T. *Mobile Cartography: Adaptive Visualisation of Geographic Information on Mobile Devices*. Tese (Doutorado) — Technische Universität München, 2004. Disponível em: <http://tumb1.biblio.tu-muenchen.de/publ/diss/bv/2004/reichenbacher.pdf>.

[Reiter e Rubin 1998]Reiter, M. K.; Rubin, A. D. Crowds: Anonymity for Web Transactions. *ACM Trans. Inf. Syst. Secur.*, ACM, New York, NY, USA, v. 1, n. 1, p. 66–92, 1998. ISSN 1094-9224. Disponível em: <http://avirubin.com/crowds.pdf>.

[Ripeanu 2001]Ripeanu, M. *Peer-to-Peer Architecture Case Study: Gnutella Network*. 2001. Disponível em: <citeseer.ist.psu.edu/ripeanu01peertopeer.html>.

[Ripeanu e Foster 2002]Ripeanu, M.; Foster, I. T. Mapping the gnutella network: Macroscopic properties of large-scale peer-to-peer systems. In: *IPTPS '01: Revised Papers from the First International Workshop on Peer-to-Peer Systems*. London, UK: Springer-Verlag, 2002. p. 85–93. ISBN 3-540-44179-4. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.16.5045&rep=rep1&type=pdf>.

[Romesburg 2004]Romesburg, H. C. *Cluster Analysis for Researchers*. [S.l.]: Lulu Press, 2004.

[Sá e Bertocchi 2007]Sá, A.; Bertocchi, D. A Web 2.0 no ano de 2006. In: Pinto, M.; Fidalgo, J. (Ed.). *Anuário 2006 - A comunicação e os media em análise*. Universidade do Minho, 2007. p. 33–43. Disponível em: <http://193.137.91.100/ojs/index.php/anuario2006/article/viewFile/380/356>.

[Sagan 1994]Sagan, H. *Space-Filling Curves*. [S.l.]: Springer, 1994.

[Salomon 2004]Salomon, D. *Data Compression: The Complete Reference*. 3. ed. [S.l.]: Springer, 2004.

[Samarati 2001]Samarati, P. Protecting respondents' identities in microdata release. *IEEE Trans. on Knowl. and Data Eng.*, IEEE Educational Activities Department, Piscataway, NJ, USA, v. 13, n. 6, p. 1010–1027, 2001. ISSN 1041-4347. Disponível em: <http://spdp.dti.unimi.it/papers/tkde_k-anonymity.pdf>.

[Samarati e Sweeney 1998]Samarati, P.; Sweeney, L. *Protecting Privacy when Disclosing Information: k-Anonymity and its Enforcement through Generalization and Suppression*. [S.l.], 1998. Disponível em: <http://www.csl.sri.com/papers/sritr-98-04/>.

[Sandvine 2009]Sandvine. *2009 Global Broadband Phenomena Report*. [S.l.], 2009. Disponível em: <http://www.sandvine.com/downloads/documents/2009%20Global%20Broadband%20Phenomena%20-%20Full%20Report.pdf>.

[Sandvine 2010]Sandvine. *2010 Mobile Internet Phenomena Report*. [S.l.], 2010. Disponível em: <http://www.sandvine.com/downloads/documents/2010%20Mobile%20Internet%20Phenomena%20Report.pdf>.

[Santos et al. 2008]SANTOS, R. de O. et al. Joinus: Management of mobile social networks for pervasive collaboration. In: *SBSC*. IEEE Computer Society, 2008. p. 224–234. ISBN 978-0-7695-3500-5. Disponível em: <http://dblp.uni-trier.de/db/conf/sbsc/sbsc2008.html>.

[Schoenberg 1938]Schoenberg, I. J. The Peano-Curve of Lebesgue. *Bulletin of American Math Society*, v. 44, n. 1, p. 519, 1938.

[SecondLife 2010]SECONDLIFE. *Second Life*. 2010. Último acesso: 13/05/2010. Disponível em: <http://secondlife.com>.

[Shin, Jung e Balakrishnan 2006]Shin, S.; Jung, J.; Balakrishnan, H. Malware prevalence in the KaZaA file-sharing network. In: *IMC '06: Proceedings of the 6th ACM SIGCOMM conference on Internet measurement*. New York, NY, USA: ACM, 2006. p. 333–338. ISBN 1-59593-561-4. Disponível em: <http://www.imconf.net/imc-2006/papers/p34-shin.pdf>.

