مجله محاسبات نرم
SOFT COMPUTING JOURNAL
تارنمای مجله: scj.kashanu.ac.ir
دانشگاه کاشان
University of Kashan

روشی برای حفاظت از الگوی دسترسی در داده‌های برون‌سپاری‌شده✧

معصومه باباخانی[۱]، کارشناسی ارشد، داود محمدپورزنجانی[۲*]، استادیار، لیلا صفری[۳]، استادیار
[۱،۲ و ۳] دانشکده فنی و مهندسی، دانشگاه زنجان، زنجان، ایران.

اطلاعات مقاله

تاریخچه مقاله:
دریافت ۲۴ شهریور ماه ۱۳۹۸
پذیرش ۲۳ خرداد ماه ۱۳۹۹

کلمات کلیدی:
برون‌سپاری داده‌ها
محرمانگی داده
حفاظت از الگوی دسترسی
جلوگیری از افشای اطلاعات

چکیده

حفاظت از الگوی دسترسی اطلاعات، به‌معنی جلوگیری از افشای جزئیات داده‌ای و ساختاری بانک‌های اطلاعاتی، در کار با داده‌ها به‌ویژه در حالت بانک‌های اطلاعاتی برون‌سپاری‌شده و بانک‌های اطلاعاتی با دسترسی اینترنتی، از اهمیت بسیاری برخوردار است. حفاظت از الگوی دسترسی اطلاعات به این نکته اشاره دارد که محرمانگی داده، به‌تنهایی کافی نیست و باید محرمانگی پرس‌وجوها و دسترسی‌ها نیز تأمین شوند؛ زیرا مهاجمان می‌توانند با مشاهدهٔ پرس‌وجوهای کاربران، روابط بین آن‌ها را استخراج و با دانشی که به دست می‌آورند، اقدام به رمزگشایی و استنتاج جزئیات داده‌ای و ساختاری بانک‌های اطلاعاتی کنند. در این مقاله، در مدل برون‌سپاری داده‌ها، روش‌های ذخیره‌سازی مناسب برای تأمین محرمانگی و حفاظت از الگوی دسترسی تشریح می‌شوند و در نهایت، روشی مبتنی بر قطعه‌بندی برای حفاظت از الگوی دسترسی در داده‌های برون‌سپاری‌شده، پیشنهاد و مورد ارزیابی قرار می‌گیرد. ارزیابی‌های انجام‌شده نشان از دستیابی به سطح قابل قبولی از جلوگیری از افشای اطلاعات نسبت به روش‌های پیشین این حوزه در کنار عدم تحمیل سربارهای زیاد ذخیره‌سازی، محاسباتی و ارتباطی دارد.

۱. مقدمه

با توجه به افزایش روزافزون حجم داده‌ها و نیاز به تخصص بالا و تجهیزات خاص برای نگهداری و فراهم کردن دسترسی مناسب به آن‌ها، به‌عنوان یک راهکار مناسب، مالکان داده، نگهداری و مدیریت داده‌های خود را به کارپذیر خارجی واگذار می‌کنند. به‌بیان دیگر، کاربران تمایل زیادی دارند که از خدمات رایانش ابری، به‌ویژه خدمت داده به‌عنوان خدمت استفاده کنند که از دیدگاه پایگاه‌داده‌ای با عنوان برون‌سپاری داده‌ها شناخته می‌شود [۱]. در تعریف برون‌سپاری داده‌ها چهار موجودیت اصلی وجود دارد که در شکل (۱) نشان داده شده‌اند [۲]:

