وزارت ارتباطات و فناوری اطلاعات
شرکت ارتباطات زیرساخت

عنوان مقاله :ایجاد یک روش کارا برای حفظ حریم خصوصی در پایگاه داده‌های مکان محور

تهیه کننده/گان:	مدرک و رشته تحصیلی	رشته شغلی	اداره کل/دفتر
گیتی قهرمانی	فوق لیسانس فناوری اطلاعات	کارشناش شبکه	کنترل,هماهنگی و پایش
حسین عبدالهی	لیسانس فیزیک	فنی و مهندسی	کنترل,هماهنگی و پایش

عنوان حوزه تحقیقاتی مورد نیاز شرکت :اینترنت اشیاء (IOT) شامل معماری,پروتکل ها, کاربرد, چالش های امنیتی و سایر موارد مطروحه

شماره ردیف حوزه تحقیقاتی مورد نیاز شرکت ۸۸

چکیده

سرویسهای مبتنی بر موقعیت مکانی (LBS) با توسعه سریع فناوری اینترنت اشیا به حوزه‌ای مهم برای تحقیقات تبدیل‌شده است. اگرچه کاربران می‌توانند از طریق LBS در شبکه‌های اینترنت اشیا از راحتی زیادی برخوردار شوند, اما ممکن است حریم خصوصی را از دست بدهند. سرورهای LBS مخرب با اطلاعاتی که از کاربران در اختیاردارند، می‌توانند کاربران را ردیابی کنند یا داده‌های شخصی را به اشخاص ثالث انتقال دهند. درابتدا الگوریتم DLS که یک روش حفظ حریم خصوصی مکان است را تحلیل می‌کنیم. برای حفظ کارایی حریم خصوصی کاربر، با در نظر گرفتن هزینه‌های محاسباتی و نیازهای حفظ حریم خصوصی کاربران ، الگوریتم DLP را معرفی می‌نماییم. این روش یک روش کارآمد به شمار می‌آید اما در این روش مکان واقعی کاربر در محاسبات وارد نمی شود. در اینجا روشی ارائه نموده ایم که از مزایای روش DLP بهره می‌برد و قیود هندسی در محاسبات دخیل می‌شود. نتایج نشان می‌دهد که الگوریتم DLP ارتقا یافته نسبت به الگوریتم DLS مکان کاربر را کمتر آشکار می کند و کارایی بهتری دارد , در عین حال همان میزان حریم خصوصی را با الگوریتم DLP حفظ می‌کند . شاخص CR در الگوریتم پیشنهادی ما نسبت به سایر الگوریتم‌ها کیفیت بالاتری از خود نشان داده است.

کلید واژه- اینترنت اشیا، سرویس‌های مبتنی بر موقعیت ، روش DLS، روش DLP

۱- مقدمه

اینترنت شبکه ی جهانی است که ارتباط تمامی کاربران را با هم برقرار می کند، اما ساختار این شبکه در حال تغییر است .همه ی ما از لپ تاپ ها، تبلت ها و گوشی های هوشمند جهت برقراری ارتباط بین دوستان خود استفاده می کنی . اغلو اطلاعات بین ما و دوستان از طریق سرورهایی که وظیفه ی اجرای سایت را برعهده دارند و نرم افزارهای ایمـی رد و بـدل مـی شـود . درنهایت می توان گفت اینترنت از کاربران، دستگاه های سمت کلاینت و سرورها تشکی شده است؛ اما عضو جدیدی در حال اضافه شدن به این مجموعه است .این عضو جدید کاربر نیست و از آن به عنوان اشیا یاد می شود، این کلمه از عبارت اینترنت اشیا برگرفته شده است. ممکن است به هر دستگاهی که دارای سنسوری جهت تبادل اطلاعات است خطاب شود « .شـی » یک مثال هایی از این سنسورها شام سنسور دما، سنسورهای ترافیک و سنسورهای اندازه گیری انرژی است. به واضح می توان گفت که اینترنت اشیا تاثیر بسیاری را در زندگی همه ی کاربران خواهد داشت . در حالت کلی می توان گفت اینترنـت اشـیا شبکه ای از اشیاء فیزیکی تعبیه شده با قطعات الکترونیکی، نرم افزار، سنسورها و اتصالات است تا آن ها توسط تبادل اطلاعات با تولید کننده، اپراتور و یا دستگاه های دیگر قادر به ارائه ارزش و خدمات بیشتر باشند (ازبک زایی، ۱۳۹۵) یکی از مهمترین سرویس ها در شبکه اینترنت اشیا سرویس های مبتنی بر مکان می باشند. این سرویس ها سرویس هایی هستند کـه در هـر زمان و مکان به کاربران خود سرویس هایی را بر اساس مکان قرارگیری آنها ارائه می کنند, و با توجه به گسـترش روز افـزون دستگاه هایی که به صورت سیار و در بستر ارائه کنندگان سرویس های موبایل به تبادل اطلاعات می پردازند, استفاده از این سیستم ها مورد بررسی قرار گیرد و برای هر قسمت از آن راهکارهایی ارائه شود, یکی از این جوانب که از اهمیت حیـاتی بـای

کاربران برخوردار است, حفظ حریم خصوصی افراد در این محیط ها می باشد زیرا کاربرانی که از این سرویس ها استفاده مـی کنند باید درخواست خود را در قالب یک پرس و جو به ارائه کنندگان سرویس هـا بفرسـتند, ایـن پـرس و جوهـا عـلاوه بـر درخواست کاربر شامل هویت (یک شماره منحصر به فرد مانند شماره تلفن کاربر) و مکان کاربر نیز می باشد. حال فرض کنیم یک فرد مهم در این میان طعمه افراد بدخواه باشد, واضح است که این افراد یکسـری اطلاعـات قبلـی از ایـن فـرد و زنـدگی خصوصی او دارند, حال اگر این فرد از این سرویس ها استفاده کند و در هر مکانی که قرار بگیرد, پرس و جوهایی را برای ارائه کنندگان سرویس های مبتنی بر مکان ارسال می کند, با افشای این اطلاعات توسط افراد بدخواه بسیاری از عادت های روزمره این فرد مانند مسیرهای رفت و آمد, مکان هایی که در آنجا رفت و آمد دارد, مراسم هایی که شرکت مـی کنـد و بسـیاری از عادت های دیگر او فاش خواهند شد و خطرات بسیاری را برای او از طرف این افـراد بـه دنبـال خواهـد داشـت. پـرداختن بـه مشکلات امنیتی اینترنت اشیاء یکی از چالش‌های اصلی محققان در چند سال اخیر بوده است. در این مقاله نیز هدف ما یافتن راهکارهای ارتقای امنیت کاربران در حوزۀ اینترنت اشیاء است.(درگلاله,۱۳۹۲)

۲-سوابق تحقیق

اینترنت اشیاء را می توان شبکه ای از اشیای فیزیکی تعبیه شده با قطعات الکترونیکی, نرم افزار, سنسور و اتصالات دانست که با تبادل اطلاعات مابین تولید کننده, اپراتور و یا دستگاه های دیگر فناوری های بی سیم و سامانه های میکروالکترونیک است.

پیشرفت های اینترنت به این معنی است که محیط, شهر, ساختمان, وسایل نقلیه, لباس, دستگاه های قابل حمل و... بیشتر و بیشتر در ارتباط با یکدیگر و تولید اطلاعات جدید باشند. علاوه بر این, فناوری های شبکه باید با چالش های جدید کنار بیایند, مثلا نرخ داده های بسیار بالا, جمعیت انبوه کاربران, تاخیر, انرژی و هزینه کم و تعداد زیاد دستگاه ها. با پیشرفت های فناوری ارتباطات همراه و استفاده گسترده از ابزار های همراه مجهز به GPS خدمات مبتنی بر مکان به سرعت در حال رشد هستند. خدمات مبتنی بر مکان برای ارائه سرویس به کاربران خود از سرورهای مبتنی بر مکان استفاده می کنند.

سرویس LBS یک سرویس در سطح نرم افزار است که از داده های مربوط به مکان جغرافیایی برای کنترل برخی ویژگی های نرم افزاری استفاده می کند. در واقع , LBS یک سرویس اطلاعاتی است که امروزه کاربردهای مختلفی به عنوان اطلاعات بدست آمده از موقعیت جغرافیایی- چه برای مصارف سرگرمی و چه برای مصارف امنیتی- در شبکه های اجتماعی مورد استفاده در گوشی های موبایل و از طریق شبکه بی سیم دارد. این تعاریف LBS را به عنوان محل تقاطع سه تکنولوژی توصیف می کند این تکنولوژی ارتباطی و اطلاعات جدید (NICTs: New Information and Communication Technologies) شامل سیستم های ارتباطات موبایل و ابزار های موبایل, تکنولوژی اینترنت و سیستم های اطلاعات مکانی (GIS) با پایگاه داده های مکانی هستند.(سبکبار,۱۳۹۱)

وجود انبوه اطلاعات در پایگاه داده های مکان مبنا از مهمترین دلایل علاقه مندی مهاجمین به سرقت اطلاعات می باشد. یک متخاصم می تواند با دستیابی به مکان کاربر امنیت او را تهدید کند. به‌منظور افزایش امنیت حریم خصوصی Data Base های مکان‌محور روش‌های زیادی مانند کدگذاری, ناشناس‌سازی و... وجود دارد.