[Steiniger, Neun e Edwardes 2006]Steiniger, S.; Neun, M.; Edwardes, A. *Foundations of Location Based Services*. Department of Geography, 2006. Lesson 1 CartouCHe 1- Lecture Notes on LBS, V. 1.0. Disponível em: <http://www.geo.unizh.ch/publications/cartouche/lbs_lecturenotes_steinigeretal2006.pdf>.

[Stoica et al. 2003]Stoica, I. et al. Chord: a Scalable Peer-to-Peer Lookup protocol for Internet Applications. *IEEE/ACM Trans. Netw.*, IEEE Press, Piscataway, NJ, USA, v. 11, n. 1, p. 17–32, 2003. ISSN 1063-6692. Disponível em: <http://pdos.csail.mit.edu/papers/chord:sigcomm01/chord_sigcomm.pdf>.

[Sweeney 2002]Sweeney, L. k-anonymity: a model for protecting privacy. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.*, World Scientific Publishing Co., Inc., River Edge, NJ, USA, v. 10, n. 5, p. 557–570, 2002. ISSN 0218-4885. Disponível em: <http://epic.org/privacy/reidentification/Sweeney_Article.pdf>.

[Vieira, Martinello e Marcondes 2009]Vieira, R. D. N.; Martinello, M.; Marcondes, C. A. C. Privacidade de Localização em Serviços Móveis: Anonimidade-k baseada em Triângulo Pontualizado. In: *SBCUP 2009: Anais do Simpósio Brasileiro de Computação Ubíqua e Pervasiva*. Porto Alegre, RS, Brasil: Sociedade Brasileira de Computação, 2009. p. 1163–1172. ISSN 2175-2761. Disponível em: <http://www.sbc.org.br/bibliotecadigital/download.php?paper=1388>.

[Vieira et al. 2010]Vieira, R. D. N. et al. Privacidade de Localização: Uma abordagem baseada em médias aleatórias. In: *SBCUP 2010: Anais do II Simpósio Brasileiro de Computação Ubíqua e Pervasiva*. Porto Alegre, RS, Brasil: Sociedade Brasileira de Computação, 2010. p. 296–305. ISSN 2175-2761.

[Voorhies 1994]Voorhies, D. Space-Filling Curves and a Measure of Coherence. In: ARVO, J. (Ed.). *Graphcis Gems II*. AP Professional, 1994. p. 26–30. Disponível em: <http://tog.acm.org/resources/GraphicsGems/>.

[Walther e Fischer 2002]Walther, U.; Fischer, S. Metropolitan area mobile services to support virtual groups. *IEEE Transactions on Mobile Computing*, I, n. 2, p. 96–110, 2002. Disponível em: <http://www.computer.org/portal/web/csdl/doi/10.1109/TMC.2002.1038346>.

[Wikipédia 2010]Wikipédia. *Web 2.0*. 2010. Último acesso: 02/04/2010. Disponível em: <http://pt.wikipedia.org/wiki/Web_2.0>.

[Xu 2002]Xu, J. On the Fundamental Tradeoffs between Routing Table Size and Network Diameter in Peer-to-Peer Networks. *College of Computing Technical Reports*, Georgia Institute of Technology, 2002. Disponível em: <http://hdl.handle.net/1853/6547>.

[Yeung e Ruzzo 2000]Yeung, K. Y.; Ruzzo, W. L. *An Empirical Study on Principal Component Analysis for Clustering Gene Expression Data*. [S.l.], 2000. Disponível em: <http://faculty.washington.edu/kayee/pca/pca.pdf>.

[Zhao, Kubiatowicz e Joseph 2001]Zhao, B. Y.; Kubiatowicz, J. D.; Joseph, A. D. *Tapestry: An Infrastructure for Fault-tolerant Wide-area Location and Routing*. Berkeley, CA, USA, 2001. Disponível em: <http://www.comp.nus.edu.sg/ cs6203/guidelines/topic1/tapestry.pdf>.

Livros Grátis

(http://www.livrosgratis.com.br)

Milhares de Livros para Download:

Baixar livros de Literatura
Baixar livros de Literatura de Cordel
Baixar livros de Literatura Infantil
Baixar livros de Matemática
Baixar livros de Medicina
Baixar livros de Medicina Veterinária
Baixar livros de Meio Ambiente
Baixar livros de Meteorologia
Baixar Monografias e TCC
Baixar livros Multidisciplinar
Baixar livros de Música
Baixar livros de Psicologia
Baixar livros de Química
Baixar livros de Saúde Coletiva
Baixar livros de Serviço Social
Baixar livros de Sociologia
Baixar livros de Teologia
Baixar livros de Trabalho
Baixar livros de Turismo