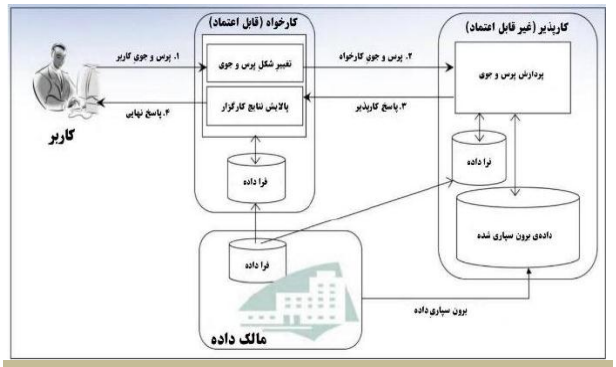
- مالک داده: فرد یا سازمانی که داده‌ها را برون‌سپاری و سیاست‌گذاریِ دسترسی به داده‌ها را تعیین می‌کند.
- کارپذیر: مسئول ذخیره‌سازی داده‌ها و ارائهٔ پاسخ به پرس‌وجوی کاربران است.
- کاربر: فردی است که پرس‌وجویی از پایگاه داده دارد.
- کارخواه: وظیفهٔ دریافت پرس‌وجوی کاربران و تبدیل آن‌ها به پرس‌وجوهای قابل فهم برای کارپذیران را دارد.

داده‌های برون‌سپاری‌شده اغلب شامل داده‌های حساس‌اند که باید از دسترسی و تغییرات مهاجمان در امان باشند. یک گام الزامی برای به‌کارگیری مفهوم برون‌سپاری داده‌ها، فراهم کردن

✧ نوع مقاله: پژوهشی
* نویسنده مسئول
پست‌های الکترونیک: m.babakhani@znu.ac.ir (باباخانی)
dmp@znu.ac.ir (محمدپور زنجانی)
lsafari@znu.ac.ir (صفری)

نحوه ارجاع به مقاله: باباخانی، معصومه، محمدپورزنجانی، داود، صفری، لیلا، «روشی برای حفاظت از الگوی دسترسی در داده‌های برون‌سپاری‌شده»، مجله محاسبات نرم، جلد ۸، شماره ۱، ص۲ـ۱۳، بهار و تابستان ۱۳۹۸.

تمهیدات لازم برای تأمین ویژگی محرمانگی داده‌ها و حفاظت از الگوی دسترسی است.



شکل (۱): برون‌سپاری داده‌ها

تأمین محرمانگی داده، به‌معنای جلوگیری از نشت محتوای داده و مصون نگاه داشتن داده از دسترسی توسط کاربران غیرمجاز است. به‌بیان دیگر، فقط کاربران مجاز اجازهٔ دسترسی به داده‌های برون‌سپاری‌شده را دارند و حتی کارپذیر نباید دربارهٔ محتوای داده‌ها اطلاعاتی به دست آورد؛ زیرا مالک داده انتظار دارد کاربران بتوانند بدون نشت داده‌ها از خدمات بهره‌مند شوند.

علاوه بر این، بسیاری از کاربران در هنگام استفاده از خدمات ابری تمایل دارند که هویت و عملیات خود را مخفی نگه دارند. برای مثال، در درخواست کاربران، کلمات کلیدی برای بازیابی داده‌ها و نتایج درخواست که توسط کارپذیر خارجی ارائه می‌شود، نباید در معرض دید کارپذیر و دیگر کاربران باشند.

در برون‌سپاری داده‌ها، مهاجم می‌تواند با تحلیل الگوی دسترسی کاربران[1] و انجام حملات مبتنی بر استنتاج[2] [۳]، به ویژگی‌های مهمی از داده دسترسی پیدا کند که موجب افشای اطلاعات و نقض محرمانگی داده‌ها می‌شود. در این مقاله، نحوهٔ افشای داده با تحلیل الگوی دسترسی کاربران مورد توجه قرار گرفته و مدل امنیتی مناسبی برای حفاظت از افشای داده ارائه شده است.

۱. ۱. امنیت داده‌های برون‌سپاری‌شده

به‌طور کلی در انتخاب مدل برون‌سپاری، مالک داده باید به دو نوع مهاجم داخلی و خارجی توجه کند. مهاجم داخلی همان کارپذیری است که وظیفهٔ ذخیره‌سازی داده‌ها را بر عهده دارد و از محتوای آن‌ها آگاه است. کارپذیری را که داده‌ها در آن ذخیره شده‌اند، «درستکار اما کنجکاو» می‌نامند [۴]. کارپذیر، وظایف خود را به‌درستی انجام می‌دهد و داده‌ها را حذف و دستکاری نمی‌کند، اما دربارهٔ محتوا و اهمیت داده‌ها کنجکاو است. اما مهاجمان خارجی اغلب سعی در انجام عملیات خرابکارانه دارند، ابتدا شمای پایگاه داده و محتوای داده‌های ذخیره‌شده را یاد می‌گیرند و سپس اقدام به دستکاری و حذف داده‌ها می‌کنند. البته ممکن است این مهاجمان نیز مانند مهاجمان داخلی فقط دربارهٔ محتوای داده‌ها و کاربران کنجکاو باشند.