۲-۱- کارهای صورت‌گرفته

روش‌های ناشناس‌کردن مکان یکی از مهم‌ترین تکنیک‌ها برای حفظ حریم خصوصی است که تلاش دارد مکان کاربر را برای کاربران دیگر، غیرقابل‌شناسایی و ردیابی کند. روش K-anonymity یکی از مهم‌ترین آن‌ها در این زمینه است. روش K-anonymity شخصی‌سازی‌شده برای اولین بار توسط (Gedik-۲۰۰۸) ارائه شد. این روش کاربر را قادر می‌سازد تا بر اساس نیازش در سطح خاصی از حریم خصوصی قرار گیرد.

ناشناس‌سازی می‌تواند توسط سرور پنهان‌ساز [1] یا توسط خود کاربری صورت گیرد که درخواست خود را برای سرور مکان‌محور ارسال می‌کند. در روش اول، سرور قابل اعتماد پنهان‌ساز، یک موتور مؤثر پیام‌رسان را فعال می‌کند. در هر دو این روش‌ها با ایجاد یک توازن میان حریم خصوصی و کیفیت سرویس QoS [2] کار مخفی‌سازی کاربر انجام می‌شود.

در (Beresford-۲۰۰۳) و (Jiang-۲۰۰۷) تعداد زیادی از روش‌های مختلفی که در دو دهۀ اخیر در این زمینه ارائه شده‌اند، آورده شده است. بیشتر این روش‌ها بر پایۀ آشوب موقعیت [3] و روش K-anonymity و سرور پنهان‌سازند. از آنجایی که این روش وابستگی زیادی به پنهان‌ساز موقعیت LA [4] دارد، یک ضعف اساسی برای این روش به شمار می‌آید زیرا اگر این سرور دچار مشکل شود این روش قابل استفاده نخواهد بود. همچنین، به‌دلیل اینکه LA نیاز به پردازش بر روی همۀ داده‌ها از تمامی کاربرها دارد، مشکل حجم محاسباتی نیز یکی دیگر از مشکلات این روش به شمار می‌آید(Gang, Sun-۲۰۱۷).

یکی از مهم‌ترین مباحث در زمینۀ امنیت، یافتن معیاری برای سنجش میزان ناشناختگی است. ما در این مقاله قصد داریم از معتبرترین معیار موجود که همان معیار آنتروپی است استفاده کنیم. برای مثال، در (Meyerowitz-۲۰۰۹) یک روش حفظ حریم خصوصی افراد با نام CachCloak ارائه شده است که از معیار آنتروپی استفاده کرده است. یا در (Niu-۲۰۱۳) دو روش K-anonymity مبنی بر مکان هندسی ارائه شده است که معیار اصلی سنجش کیفیت الگوریتم مطرح شده، آنتروپی است. همچنین، این معیار در مراجع (Niu-۲۰۱۳)و(Zhu-۲۰۱۳)و (Niu-۲۰۱٤) نیز مورد استفاده قرار گرفته است.

در مقالاتی که در این حوزه به چاپ رسیده‌اند شاهد ترکیب روش‌های گوناگونی با روش K-anonymity هستیم. مثلاً در (Niu-۲۰۱۳) طرحی به نام ۳plus مطرح شده است. به این صورت که از یک سرور ثبت هویت استفاده کرده است, هر کاربر هویت خود را در این سرور ثبت می‌کند و در زمان ارسال مکان به سرور LBS , ابتدا تعیین هویت می‌شود و سپس اجازۀ ارسال مکان جعلی را به LBS دارد. یا در (Aryan-۲۰۱۰) از سروری به نام Trusted Third Party یا شخص سوم قابل اعتماد استفاده می‌کند که از ترکیب یکسری اطلاعات کاربر نام مستعار ایجاد کند.

با وجود ترکیب روش‌های گوناگون با K-anonymity , بیشترین کار بر روی چگونگی انتخاب نقاط "Dummy Location" صورت گرفته است. در(Niu-۲۰۱۳)عنوان شده است که کاربر مکان‌های جعلی خود را از حافظۀ بافر خود و مکان‌هایی که در گذشته در آن نقاط بوده است انتخاب می‌کند. یکی از مشکلاتی که این روش با آن مواجه بود وجود مکان‌های یکسان در حافظۀ بافر آن کاربر است. مثلاً نقاطی که یک پزشک در آن‌ها در طی چند ساعت حضور داشته است؛ نقاطی مانند کلینیک, بیمارستان, داروخانه و... است که این مسئله تعیین مکان کاربر را برای مهاجم آسان می‌کند. در (Niu-۲۰۱۴) مطرح می‌شود که این نقاط بهتر است دور از هم انتخاب شوند ولی احتمال قرارگیری این نقاط در مناطقی که احتمال وجود کاربر در آن‌ها بسیار پایین است (مانند مرداب, دره و...) وجود دارد. یا مثلاً می‌توان به ترکیب قیود هندسی با K-anonymity در مقالۀ اشاره کرد که به‌منظور رفع مشکل وجود نقاط جعلی در مناطقی مانند دریاچه, کوه و... مطرح شده بود, به این صورت که نقاط جعلی بر روی سطوح یک کره, هرم و... قرار بگیرند. اما مشکلی که این مقاله ایجاد کرده است وجود نقاط جعلی‌ای است که از نظم خاصی پیروی می‌کنند که از طرف مهاجم قابل شناسایی است. یا در (Zhu-۲۰۱۳)عنوان شده است که این نقاط نزدیک همدیگر انتخاب شوند. یا در (Ying-۲۰۱۴) یک روش K-anonymity با درجۀ ناشناختگی k را مطرح شده است. در این روش، علاوه‌بر استفاده از k ناشناس از تنوع l نیز استفاده می‌شود. زمانی که این روش اجرا می‌گردد، به کاربر این تضمین را می‌دهد که میزان ناشناختگی k با تنوع L می‌تواند امنیت حریم خصوصی را تا حد زیادی تأمین کند.

مراجع (Vu-۲۰۱۲)و(Liu-۲۰۱۳) سعی دارد تا به‌صورت ساختگی، مکان‌هایی را برای کاربران تولید کند که برای هکر غیرقابل‌ردیابی باشد. در بسیاری از این روش‌ها مانند DLS [5] اطلاعات جانبی کاربران مانند زمان درخواست، نوع درخواست، جنسیت درخواست‌کننده و... درنظر گرفته نمی‌شود. این اطلاعات جانبی به‌نوعی به مهاجم کمک می‌کند تا به مکان کاربر

---

۱ Annonymizer
۲ Quality of Service
۳ Location Perturbation
٤ Location annonymizer
٥ Dummy Location Selection

دسترسی پیدا کنند. در (Niu-۲۰۱۴) روشی مطرح شده که در آن الگوریتم DLS با درنظرگرفتن اطلاعات جانبی کاربران عمل می‌کند. معیار آنتروپی محاسبه در این روش به‌مراتب بالاتر از روش‌های مطرح‌شدهٔ قبلی است. اما مشکل این روش بالابودن هزینه‌های محاسباتی و زمان محاسباتی است. مرجع (Sun-۲۰۱۷) یکی از جدیدترین روش‌های مطرح‌شده در این زمینه را ارائه می‌کند که به‌نوعی روش پایه‌ای کار ما در این رساله است. روش مطرح‌شده در این مرجع روش DLP[۶] است. این الگوریتم ازجمله روش‌هایی است که نقاط جعلی توسط خود کاربر تعیین می‌شود؛ به این معنا که از سرورهای anonymizer به این منظور استفاده نمی‌کند. روند کار به این صورت است که کاربر که نیاز به استفاده از سرور LBS دارد (مثلاً دریافت اطلاعاتی درخصوص رستوران‌های اطراف مکان اقامت) یک درخواست مبنی بر ارسال K مکان (که در ادامه توضیح داده خواهد شد) که احتمال درخواستی مشابه مکان کاربر دارند به سرور LBS ارسال می‌کند. بعد از دریافت این K مکان به روش‌هایی که در ادامه درخصوص آن‌ها توضیح داده خواهد شد, K-۱ مکان انتخاب کرده و به‌همراه مکان خود به سرور LBS ارسال می‌کند. دراین‌صورت عملاً سرور LBS, K مکان برای این کاربر متصور می‌شود و رستوران‌هایی که به این K مکان نزدیک است را به کاربر معرفی می‌کند. همچنین درخصوص مزایای DLP می‌توان به بهینه‌سازی حجم محاسبات نیز اشاره کرد.

در مقاله (Hossain-۲۰۱۲) یک روش حفظ حریم خصوصی با نام معروف ClusterCloak ارائه گردید. در این روش یک کاربر تلفن همراه به کمک خوشه‌بندی می‌تواند به سطح مطلوبی از ناشناختگی دست یابد. مفهوم ترکیب مناطق توسط (Liu-۲۰۱۲) برای اولین بار ارائه گردید. این مفهوم بیانگر مناطق جغرافیایی است که کاربر یک اپلیکیشن درخواستی دارد. در این روش یک روش گراف وزنی مبتنی بر K-anonymity ارائه شده که نه‌تنها می‌تواند حریم خصوصی فرد را تضمین کند بلکه پهنای باند را نیز کاهش می‌دهد. سپس این مفهوم توسط (Jung-۲۰۱۳) به کاربران اجازهٔ جابه‌جایی هویت نامشخص خود با سایر کاربران در منطقهٔ ترکیبی را می‌دهد. این عمل تضمین می‌کند که کاربران نیاز نیست تا یک ناشناختگی زیاد داشته باشند. بنابراین رابطهٔ میان ناشناختگی و مکان جغرافیایی با جابه‌جایی مکان کاربران قابل مسامحه است.

در (Zhang-۲۰۱۹) یکی از جدیدترین روش‌های مطرح‌شده در سال ۲۰۱۹ در این حوزه معرفی شد که در آن نویسندگان مقاله، ابتدا مکان تصادفی کاربر توسط سرور قابل اطمینان anonymizor ایجاد می‌شود و سپس این مکان‌ها برای استفادهٔ دوبارهٔ کاربر ذخیره می‌شوند. مقالاتی قبل از این مقاله که بر روی ذخیرهٔ نقاط جعلی کار کرده بودند از حافظهٔ دو سطحی استفاده شده بوده است که این احتمال وجود داشته که با جابه‌جایی کاربر این نقاط دیگر قابل استفاده نباشند. در این مقاله سعی شده است با استفاده از زنجیرهٔ مارکوف مکان بعدی کاربر حدس زده شود و همچنین از حافظهٔ چندسطحی استفاده شود. از مشکلات این روش به حجم محاسباتی بالا و همچنین استفاده از سرور anonymizer می‌توان اشاره کرد.

در (Zhang-۲۰۱۸) یک روش حفظ حریم خصوصی دوگانه که برای محافظت از مسیر کاربر و اطلاعات مربوط به کاربر است ارائه شده است. در این مقاله، کاربر می‌تواند ناشناس‌ساز مختلف برای پنهان‌ساز موقعیت خود انتخاب کند. همچنین در (Zhao-۲۰۱۸) روشی ارائه شده که در آن به مهاجم اجازهٔ دسترسی به LBS را نیز نمی‌دهد. این امر موجب شده تا میزان حریم خصوصی به‌طرز قابل‌ملاحظه‌ای افزایش یابد. زیرا مهاجم می‌تواند با استفاده از بعضی برنامه‌ها بر روی تلفن همراه کاربر چند نقطهٔ جعلی ایجاد کند و زمانی که کاربر مکان خود و مکان‌های جعلی را برای سرور LBS ارسال می‌کند مهاجم با حذف نقاط جعلی که خود ایجاد کرده و همچنین محاسبهٔ سرعت حرکت کاربر (مثلاً در جاده) محدودهٔ کاربر را شناسایی کند. دراین‌صورت یافتن مکان واقعی کاربر از بین دو نقطه راحت‌تر از انتخاب از بین مثلاً چهار نقطه است. مشکل این مقاله استفاده از سرور anonymizer است.

در یکی از جدیدترین مقاله‌های ارائه شده در سال ۲۰۱۹، روشی مشابه روش DLP ارائه شده که (Du-۲۰۱۹) قصد ارتقای این روش را دارد. در این مقاله، در مرحلهٔ ابتدایی روش DLP که درجهٔ آزادی برای انتخاب مکان وجود دارد و نقاط به‌صورت حریصانه انتخاب می‌شوند، روشی ارائه شده است که این انتخاب حریصانه را ارتقا می‌دهد. در این طرح پیشنهادی برای

---

۶ Dummy location privacy-preserving

الگوریتم حریصانه ارتقا یافته، نیاز به جستجوی تمام فضای جواب برای یافتن مکان‌هایی با بیشترین شباهت احتمال درخواست با کاربر واقعی وجود ندارد. از طرفی ضعف این روش در این است که تنها معیار آنتروپی را برای یافتن این نقاط درنظر گرفته است.

در تحقیق دیگری در(Fei-۲۰۱۷) در مواجهه با چالش بیشینه‌سازی آنتروپی، یک طرح تقسیم هزینهٔ جدید پیشنهاد شده است. در این مقاله برای به‌دست‌آوردن حداکثر حریم خصوصی کاربران، ابتدا منطقه را بر اساس احتمال درخواست به گروه‌های کوچک تقسیم می‌کند و به جای اینکه anonymizer برای تک‌تک اعضای این گروه‌ها کار گمنام‌سازی را انجام دهد تنها برای یک کاربر در هر گروه نقاط جعلی را تعیین می‌کند. اما مشکلی که این مقاله با آن مواجه است مصرف انرژی برای انتخاب سرگروه است که این مشکل را با بیان الگوریتمی که در آن تعداد کاربران در هر منطقه به طریقی مشخص شود که آنتروپی هر منطقه بیشترین حد ممکن و انرژی به کمترین حد ممکن برسد حل کرده است. البته مشکل اصلی این مقاله استفاده از سرور anonymizer است.

در (El Ouazzani-۲۰۱۸)، یک روش گمنام‌سازی ارائه شده است که به خصوصیات شبه‌شناسایی می‌پردازد. علاوه‌براین، در این تحقیق برخی از کارها در زمینهٔ روش ناشناسی k ارائه شده است. در این مقاله نیز از روش K-anonymity استفاده می‌کند و k گمنامی توسط کاربر مشخص می‌شود. در همهٔ مقالاتی که تا اینجا به آن پرداخته شد k یک عدد ثابت به حساب می‌آمد که به مقدار آن نقاط جعلی برای کاربر مشخص می‌شود. این مقاله به تعداد k مکان جعلی ایجاد می‌کند با این تفاوت که هر مکان, محدوده‌ای از نقاط را شامل می‌شود.

در سال ۲۰۱۸، (Jiang-۲۰۱۸) یک الگوریتم حفظ حریم خصوصی مکان به نام RobLoP پیشنهاد کرد. این مقاله برخلاف کارهایی که تا اینجا به آن پرداختیم, یعنی discrete location به continuous location می‌پردازد و بیان می‌کند که مکان کاربر با توجه به ارتباطش با نودهای کناری و سرعت حرکت مشترک و زمان حضور در هر مکان توسط مهاجم حدس زده شود. این مقاله الگوریتمی را برای جلوگیری از شناسایی مکان کاربر مطرح می‌کند.

درنهایت می‌توان گفت روش‌هایی که ما در این تحقیق بررسی کرده‌ایم عمدتاً سعی دارند تا احتمال دسترسی مهاجم به مکان کاربر را تا جای ممکن کاهش دهند. هرچند برخی از این روش‌ها به اطلاعات جانبی یا حجم محاسباتی روش مطرح‌شده توجه ویژه‌ای نداشته‌اند.

تا این قسمت ما به بررسی جدیدترین روش‌های موجود در این حوزه پرداختیم. در ادامه قصد داریم تا روش پایه‌ای DLS را که مبنای اصلی کار ماست در این فصل بیان کنیم. الگوریتم DLS به ما کمک می‌کند تا درک بهتری از مسئلهٔ پیش رو داشته باشیم. در ادامه روش DLS ارتقایافته را مطرح می‌کنیم که این روش می‌تواند به ما در رسیدن به روش پیشنهادی مد نظر کمک شایانی کند.

۲-۲- روش DLS

هدف اصلی این روش همانند سایر روش‌های حفظ حریم خصوصی، تلاش برای تولید مجموعه موقعیت‌های کورکورانه است که مهاجم را برای حمله گمراه کند. فرض کنید که نقشهٔ مکانی که در اختیار داریم به $n \times n$ سلول تقسیم شود. هریک از این سلول‌ها دارای یک معیار احتمال درخواست[۷] هستند که به این صورت محاسبه می‌شود:

(۱-۲)

$$q_i = \frac{\text{Number of Queries in Cell i}}{\text{Number of All Queries in Network}} \quad i=۱و۲و۳و\ldots \qquad \sum_{i=۱}^{n^۲} q_i = ۱$$

۷ Query Probability

در حالت کلی، روش DLS می‌بایست در مجموعهٔ کامل دیتابیس به دنبال یافتن k مکان کورکورانه باشد که ایـن k درجـهٔ ناشناختگی روش DLS است. در این k مکان ساخته‌شده یکی از آن‌ها مکان حقیقـی کـاربر و k-۱ مکـان دیگـر بـه‌صـورت کورکورانه انتخاب می‌شود.