استفاده از الگوریتم‌های رمزنگاری قابل جست‌وجو، یکی از مکانیزم‌های مرسوم برای حفاظت از داده‌هاست. در واقع از آنجایی که کارپذیر «درستکار اما کنجکاو» است، مالک داده، قبل از برون‌سپاری داده‌ها، آن‌ها را در کارخواه (بخش قابل اعتماد) رمزنگاری کرده، سپس در کارپذیر (غیرقابل اعتماد) ذخیره می‌کند. این کار برای تأمین محرمانگی داده‌ها صورت می‌گیرد، اما رمزنگاری قابل جست‌وجوی داده‌ها، به‌تنهایی برای تأمین امنیت آن‌ها کافی نیست. برای مثال فرض کنید کاربران پرس‌وجوی خود را با کلمات کلیدی به کارپذیر ارسال می‌کنند و داده‌ها و پرس‌وجوی کاربران رمزنگاری شده‌اند، اما باز هم امکان نشت مقدار قابل توجهی از اطلاعات وجود دارد؛ زیرا مهاجم (کارپذیر) می‌تواند تعداد تکرار کلمات کلیدی را بشمارد و فراوانی آن‌ها را به دست آورد و سپس با دانش قبلی خود مقایسه کرده و اقدام به رمزگشایی کند. بنابراین محرمانگی داده‌ها به‌تنهایی کافی نیست و علاوه بر آن باید محرمانگی دسترسی و حفاظت از الگوی دسترسی نیز تأمین شود.

۲.۱. روش‌های تأمین محرمانگی داده‌ها

روش‌هایی مانند رمزنگاری نرم‌افزاری، قطعه‌بندی داده‌ها یا ترکیب هر دو و نیز رمزنگاری سخت‌افزاری برای تأمین محرمانگی داده‌ها استفاده می‌شوند [۵ و ۸].

در رمزنگاری نرم‌افزاری، داده‌ها در بخش امن (کارخواه) رمزنگاری شده و سپس در کارپذیر ذخیره می‌شوند، در نتیجه

1. Data Access Patterns
2. Inference Attack

همانند مهاجمان خارجی، کارپذیر از محتوای داده‌ها آگاه نیست. اما در حالت پایگاه‌داده‌ای، استفاده از رمزنگاری قابل جست‌وجو پیشنهاد می‌شود تا اجرای پرس‌وجوها و جست‌وجو روی داده‌های رمزشده به‌صورت مستقیم در کارپذیر (خدمت‌دهندهٔ پایگاه داده) انجام گیرد و در نتیجه نیاز به انتقال کل داده به کارخواه نیست [۵].

علاوه بر رمزنگاری می‌توان از تکنیک قطعه‌بندی داده‌ها نیز استفاده کرد. در این روش، عمل تفکیک داده‌ها با هدف از بین بردن ارتباطات حساس بین داده‌ها انجام می‌گیرد. با توجه به اینکه با قطعه‌بندی داده‌ها نیز همچنان امکان نشت اطلاعات وجود دارد، استفاده از ترکیب دو تکنیک رمزنگاری و قطعه‌بندی توصیه شده است. بدین صورت که اگر بعد از قطعه‌بندی داده‌ها، باز هم امکان نشت اطلاعات وجود داشته باشد، برخی از داده‌ها رمزنگاری می‌شوند [۲].