روند الگوریتم DLS به‌صورت زیر اجرا می‌شود:

در گام اول، یک کاربر مشخص نیاز است تا درجهٔ ناشناختگی k را مشخص کند. هرچه این مقدار k انتخابی برای یک موقعیت بزرگ‌تر انتخاب شود، میزان درجهٔ ناشناختگی موقعیت کاربر بیشتر می‌شود. اما از طرفی، انتخاب مقادیر بزرگ‌تر k، هزینهٔ محاسبات را نیز به‌مراتب افزایش می‌دهد.

زمانی بیشترین درجهٔ ناشناختگی به دست می‌آید که تمامی k مکان تولیدی دارای احتمال یکسان در طرف سرور باشند. دلیل این امر کاملاً منطقی است زیرا داشتن احتمال یکسان موجب می‌شود تا مهاجم در انتخاب مکان حقیقی دچار مشکل شود. در ابتدای الگوریتم DLS، کاربر می‌بایستی همهٔ احتمال درخواست‌های شبکه را فراخوانی کرده و به‌صورت نزولی جای‌گذاری کند. در لیست مرتب‌شده، اگر چند مکان مختلف با مکان واقعی احتمال یکسان داشتند، نیمی از این مکان‌ها در سمت راست و نیمی دیگر در سمت چپ مکان واقعی قرار می‌دهیم (Niu-۲۰۱۴). در این دسته‌بندی k سلول در سمت چپ و k سلول در سمت راست سلول حقیقی قرار می‌گیرند. درنتیجه در ۲k مکان کاندید جدید، m مجموعه از سلول‌ها با k سلول در هر مجموعه تولید می‌شوند. در هریک از این m مجموعه، یکی از درایه‌ها مکان واقعی و k-۱ درایهٔ دیگر به‌صورت تصادفی انتخاب می‌شوند. سلول‌های دیگر به‌صورت تصادفی از ۲k کاندیدای دیگر انتخاب می‌شوند. j امین مجموعه را می‌توان به‌صورت C نشان داد:

$C_j=[c_{j1},c_{j2},\ldots,c_{jk}]$

بر اساس احتمال درخواست اصلی سلول‌های انتخاب‌شده، احتمالات درخواست نرمال از سلول‌های موجود شامل pj۱، pj۲، ۰۰ ،۰ pji، ۰۰ ،pjk و محاسبه شده توسط:

(۲-۲)

$$p_i=\frac{q_{ij}}{\sum_{l=1}^{k} q_{ik}} \quad i=1,2,3,\ldots$$

دلیل انتخاب مکان‌های ۲k به‌عنوان کاندیداهای کورکورانه[۸]، افزایش میزان ناشناختگی است و اندازهٔ این مجموعه را می‌توان با توجه به نیاز کاربر تغییر داد.

اکنون، ما باید یک مجموعهٔ بهینه را برای دستیابی به مقدار بهینهٔ k برای کاربر تعیین کنیم. درجهٔ حریم خصوصی از راه حل ما با استفاده از معیار آنتروپی، که به‌طور گسترده‌ای در اندازه‌گیری حریم خصوصی کاربر استفاده می‌شود تضمین شده است. به‌طور خاص، برای یک انتخاب خاص مجموعه C_j ما معیار آنتروپی را به شکل زیر تعریف می‌کنیم:

(۳-۲)

$$H_j = -\sum_{i=1}^{k} p_{ji}.log_2 p_{ji}$$

در انتها، مجموعهٔ مکان کاربر را بر اساس معادلهٔ زیر تعیین می‌کند:

(۴-۲)

$$C = argmax\, H_j$$

شبه کد الگوریتم DLS به صورت زیر است(Niu-۲۰۱٤):

---

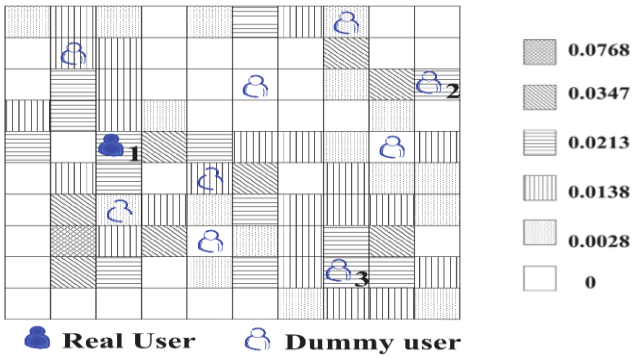
۸ Dummy

DLS Algorithm

Input : query probabilities in history q_i,
real location l_{real}, number of sets m, k
Output: an optimal set of dummy locations
۱ sort cells based on their query probability;
۲ choose $2k$ dummy candidates among which k candidates are right before l_{real} and k candidates are right after l_{real} in the sorted list;
۳ for $(j = 1; j \leq m; j++)$do
۴ construct set C_j which contains lreal and $k-1$ other cells randomly selected from the $2k$ candidates;
۵ compute the normalized probability p_{ji} for each cell c_{ji} in the set;
۶ $Hj \leftarrow -\sum_{i=1}^{k} p_{ji} \log_2 p_{ji}$
۷ end
۸ output $argmax\, H_j$;

شکل ۲-۱- شبه‌کد الگوریتم DLS (Niu-۲۰۱۴)

شکل ۲-۱ روند عمومی الگوریتم DLS را نشان می‌دهد. این الگوریتم می‌تواند به‌راحتی و به‌طور مؤثر ناشناس‌بودن با درجهٔ k را فراهم کند. اگرچه الگوریتم DLS می‌تواند به لحاظ انتروپی به میزان بیشتری از حریم خصوصی دست یابد، هنگام انتخاب مکان‌های ساختگی، بهتر است این مکان‌های ساختگی را دور از هم گسترش دهیم. به‌عنوان یک نتیجه، زمانی که مکان‌های ساختگی به یک منطقهٔ بزرگ‌تر گسترش می‌یابد مقدار آنتروپی DLS می‌تواند افزایش یابد. در این مقاله قصد داریم تا الگوریتم DLS و DLP را پیاده‌سازی کنیم و سپس روشی را که در (Niu-۲۰۱٤) برای ارتقای الگوریتم DLS استفاده شده بود را بر روی الگوریتم DLP که آنتروپی بسیار بهتری نسبت به الگوریتم DLS داشته است را اجرایی کنیم که به این وسیله احتمال شناسایی مکان کاربر را توسط مهاجم کاهش دهیم.

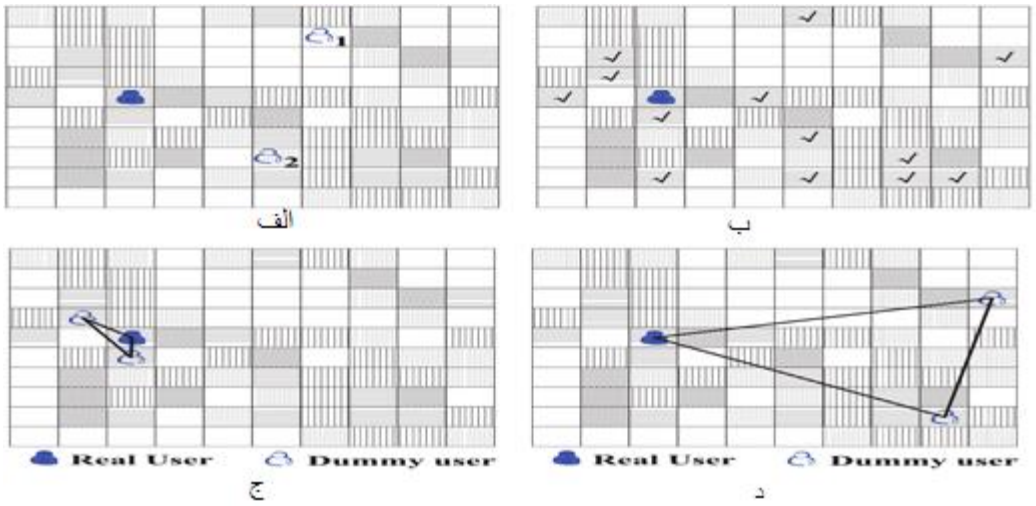


شکل ۲-۲ نمایش توزیع احتمال درخواست کاربر در یک مکان جغرافیایی همراه با مکان واقعی و موقعیت‌های کورکورانه

۲-۳- ارتقای الگوریتم DLS

برای دستیابی به درجهٔ ناشناختگی k، بسیاری از رویکردهای موجود به کاربران این امکان را می‌دهند تا اطلاعات کاربران اطراف را از طریق حالت نقطه به نقطه[۹] P۲P , یا روش مبتنی بر سرور مرکزی جمع‌آوری کنند. به‌طور شهودی، آن‌ها از نظر پراکندگی خوب نیستند. ما برای نشان‌دادن نگرانی‌های خود از یک مثال ساده در شکل ۲-۳ استفاده می‌کنیم.

[۹] Point to Point

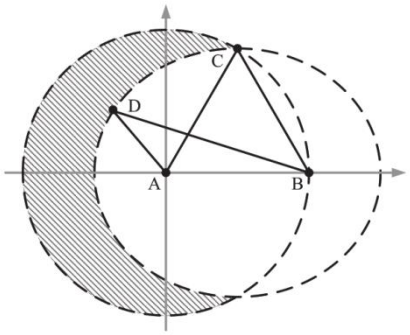


شکل۲-۳ الف) روش معمولی k-annonymity ب) روش DLS پیاده‌سازی شده ج) مکان‌های انتخابی در محدودهٔ کوچک‌تر د) مکان‌های انتخابی در محدودهٔ بزرگ‌تر(Niu-۲۰۱۴)

برای مثال، در شکل ۲-۳ قسمت (ج)، ممکن است مهاجم نیازی به حدس‌زدن مکان واقعی نداشته باشد، او می‌تواند کاربر واقعی را در یک منطقهٔ بسیار کوچک مانند بیمارستان یا کلینیک مستقر کند که این هم نوعی حریم خصوصی است. واضح است ما موردی را که در شکل ۲-۳ قسمت (د) نشان داده شده است ترجیح می‌دهیم زیرا می‌تواند زمینهٔ حفظ حریم خصوصی بیشتری را برای کاربر واقعی فراهم کند. اگرچه روش مرکزی مبتنی بر سرور می‌تواند منطقه‌ای از حریم خصوصی بزرگ‌تر را همان‌طور که در شکل ۳-۱ قسمت (د) نشان داده شده است، فراهم کند، اما سرور به‌تنهایی به مشکلی برای کل سیستم از نظر حریم شخصی و عملکرد سیستم تبدیل می‌شود. مکان‌های ساختگی می‌توانند مشکلات فوق را با توجه به مسئلهٔ حوزهٔ حریم خصوصی یا نگرانی‌های سرور مرکزی حل کنند.

در نتیجه برای بهبود سطح حریم خصوصی، الگوریتم DLS را می‌توان با درنظرگرفتن آنتروپی و CR[۱۰] (افزایش وسعت جغرافیایی) افزایش داد. از آنجا که دو عامل در نظر گرفته شده است، مسئلهٔ انتخاب ساختگی می‌تواند به‌عنوان یک مسئلهٔ بهینه‌سازی چند هدف (MOP) تنظیم شود. از یک طرف، ما می‌خواهیم بر اساس معیار آنتروپی، میزان حریم خصوصی مورد نیاز را به حداکثر برسانیم. از طرف دیگر، قصد داریم با پخش‌کردن مکان‌های ساختگی شاخص CR را افزایش دهیم. یک سؤال اساسی چگونگی اندازه‌گیری CR است. به‌طور شهودی، می‌توان از جمع فاصله بین جفت محل‌های ساختگی برای اندازه‌گیری CR استفاده کرد به‌صورت $\sum_{i\neq j} d(c_i,c_j)$ به‌طوری‌که $d(c_i,c_j)$ فاصلهٔ هندسی دو مکان c_i و c_j است. بااین‌حال، ممکن است حاصلجمع فاصله‌ها به‌خوبی حاصلضرب فاصله بین جفت محل‌های ساختگی نباشد. این حاصلضرب عبارت است از: $\prod_{i\neq j} d(c_i,c_j)$. مثالی را در شکل ۳-۲ در نظر بگیرید. در این مثال، A مکان واقعی کاربر است. B به‌عنوان یک مکان ساختگی انتخاب می‌شود زیرا دورترین مکان از A است. فرض کنید ما دو گزینه برای اختصاص مکان ساختگی سوم، C و D داریم. اگر آن را بر اساس میزان فاصله بین جفت مکان‌های ساختگی انتخاب کنیم، می‌توانید هریک از آن‌ها را انتخاب کنید، زیرا CA + CB = DA + DB. بااین‌حال، از نظر حریم خصوصی، ما C را نسبت به D ترجیح می‌دهیم زیرا مکان‌های ساختگی را بیشتر گسترش می‌دهد. درنتیجه، به‌جای استفاده از مجموع فاصله بین جفت محل‌های ساختگی، از ضرب‌های آن‌ها استفاده می‌کنیم. در این حالت CA.CB> DA.DB و ازاین‌رو C را به‌عنوان مکان ساختگی انتخاب می‌کنیم. شکل ۲-۴ مثالی از توضیحات گفته شده را به تصویر می کشد.

۱۰.Cloaking Region



شکل ۲-۴- مثالی از فاصله‌های مکانی و شکل هندسی آن‌ها (Niu-۲۰۱۴)

$C = [c_1, c_2, \dots c_k]$ را به‌عنوان مجموعه مکان‌های انتخابی درنظر می‌گیریم. درنتیجه، به شرط الگوریتم DLS در فصل قبل یک قید هندسی نیز اضافه می‌شد که به‌صورت زیر درمی‌آید.

(۵-۲)

$$\text{Max}\{-\sum_{i=1}^{k} p_{ji}.\log_2 p_{ji}, \prod_{i\neq j} d(c_i,c_j)\}$$

به‌طوری‌که p_i و p_j احتمال درخواست در الگوریتم DLS به‌ترتیب برای مکان‌های c_i و c_j را نشان می‌دهد. برقراری هم‌زمان این شروط در برخی موارد بسیار دشوار است. بنابراین برای ما ابتدا اولویت بر این است تا میزان آنتروپی را به حداکثر برسانیم و این کار را با معرفی مکان‌هایی انجام دهیم که بیشترین گستردگی و پراکندگی مکان را به ارمغان آورند.

تا این قسمت روش DLS ارتقایافته توضیح داده شد. در قسمت بعد، قصد داریم به معرفی روش موفق و قدرتمند DLP که در سال ۲۰۱۷ ارائه شد بپردازیم (Sun-۲۰۱۷). این روش به‌خودی‌خود در بسیاری از کاربردها استفاده می‌شود. اما در این مقاله قصد داریم تا روش جدیدی بر پایهٔ الگوریتم DLP ارائه کنیم تا به نوعی قیود هندسی نیز در محاسبات وارد شوند. برای این منظور ابتدا نیاز است تا روش DLP به‌طور کامل بحث شود.

ایدهٔ اصلی الگوریتم حفظ حریم خصوصی (DLP) انتخاب مکان‌های بهینهٔ ساختگی است. با توجه به اینکه اگر تعداد مکان‌های جعلی که احتمال درخواست آن‌ها با مکان واقعی کاربر مشابه است اندک باشد الگوریتم DLS دچار مشکل می‌شود اما الگوریتم DLP برای رفع این مشکل راه حل اندیشیده است. همچنین این الگوریتم برای یافتن مجموعه‌ای بهینه از مکان‌های ساختگی، یک روش حریصانه را جستجو می‌کند. برای دستیابی به ناشناس‌بودن k، ما به‌صورت پیوسته k-۱ مکان‌های دیگر را از همهٔ مکان‌های موجود در نقشه انتخاب می‌کنیم که باید مطمئن شویم آنتروپی فعلی بزرگ‌ترین است. برای مثال، اگر الگوریتم DLP قبلاً مکان‌های i را انتخاب کرده است (جایی که $i < k$)، هنگام انتخاب مکان (i + ۱)، باید اطمینان حاصل کنید که آنتروپی مکان بزرگ‌ترین آنتروپی را برای همهٔ مکان‌های باقیمانده تضمین می‌کند. به‌صورت زیر:

(۶-۲)

$$H_{i+1} = -\sum_{j=1}^{i+1} \frac{p_j}{\sum_{l=1}^{i+1} p_l} \log_2 \frac{p_j}{\sum_{l=1}^{i+1} p_l}$$

که p_j بیانگر تاریخچهٔ احتمال درخواست در مکان j است. در ادامه نشان خواهیم داد که چگونه الگوریتم DLP به بیشینهٔ مقدار آنتروپی دست می‌یابد.

ابتدا یک کاربر باید درجهٔ ناشناس‌بودن مناسب k را تنظیم کند که با نیاز کاربر برای حفظ حریم خصوصی ارتباط نزدیکی دارد. اگرچه یک k بزرگ‌تر به درجهٔ ناشناسی بیشتر منجر می‌شود، اما به‌دلیل هزینهٔ انتخاب مکان‌های ساختگی، باعث بالاتررفتن حجم محاسبات نیز می‌شود.