۳.۱. روش‌های حفاظت از الگوی دسترسی

همان طور که عنوان شد، محرمانگی داده‌ها به‌تنهایی کافی نیست، حفاظت از الگوی دسترسی، محرمانگی را تقویت می‌کند [۹]. از روش‌های حفاظت از الگوی دسترسی به موارد زیر می‌توان اشاره کرد:

- ارسال پرس‌وجوهای ساختگی: مهاجم می‌تواند با استفاده از تحلیل فراوانی و نیز ترتیب درخواست‌ها، اطلاعاتی را استنتاج کند. بنابراین نیاز است تا ترتیب درخواست‌ها و تعدادشان تغییر کند. بدین منظور پرس‌وجوهای ساختگی با هدف بازیابی داده‌هایی که به دفعات کمتری درخواست شده‌اند، به کارپذیر ارسال می‌شوند تا مهاجم را گمراه کنند [۹].
- ذخیره‌سازی داده‌های جعلی: در این روش، داده‌های مالک، همراه با داده‌های ساختگی در کارپذیر ذخیره می‌شوند. داده‌های جعلی می‌توانند برای پوشش درخواست‌ها از طریق درخواست ساختگی بازیابی شوند. این کار تشخیص داده‌های هدف را برای مهاجم سخت می‌کند [۹ و ۲].
- جابه‌جایی: فراوانی دسترسی به یک بلوک حافظه، نشان‌دهندهٔ اهمیت آن بلوک است. در روش جابه‌جایی، رابطهٔ بین بلوک‌ها و داده‌ها شکسته می‌شود و آدرس بلوک در هر دسترسی تغییر می‌یابد. این کار موجب تصادفی شدن تمام درخواست‌ها می‌شود و تکرار آدرس یک بلوک به‌معنای تکرار یک داده نیست [۹ و ۲].

استفاده از حافظهٔ نهان: در این روش، کارخواه حافظهٔ امنی فراهم می‌کند تا داده‌هایی که به دفعات زیادی مورد درخواست قرار می‌گیرند، برای مدتی در حافظهٔ نهان نگهداری و از آنجا به کاربر تحویل داده شوند. با این اقدام از استنتاج مهاجم بر اساس تعداد دفعات درخواست، جلوگیری می‌شود. البته از آنجایی که مهاجم از زمان خواندن داده‌ها در کارپذیر آگاه است، در هنگام بازیابی داده‌ها از حافظهٔ نهان، باید پرس‌وجوهای ساختگی نیز به کارپذیر ارسال شود [۹ و ۳].

۲. روش پیشنهادی

همان طور که پیش‌تر بیان شد، رمزنگاری قابل جست‌وجو راه‌حلی برای محافظت از داده‌های برون‌سپاری‌شده است. اما امکان نشت اطلاعات وجود دارد که به یادگیری مهاجمان در مورد داده‌های حساس کمک می‌کند. به‌طوری که مهاجمان با مشاهده و ردیابی محل فیزیکی داده‌های ذخیره‌شده و مشخص شدن نوع و تکرار پرس‌وجوها و نتایج آن‌ها می‌توانند اطلاعاتی استنتاج کنند. این یادگیری بر اساس الگوی دسترسی کاربران صورت می‌گیرد، بنابراین لزوم حفاظت از الگوی دسترسی برای جلوگیری از نشت اطلاعات بیان شد.

روش ارائه‌شده در مقالهٔ حاضر که اصلاح‌شدهٔ روش APHQ از مقالهٔ [۱۰] می‌باشد، برای حفاظت از الگوی دسترسی پیشنهاد شده است، به‌طوری که بتواند با استفاده از چندین کارپذیر و با قطعه‌بندی افقی داده‌ها، نسبت فراوانی کلمات کلیدی در بین داده‌های یک کارپذیر را تغییر دهد. در این صورت هر کارپذیر، فراوانی کلمات کلیدی را به‌صورت نادرستی مشاهده می‌کند. مزیت این نوع قطعه‌بندی این است که با تکنیک‌های نرمال‌سازی داده‌ها همراه است؛ یعنی در هر کارپذیر یک نمونه پایگاه دادهٔ نرمال‌شده ذخیره می‌گردد به‌طوری که محرمانگی و حریم خصوصی آن‌ها تأمین شده

است. این قطعه‌بندی به‌گونه‌ای است که پرس‌وجوها با نتایج یکسان یا مشابه، به یک کارپذیر تخصیص داده می‌شوند. این نوع قطعه‌بندی علاوه بر بهبود حفاظت از الگوی دسترسی، موجب بهبود افشای اطلاعات و زمان پاسخ می‌شود.

در روش پیشنهادی با فرض سه کارپذیر، رابطهٔ اصلی R به‌صورت نمونه، به سه قطعه تقسیم می‌شود. دلیل انتخاب سه قطعه بیان ساده‌تر مراحل روش است. در ادامه، نحوهٔ تخصیص رکوردها و پرس‌وجوها به کارپذیران تشریح می‌شود. همچنین بهبودهای اعمال‌شده نسبت به مرجع [۱۰] از لحاظ استراتژی توزیع رکوردها و پرس‌وجوها، با لحاظ کردن پارامترهای مناسب‌تری در محاسبهٔ تابع هدف بیان می‌گردد؛ که نشان از نتایج بهتری به‌لحاظ جلوگیری از افشای اطلاعات دارد.

۱.۲. استراتژی توزیع رکوردها و پرس‌وجوها

توزیع رکوردها و پرس‌وجوها در چندین کارپذیر باید به‌گونه‌ای باشد که زمان پاسخ‌گویی حداقل باشد و علاوه بر آن الزامات امنیتی را ارضا کند. فرض کنید $R=\{r_1,r_2,r_3,...,r_N\}$ مجموعه ای از رکوردهای پایگاه داده با چندین صفت باشد و $Q=\{q_1,q_2,q_3,...,q_m\}$ دنباله‌ای از پرس‌وجوها باشد که نتیجهٔ هریک تعدادی از رکوردهای R باشد. مجموعهٔ $S=\{S_1,S_2,S_3,...,S_k\}$ کارپذیران خارجی هستند که رکوردها و پرس‌وجوها به آن‌ها تخصیص داده می‌شوند. قبل از توزیع رکوردها و پرس‌وجوها، مالک داده، ماتریسی به‌صورت زیر به نام B می‌سازد که شامل نتایج همهٔ پرس‌وجوهاست:

$$B=\begin{pmatrix} x_{1,1} & \cdots & x_{1,m} \\ \vdots & \ddots & \vdots \\ x_{N,1} & \cdots & x_{N,m} \end{pmatrix} \qquad (۱)$$

در معادلهٔ (۱)، هر ستون ماتریس B نشان‌دهندهٔ یک پرس وجو از مجموعه $Q=\{q_1,q_2,...,q_m\}$ است و هر سطر ماتریس B نشان‌دهندهٔ یک رکورد از $R=\{r_1,r_2,...,r_N\}$ است. برای هر عنصر از این ماتریس ($x_{r,q}\in B$)، اگر رکورد $r\in R$ در مجموعه پاسخ $q\in Q$ باشد، مقدار آن عنصر ($x_{r,q}$)، یک و در غیر این صورت مقدار آن، صفر تنظیم می‌شود.

با توجه به ماتریس B توزیع رکوردها طبق مراحل زیر انجام می‌شود:

- مرحلهٔ ۱: برای توزیع رکوردها و پرس‌وجوها در چند کارپذیر، درایه‌هایی از ماتریس B که دارای مقدار یک هستند انتخاب و به‌صورت تصادفی به یکی از کارپذیران تخصیص داده می‌شوند. این فرایند تا زمانی تکرار می‌شود که همهٔ درایه‌های حاوی مقدار یک، از ماتریس B به همهٔ کارپذیران اختصاص داده شوند. در تخصیص رکوردها باید به این نکته توجه کرد که تا حد امکان هیچ کارپذیری خالی نماند.
- مرحلهٔ ۲: برای هر کارپذیر، ماتریس دودویی به نام A^S ساخته می‌شود. نمایشی از آن در معادلهٔ (۲) آورده شده است:

$$A^S=\begin{pmatrix} b_{1,1} & \cdots & b_{1,m} \\ \vdots & \ddots & \vdots \\ b_{N,1} & \cdots & b_{N,m} \end{pmatrix} \qquad (۲)$$

$A_Q^S=\left[q_{1,s},q_{2,s},...,q_{m,s}\right]$ نشان‌دهندهٔ ستون‌های ماتریس A^S است که شامل پرس‌وجوهای اختصاص داده‌شده به کارپذیر S می‌باشد.