در ابتدا، الگوریتم DLP باید تمام احتمالات درخواست به دست‌آمده را از سرور LBS بخواند و سپس احتمالات درخواست را به‌ترتیب صعودی مرتب کند. p احتمال درخواست موقعیت واقعی کاربر را مشخص می‌کند. در لیست مرتب‌شدهٔ احتمالات درخواست، الگوریتم DLP تعداد مکان هایی که احتمال درخواست یکسان با p است را محاسبه می‌کند که توسط $\bar{k}$ مشخص شده است. اگر $\bar{k}$ به اندازهٔ کافی بزرگ باشد، نیمی از آن‌ها را قبل و نیم دیگر را بعد از موقعیت واقعی قرار می‌دهد.

اگر $\bar{k} \geq k$ باشد، الگوریتم DLP مکان‌های k-۱ را انتخاب می‌کند که احتمال درخواست یکسان با p از لیست مرتب‌شده را دارند. سپس، آن را از محل انتخاب ساختگی k-۱ و محل واقعی کاربر خارج می‌کند.

اگر $k \leq \bar{k} \leq k / ٤$، الگوریتم مکان‌های $\bar{k}$-۱ را که احتمال درخواست یکسان با p است از لیست مرتب‌شده انتخاب می‌کند. ما از مجموعهٔ C برای مشخص‌کردن مکان‌های ساختگی $\bar{k}$ -۱ و مکان واقعی کاربر استفاده می‌کنیم. در لیست مرتب‌شده، الگوریتم k-$\bar{k}$ سمت چپ و سایر k-$\bar{k}$ را بلافاصله پس از مکان واقعی به‌عنوان $٢(k-\bar{k})$ مکان نامزد انتخاب می‌کند که احتمالات درخواست با p متفاوت است. مجموعهٔ S بیانگر $٢(k-\bar{k})$ است. دلیل انتخاب نامزد $٢(k-\bar{k})$ برای مکان‌های ساختگی این است که حتماً آنتروپی به اندازهٔ کافی بزرگ به دست آید.

برای دستیابی به ناشناس‌بودن k، لازم است مکان‌های k-$\bar{k}$ باقیمانده را از مجموعهٔ S انتخاب کنید. وقتی اندازهٔ C برابر k باشد، DLP مجموعهٔ C را به‌عنوان خروجی نهایی ارائه می‌دهد.

اگر $\bar{k}<k / ۴$ ، DLP مکان‌های $٢k-\epsilon$ سمت چپ و سایر مکان‌های $٢k-\epsilon$ را بلافاصله پس از مکان واقعی به‌عنوان نامزدهای $٤k-\epsilon-\omega$ از لیست مرتب‌شده انتخاب می‌کند. ما از مجموعهٔ $\bar{S}$ برای مشخص‌کردن نامزدهای $٤k-\epsilon-\omega$ استفاده می‌کنیم. هردو ω و ε توسط کاربران براساس شرایط حفظ حریم خصوصی آن‌ها تنظیم شده است. به‌طور کلی، ω کوچک‌تر از ε است. بگذارید مجموعهٔ $\bar{C}$ مکان واقعی کاربر را مشخص کند. به‌طور تصادفی یک مکان را به‌عنوان یک مکان ساختگی از مجموعهٔ S انتخاب می‌کند و این مکان را در مجموعهٔ $\bar{C}$ قرار می‌دهد.

برای دستیابی به درجهٔ ناشناس‌بودن k، پیاپی مکان‌های K-۲ باقیمانده را از مجموعهٔ $\bar{S}$ انتخاب می‌کند. برای دستیابی به امین موقعیت مکانی، باید اطمینان حاصل شود که H_i برای همهٔ مکان‌های باقیمانده در مجموعهٔ C, l مقدار بیشینه است.

وقتی اندازهٔ $\bar{C}$ برابر k باشد، DLP مجموعهٔ $\bar{C}$ را به‌عنوان خروجی نهایی تحویل می‌دهد. شبه کد الگوریتم DLP در شکل ۲-۵ آورده شده است.

DLP Algorithm

Input: The set of historical query probabilities P; users' real location.
Output: The optimal set of dummy locations, C.
۱: Sort P in ascending order;
۲: $H \leftarrow$ select the locations which have the same query probability as users' real location from sorted P;
۳: **if** (*size* $(H) \geq k$) **then**
٤: $C \leftarrow$ randomly select k locations including the user real location from H;
٥: **else if** ($k/٤ <$ *size*$(H) < k$) **then**
٦: $\bar{k} \leftarrow$ *size*(H), $C \leftarrow H$;
٧: $S \leftarrow$ choose ٢$(k-\bar{k})$ candidate locations whose query probabilities are similar to the user's real location;
٨: **for** ($j = ١; j \leq k-\bar{k}; j++$) **do**
٩: Choose one location l from set S, such that $H(C, q)$ is the maximum in set S;
١٠: $C \leftarrow C \cup \{l\}$, $S \leftarrow S \setminus \{l\}$;
١١: **end for**
١٢: **else**
١٣: $S \leftarrow$ choose ٤$k-\omega-e$ candidate locations whose query probabilities are similar to the user's real location;
١٤: Randomly choose location i from S;
١٥: $C \leftarrow H \cup \{i\}$;
١٦: **for** ($j = ١; j \leq k-٢; j++$) **do**
١٧: Choose one location h from S, which makes sure that $H(C, q)$ is the maximum in set S;
١٨: $C \leftarrow C \cup \{h\}$, $S \leftarrow S \setminus \{h\}$;
١٩: **end for**
٢٠: **end if**
٢١: **return** the optimal set of dummy locations, C.

شکل۲-۵- شبه کد الگوریتم DLP (Sun-۲۰۱۷)

۳-روش پیشنهادی DLP ارتقا یافته

تا کنون روش DLP به طور کامل مطرح گردید. همچنین بر اساس کار صورت گرفته در (Niu-۲۰۱٤)، که در آن الگوریتم DLS را با در نظر گرفتن قیود هندسی پیاده‌سازی نموده، روش DLP را ارتقا دهیم. در این الگوریتم زمانی که مقادیر

مجموعه H تعیین گردید نیاز است تا شرط هندسی بررسی گردد و مکان‌های انتخابی در مجموعه H بر اساس موقعیت جغرافیایی نیز تجزیه و تحلیل گردند. الگوریتم روش پیشنهادی در شکل ۳-۱ آورده شده است.

Enhance DLP Algorithm

Input: The set of historical query probabilities P; users' real location.
Output: The optimal set of dummy locations, C.
۱: Sort P in ascending order;
۲: $H \leftarrow$ select the locations which have the same query probability as users' real location from sorted P;
۳: **if** ($size$ (H) $\geq k$) **then**
۴: Calculate $H.dist$ as product of distance from real location for each member of H

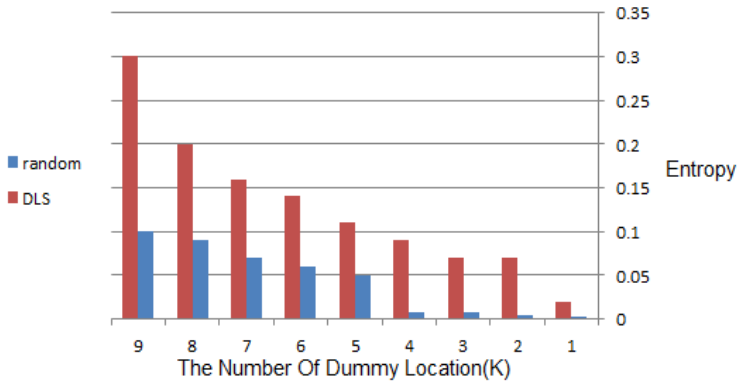
$$H.dist_i = \prod_{1 \le i \le k} d(H_i(c_i, c_{real}))$$

۵: Sort $H.dist$ in ascending order;
۶: $C \leftarrow$ select first k-۱ locations of H and the user real location;
۷: **else if** ($k/4 < size(H) < k$) **then**
۸: Calculate $H.dist$ as line ۴ and sort $H.dist$ in ascending order;
۹: $\bar{k} \leftarrow size(H)$, $C \leftarrow H$;
۱۰: $S \leftarrow$ choose ۲(k-$\bar{k}$) candidate locations whose query probabilities are similar to the user's real location;
۱۱: **for** (j = ۱; $j \le k$-$\bar{k}$; j ++) **do**
۱۲: Choose one location l from set S, such that $H(C, q)$ is the maximum in set S;
۱۳: $C \leftarrow C \cup \{l\}$, $S \leftarrow S \setminus \{l\}$;
۱۴: **end for**
۱۵: **else**
۱۶: $S \leftarrow$ choose ٤k-ω-e candidate locations whose query probabilities are similar to the user's real location;
۱۷: Choose location i from S which has longest distance to real location ;
۱۸: $C \leftarrow H \cup \{i\}$;
۱۹: **for** (j = ۱; $j \le k$-۲; j + +) **do**
۲۰: Choose one location h from S, which makes sure that $H(C, q)$ is the maximum in set S;
۲۱: $C \leftarrow C \cup \{h\}$, $S \leftarrow S \setminus \{h\}$;
۲۲: **end for**
۲۳: **end if**
۲٤: **return** the optimal set of dummy locations, C.

شکل ۳-۱- شبه‌کد الگوریتم پیشنهادی DLP ارتقایافته

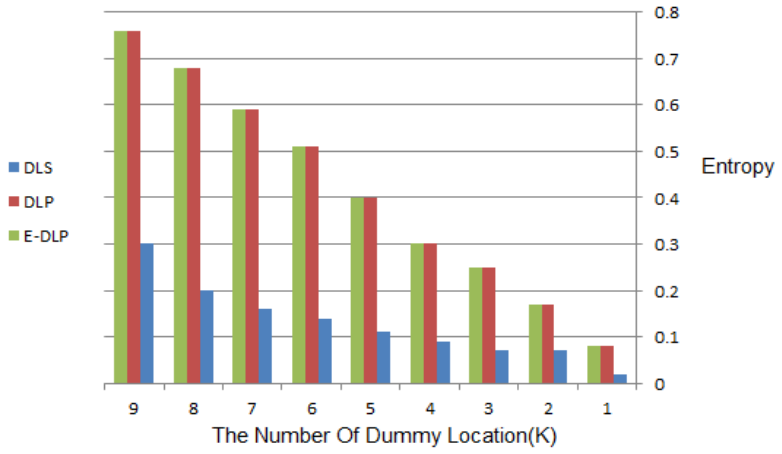
در این الگوریتم تا زمانی که تعداد مکان‌های انتخابی که احتمال درخواست برابر با کاربر اصلی را دارند از k/۴ بیشتر باشد، شرط قید هندسی بررسی می‌شود. در بررسی شروط معیار اول ما، انتخاب مکانی با آنتروپی بیشتر است. درنتیجه زمانی که k / k / ۴ ≤k̄≤ k یا $k/4 > \bar{k}$ باشد، دیگر اعمال شرط هندسی قابل اجرا نیست و ممکن است شرط آنتروپی ماکزیمم به دست نیاید. این الگوریتم به ما کمک می‌کند تا در مواقعی که شرط آنتروپی بیشینه برقرار است بتوانیم مکان‌هایی را انتخاب کنیم که بیشترین پراکندگی را برای ما به ارمغان بیاورند. این امر موجب می‌شود تا مهاجم در صورت دسترسی به اطلاعات جانبی از طرف سرور، بازهم در انتخاب محدودهٔ کاربر واقعی با احتمال خطای بیشتری روبه‌رو شود.

در اینجا ابتدا به پیاده‌سازی الگوریتم DLS بر روی یک آزمایش عملی می‌پردازیم. ابتدا نیاز است تا گستردگی کاربران در یک حوزهٔ جغرافیایی پیاده‌سازی شود. برای این شبیه‌سازی از یک منطقه با $n = ۱۰$ استفاده می‌کنیم. در این منطقه ۱۰۰ سلول وجود دارد که توزیع احتمال درخواست آن‌ها مانند شکل ۴-۱ است. این مدل در نرم‌افزار MATLAB پیاده‌سازی شده و نتایج خروجی برای روش DLS در مقایسه با انتخاب تصادفی موقعیت در شکل ۳-۲ نمایش داده شده است.



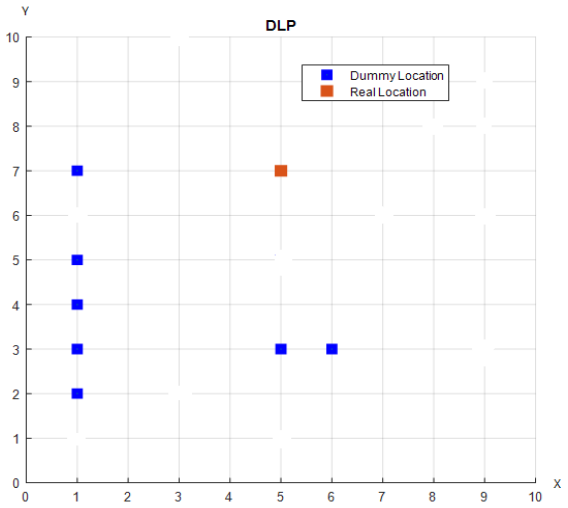
شکل ۳-۲- مقایسهٔ آنتروپی دو روش DLS و روش انتخاب مکان تصادفی به ازای درجات ناشناختگی مختلف

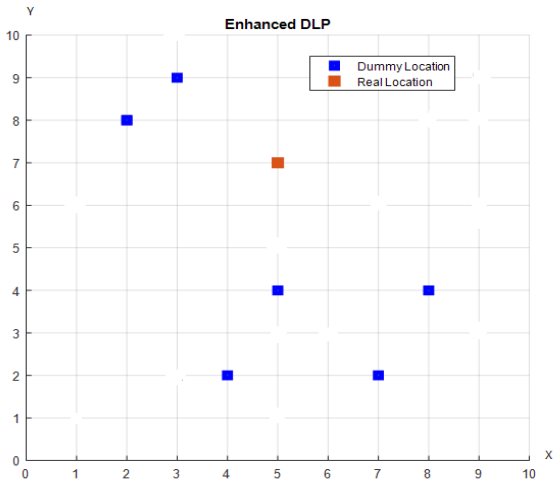
در قسمت بعد قصد داریم تا الگوریتم DLP را بر روی مدل توصیف‌شده در قسمت قبل پیاده‌سازی کنیم. شکل ۳-۳ خروجی روش‌های DLP، DLS و روش تصادفی را با یکدیگر مقایسه کرده است.



شکل۳-۳- مقایسهٔ خروجی روش‌های مختلف به ازای درجات ناشناختگی متفاوت

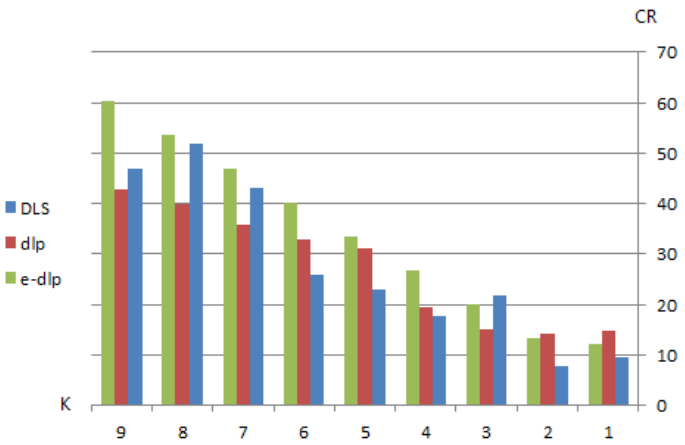
در ادامهٔ شبیه‌سازی‌های انجام‌شده هدف ما این است تا با اعمال قیود هندسی مطرح‌شده، الگوریتم DLP ارتقایافته را بررسی کنیم. همان‌طور که در شکل ۴-۳ پیداست توزیع مکان‌های انتخابی در دو روش با یکدیگر متفاوت است. حال آنکه آنتروپی به‌دست‌آمده برای هر دو مجموعه مکان‌های انتخابی برابر است. گفتنی است این مکان‌های مشترک به‌دلیل برقراری شرط آنتروپی ماکزیمم به‌دست آمده‌اند. یعنی الگوریتم پیشنهادی ابتدا به‌سبب تضمین آنتروپی بیشینهٔ مکان‌هایی که شرط لازم را برقرار می‌کنند انتخاب می‌کند و در جایی که شرط لازم آنتروپی بیشینه رعایت می‌شود می‌تواند نقاطی را به‌عنوان مکان منتخب برگزیند که توزیع وسیع‌تری دارند. این نکته در زمینهٔ روش پیشنهادی بسیار حائز اهمیت است زیرا روش پیشنهادی به‌نوعی مزیت اصلی روش DLP را حفظ می‌کند و درعین‌حال ضعف آن را نیز پوشش می‌دهد.





شکل ۳-۴- توزیع مکان‌های انتخابی به‌دست‌آمده در محیط جغرافیایی: الف) روش DLP ب) روش پیشنهادی

همان‌طور که در شکل ۳-۴ معلوم است، میزان پراکندگی داده‌ها در روش پیشنهادی به‌مراتب بهینه‌تر از روش DLP است. این مکان نشان می‌دهد که در حالتی که شرط آنتروپی رعایت می‌شود، روش پیشنهادی ما قادر است تا توزیع مناسب‌تری از مکان‌های انتخابی ارائه کند. برای مثال، فرض کنید هریک از خطوط مشبک روی سطح نشان‌دهندهٔ یک خیابان در یک منطقهٔ شهری است. همان‌طور که پیداست در روش DLP بیش از ۹۰٪ داده‌های غیرمشترک در یک خیابان تجمع یافته‌اند. حال آنکه در روش پیشنهادی ما پراکندگی داده‌ها به‌وضوح نشان می‌دهد که مکان‌های انتخابی به‌طور مناسب در خیابان‌های مختلف پراکنده شده‌اند. درنتیجه، برای یک مهاجم سخت‌تر است تا مکان واقعی کاربر را در روش پیشنهادی این مقاله تشخیص دهد و درصد خطای بالاتری خواهد داشت. این امر با شاخص مساحت منطقهٔ CR[۱۱] (تحت پوشش) قابل اندازه‌گیری است. شکل ۳ـ۵ شاخص CR را برای سه روش مطرح شده در این مقاله به ازای درجات ناشناختگی مختلف نشان می‌دهد.