$A_R^S=\left[r_{1,s},r_{2,s},...,r_{N,s}\right]$ نشان‌دهندهٔ سطرهای ماتریس A^S است که شامل رکوردهای تخصیص داده‌شده به کارپذیر S می‌باشد.

درایه‌هایی که مقدار یک دارند به‌معنی تخصیص رکورد و پرس‌وجوی متناظر با آن درایه، در کارپذیر S است.

ماتریس A^S برای هر کارپذیر ساخته می‌شود و تخصیص اولیهٔ آن به‌صورت تصادفی انجام می‌گیرد. در مراحل اجرای الگوریتم، این ماتریس به‌روزرسانی می‌شود. هدف از تعریف چنین ماتریسی این است که بدانیم چه رکوردها و پرس‌وجوهایی به هر کارپذیر تخصیص داده شده است [۱۰].

۲.۲. افشای اطلاعات

با مشاهدهٔ پرس‌وجوها و نتایج بازگشتی آن‌ها، مهاجم الگوی دسترسی کاربران را استخراج می‌کند که می‌تواند منجر به افشای اطلاعات شود. برای تعریف افشای اطلاعات هر کارپذیر، از

ماتریس A^S آن کارپذیر استفاده می‌شود. طبق توضیحاتی که قبلاً ارائه شد، می‌دانیم که A^S بخشی از ماتریس B است و ستون‌های ماتریس A^S، نشان‌دهندهٔ نتایج پرس‌وجوها در کارپذیر S می‌باشد. برای بیان مفهوم افشای اطلاعات، مفهومی به نام D(A) تعریف می‌شود؛ D(A) از موارد زیر ساخته شده است:

- با درنظرگرفتن یک حد آستانهٔ پذیرفته‌شده برای افشای اطلاعات با عنوان α که $0 \le \alpha \le 1$ می‌باشد، اختلاف بین دو پرس‌وجو در هر کارپذیر (مانند S) باید کمتر از مقدار α باشد [۱۰]. برای محاسبهٔ اختلاف بین دو پرس‌وجو، از فاصلهٔ همینگ[1] استفاده می‌شود. به این صورت که دو ستون از ماتریس A^S انتخاب شده و همینگ آن‌ها حساب می‌شود و در نهایت طبق معادلهٔ (۳)، بررسی می‌شود که آیا افشای اطلاعات وجود دارد یا خیر.

$$\psi = \frac{ham\left(A^s_{q_1}, A^s_{q_2}\right)}{rows\left(A^s\right)} \qquad (۳)$$

تابع ham که همان تابع همینگ است، برای محاسبهٔ اختلاف مقدار بین دو ستون از ماتریس A^S مورد استفاده قرار گرفته است. تابع $rows$ تعداد رکوردهای اختصاص داده‌شده به کارپذیر S است. با درنظرگرفتن $D(A^S)$ به‌عنوان مقدار افشای اطلاعات ماتریس A^S، می‌توان آن را بدین صورت محاسبه کرد:

- اگر مقدار ψ از α بزرگ‌تر باشد، می‌توان نتیجه گرفت که پرس‌وجوهای q_1, q_2 فاش‌کنندهٔ اطلاعات هستند. در این حالت، مقدار ψ، به $D(A^S)$ اضافه می‌شود.
- اگر نتایج یک پرس‌وجو به طور کامل به یک کارپذیر اختصاص داده شود، به این معناست که آن پرس‌وجو، افشاکنندهٔ اطلاعات است، پس باید به‌ازای آن پرس‌وجو، مقدار ۱ به $D(A^S)$ اضافه گردد.

در نهایت، افشای اطلاعات کلی (تمامی کارپذیران) به‌صورت معادلهٔ (۴) تعریف می‌شود:

$$D(A) = \frac{\sum_s D(A^s)}{\sum_s M(A^s) + N_Q} \qquad (۴)$$

$M(A^s)$ بزرگ‌ترین $D(A^s)$ ممکن برای کارپذیر S است. N_Q تعداد کل پرس‌وجوهای موجود در همهٔ کارپذیران است و $\sum_s M(A^S) + N_Q$ برای نرمال‌سازی مورد استفاده قرار گرفته است.