شکل۳-۵- توزیع مکان‌های انتخابی غیرمشترک در روش های DLS , DLP و پیشنهادی

همان‌طور که در این شکل پیداست الگوریتم پیشنهادی شاخص CR که در مرجع (Niu-۲۰۱٤) مورد استفاده قرار گرفته را ارتقا بخشیده است.

۱۱ Cloaking Region

۴- نتیجه گیری

در این مقاله، ابتدا به‌صورت تئوری الگوریتم انتخاب DLS را بررسی کردیم که این رویکرد فعلی برای محافظت از حریم خصوصی کاربران در LBS برای اینترنت اشیاء است. سپس، به بحث راجع به الگوریتم DLP پرداختیم. برای حفظ مؤثر حریم خصوصی موقعیت مکانی کاربران، ما نیز با درنظرگرفتن تعادل بین هزینهٔ محاسباتی (یعنی پیچیدگی زمانی) و الزامات حفظ حریم خصوصی کاربران، الگوریتم جدید حفظ حریم خصوصی مکان DLP ارتقایافته را پیشنهاد می‌کنیم. براساس اطلاعات جانبی به‌دست‌آمده و شاخص آنتروپی، الگوریتم DLP به‌صورت حریصانه و کورکورانه مکان‌های ساختگی را برای دستیابی به سطح حریم مطلوب ناشناس‌بودن k انتخاب می‌کند. ما همچنین یک الگوریتم DLP ارتقایافته را پیشنهاد کردیم که هر دو آنتروپی و منطقهٔ CR را برای حفظ آنتروپی درنظر می‌گیرد و سعی می‌کند تا اطمینان حاصل کند که مکان‌های ساختگی انتخاب‌شده تا آنجا که ممکن است پخش شوند. نتایج ارزیابی نشان می‌دهد که الگوریتم پیشنهادی می‌تواند به‌طور قابل‌توجهی سطح حریم خصوصی را از نظر آنتروپی بهبود بخشد. الگوریتم DLP ارتقایافته می‌تواند ضمن حفظ سطح حریم خصوصی مشابه الگوریتم DLP ، منطقهٔ تحت پوشش را بزرگ‌تر کند.

مراجع

- ازبک زایی، م. اینترنت اشیا و هوشمندسازی شهری، اولین کنفرانس و نمایشگاه مدیریت و فناوری اطلاعات و ارتباطات، تهران. ۱۳۹۵

-عبدالناصر درگلاله, محمود پرموزه, فرشته سپهری, حفظ حریم خصوصی کاربران در سرویس های مبتنی بر مکان با استفاده از روش K-anonymity , همایش ملی فن آوری محاسبات و اطلاعات,۱۳۹۲

-سبکبار, ح . خدمات مکان مبنا(LBS) در خدمت مدیریت بازاریابی گردشگری, فصلنامه علمی - پژوهشی اطلاعات جغرافیایی سپهر, دوره ۲۱، شماره ۸۱، بهار ۱۳۹۱

personalized k-anonymity: Architecture and algorithms. - Gedik, B., & Liu, L. Protecting location privacy with Transactions on Mobile Computing, ۷(۱), ۱–۱۸-۲۰۰۸. IEEE

- A. Beresford and F. Stajano, “Location privacy in pervasive computing,”Pervasive Computing, IEEE, vol. ۲, no. ۱, pp. ٤٦ – ٥٥, jan-mar, ۲۰۰۳.

wireless lans,” in ACM MobiSys ۲۰۰۷. - T. Jiang, H. J. Wang, and Y.-C. Hu, “Preserving location privacy in

- Sun, Gang, et al. "Efficient location privacy algorithm for Internet of Things (IoT) services and applications." Journal of Network and Computer Applications ۸۹ : ۳-۱۳- ۲۰۱۷.

location privacy through camouflage,” in - J. Meyerowitz and R. Roy Choudhury, “Hiding stars with fireworks: ACM MobiCom, ۲۰۰۹.

- B. Niu, X. Zhu, H. Chi, and H. Li, “۳plus: Privacy-preserving pseudolocation updating system in location-based services,” in IEEE WCNC, ۲۰۱۳.

- B. Niu, X. Zhu, X. Lei, W. Zhang, and H. Li, “Eps: Encounterbased privacy-preserving scheme for location-GLOBECOM, ۲۰۱۳. based services,” in IEEE

- X. Zhu, H. Chi, B. Niu, W. Zhang, Z. Li, and H. Li, “Mobicache: When k-anonymity meets cache,” in IEEE GLOBECOM ۲۰۱۳.

- B. Niu, Z. Zhang, X. Li, and H. Li, “Privacy-area aware dummy generation algorithms for location-based services,” in IEEE ICC ۲۰۱٤.

- A. Aryan and S. Singh. Protecting location privacy in Augmented Reality using k-anonymization and pseudo-id. International Conference on Computer and Communication Technology, ۱۱۹-۱۲٤, ۲۰۱۰.

- B. Ying , D. Makrakis. Protecting Location Privacy with Clustering Anonymization in vehicular networks. IEEE INFOCOM Workshops, ۳۰٥-۳۱۰, ۲۰۱٤.

- K. Vu, R. Zheng and J. Gao. Efficient algorithms for k-anonymous location privacy in participatory sensing. IEEE INFOCOM, ۲۳۹۹-۲٤۰, ۲۰۱۲.

- X. Liu, K. Liu , L. Guo. A game-theoretic approach for achieving k-anonymity in Location Based- Services. IEEE INFOCOM, ٢٩٨٥- ٢٩٩٣, ٢٠١٣.
- B. Niu, Q. Li, X. Zhu, G. Cao. Achieving k-anonymity in privacy -aware location-based services. IEEE INFOCOM, ٧٥٤ -٧٦٢, ٢٠١٤.
- Sun, Gang, et al. "Efficient location privacy algorithm for Internet of Things (IoT) services and applications." Journal of Network and Computer Applications ٨٩ : ٣-١٣- ٢٠١٧.
- A. Hossain, S. Jang, J. Chang. Privacy-Aware Cloaking Technique in Location-Based Services. IEEE the First International Conference on Mobile Services, ٩-١٦, ٢٠١٢.
- X. Liu, H. Zhao, M. Pan, H. Yue, X. Li, and Y. Fang. Traffic aware multiple mix zone placement for protecting location privacy. IEEE INFOCOM, ٩٧٢-٩٨٠, ٢٠١٢.
- T. Jung, X. Li, Z. Wan and M. Wan. Privacy preserving cloud data access with multi-authorities. IEEE INFOCOM, ٢٦٢٥-٢٦٣٣, ٢٠١٣.
- Zhang, Shaobo, et al. "A caching and spatial K-anonymity driven privacy enhancement scheme in continuous location-based services." *Future Generation Computer Systems* ٩٤: ٤٠-٥٠. ٢٠١٩.
- Zhang, Shaobo, et al. "A Dual Privacy Preserving Scheme in Continuous Location-Based Services." *IEEE Internet of Things Journal* (٢٠١٨).
- Zhao, Ping, et al. "ILLIA: Enabling $ k $-Anonymity-Based Privacy Preserving Against Location Injection Attacks in Continuous LBS Queries." *IEEE Internet of Things Journal* ٥,٢: ١٠٣٣-١٠٤٢-٢٠١٨
- Du, Yongwen, et al. "An Efficient Dummy-Based Location Privacy-Preserving Scheme for Internet of Things Services." Information ١٠,٩ (٢٠١٩): ٢٧٨.
- Fei, Fan, et al. "A K-anonymity based schema for location privacy preservation." *IEEE Transactions on Sustainable Computing* (٢٠١٧).
- El Ouazzani, Zakariae, and Hanan El Bakkali. "A new technique ensuring privacy in big data: K-anonymity without prior value of the threshold k." *Procedia Computer Science* ١٢٧(٢٠١٨): ٥٢-٥٩.
- Jiang, Hongbo, Ping Zhao, and Chen Wang. "RobLoP: Towards robust privacy preserving against location dependent attacks in continuous LBS queries." IEEE/ACM Transactions on Networking ٢٦,٢ (٢٠١٨): ١٠١٨-١٠٣٢.