۳.۲. زمان پاسخ

زمان پاسخ پرس‌وجو برای ارزیابی چگونگی توزیع رکوردها و پرس‌وجوها در کارپذیران استفاده می‌شود [۱۰]. اگر m تعداد کل پرس‌وجوها فرض شود، $q_1, q_2, q_3, \dots, q_m$ توسط مالک داده از قبل شناسایی شده‌اند. فراوانی اجرای هر پرس‌وجو توسط یک بردار $F = [f_1, f_2, f_3, \dots, f_m]$ تعریف می‌شود و به‌طوری که $\sum_{q=1}^{q=m} f_q = 1$ می‌باشد. زمان پاسخ هر کارپذیر به پرس‌وجو شامل دو بخش است: زمان پردازش محلی و زمان انتقال نتایج.

زمان پردازش محلی هر کارپذیر به‌صورت یک صف مدل می‌شود و به‌عنوان مقدا برای هر پرس‌وجو در نظر گرفته می‌شود. زمان انتقال نیز بر اساس اندازهٔ نتایج و پهنای باند شبکه است به‌صورت معادلهٔ (۵) محاسبه می‌شود.

$$Trans\,(q,s) = \frac{Size\left(A^s_q\right)}{Net\,(S)} \qquad (۵)$$

در معادلهٔ (۵)، $Trans\,(q,s)$ هزینهٔ انتقال برای پرس‌وجوی q در کارپذیر S است. $Size(A^S_q)$ تعداد رکوردهایی است که در پاسخ به پرس‌وجوی q از کارپذیر S برگردانده می‌شود و $Net(S)$ پهنای باند مورد استفاده برای انتقال نتایج است که مقدار آن با بیشترین اندازه برای پرس‌وجوهای تخصیص داده‌شده به کارپذیر S تعریف می‌شود.

از آنجایی که هر پرس‌وجو به بیش از یک کارپذیر تخصیص داده می‌شود، اجرای آن پرس‌وجو به‌صورت موازی انجام می‌گیرد. بنابراین باید حداکثر زمان پاسخ برای هر پرس‌وجو در کارپذیران مختلف به دست آید که مقدار آن از طریق معادلهٔ (۶) محاسبه می‌شود:

$$T\,(q,A) = \max\left\{\theta_{s,q} + Trans\,(q,s)\right\} \qquad (۶)$$

به این ترتیب برای دنباله‌ای از پرس‌وجوهای $Q = \{q_1, \dots, q_m\}$ میانگین زمان پاسخ به‌صورت معادلهٔ (۷) تعیین

1. Hamming Distance

می‌شود:

$$T(Q,A)=\sum_{q\in Q}T(q,A).f_q \quad (۷)$$

منظور از $T(Q,A)$، زمان پاسخ به تمامی پرس‌وجوهاست.

۴.۲. تابع هدف

در ماتریس B که برای نمایش نتایج پرس‌وجوها ساخته می‌شود، عمل تخصیص اولیهٔ رکوردها و پرس‌وجوها به‌صورت تصادفی انجام می‌گیرد. اما باید به‌دنبال بهترین انتساب رکوردها و پرس‌و جوها به کارپذیران بود. برای این منظور، نیاز به یک قطعه‌بندی بهینه است. منظور از قطعه‌بندی بهینه، قطعه‌بندی است که علاوه بر تأمین محدودیت‌های محرمانگی و الگوی دسترسی، بتواند از نظر سربار پاسخ به پرس‌وجوها کیمنه باشد. برای یافتن قطعه بندی بهینه در روش پیشنهادی (بر اساس هریک از دو معیار یادشده یعنی حداقل زمان پاسخ به پرس‌وجو و افشای اطلاعات) بر مبنای معادلهٔ (۸) عمل می‌شود که در آن γ ضریبی برای D(A) است تا مقدار آن را مثبت نگه دارد [۱۰]:

$$\min_{A\in(B\times S)}\left(T(Q,A)+\gamma.D(A)\right) \quad (۸)$$

۵.۲. قطعه‌بندی بهینهٔ داده‌ها

در قطعه‌بندی بهینه داده‌ها سعی می‌شود تا انتساب رکوردها و پرس‌وجوها به کارپذیران با حداقل زمان پاسخ و افشای اطلاعات توصیف شود. برای این منظور، معادلهٔ (۸) می‌تواند مبنای کار قرار بگیرد. اما با توجه به استدلال ارائه‌شده در مرجع [۱۰] یافتن قطعه‌بندی بهینه، مسئله‌ای NP-Hard است لذا برای حل مسئلهٔ مذکور، از الگوریتم تقریبیِ منطبق با معیارهای ارائه‌شده استفاده می‌شود. در مرجع [۱۰] الگوریتم (۱) برای حل تقریبی این مسئله ارائه شده است.

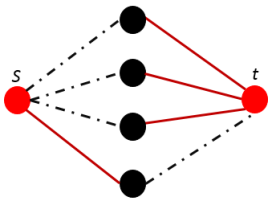
با هدف بهبود نتایج الگوریتم (۱) در روش پیشنهادی اصلاحی روی مراحل اجرایی ارائه شده است اما ابتدا به تشریح اجرای الگوریتم (۱) می‌پردازیم.

الگوریتم (۱) به‌دنبال بهینه‌سازی تابع هدف (G(A)) است. برای حل این مسئله، دوبه‌دوی کارپذیران (t, s) برای تخصیص پرس‌وجوهای B بر اساس ماتریس A انتخاب می‌شوند تا برای هر دو جفت کارپذیر، مسئلهٔ بهینه‌سازی تابع هدف به مسئلهٔ بهینه‌سازی محلی تبدیل شود. مقدار بهینهٔ محلی به‌ازای هر دو کارپذیر به دست می‌آید تا در نهایت، مقدار بهینهٔ عمومی برای کل کارپذیران محاسبه شود. پس در الگوریتم (۱)، مسئلهٔ بهینه‌سازی معادلهٔ (۸) برابر است با انتساب داده‌ها و پرس‌وجوهای ماتریس B به‌صورت $B=\{b_{1,1},\dots,b_{1,M},b_{N,1},\dots,b_{N,M}\}$ به کارپذیران $S=\{S_1,\dots,S_K\}$ به‌طوری که تابع هدف به حداقل مقدار ممکن برسد.

```
Input: matrix B; set of servers S; empty
assignment A; objective function G
Output: assignment A; G(A) is minimal
Initialization: A←an arbitrary assignment;
do
 stop←0;
 for all pairs of servers {s, t} ∈S do
  A'=exchange (s, t, A);
  if G(A') <G(A) then
   A= A';
   stop←1;
  elseif |G(A')- G(A)|<thr_d then
   A= A';
  end
 end
while stop == 1;
return A;
```

الگوریتم (۱): یافتن قطعه‌بندی بهینه [۱۰]

در فرایند تخصیص، برای هر درایهٔ موجود در B یعنی $x_{r,q}\in B$ به‌طور جداگانه تصمیم گرفته می‌شود. برای مثال، در شکل (۲)، دو کارپذیر به‌صورت گره‌های پایانی با نام‌های s و t آورده و گره‌های میانی به s و t متصل شده‌اند. خطوط پررنگ به‌معنای وجود آن عنصر در کارپذیری است که به آن متصل شده و خط‌چین به‌معنای تخصیص آن عنصر در تکرارهای بعدی است. در هر تکرار از الگوریتم (۱)، گره‌های میانی به یکی از جفت کارپذیران انتخابی متصل می‌شوند. هر گره میانی می‌تواند به یکی از گره‌های پایانی متصل شود به‌طوری که تابع هدف در حالت بهینه ممکن باشد. این مراحل برای هر جفت کارپذیر تکرار می‌شود تا زمانی که تخصیص بهینه شود.



شکل (۲): انتساب بهینهٔ عناصر به یک جفت کارپذیر

الگوریتم (۱) به‌صورت مکرر، یک حلقه را اجرا می‌کند تا زمانی که نتواند هیچ پیشرفتی در جهت بهبود تابع هدف، نسبت به آخرین انتساب به دست آورد. تابع exchange در الگوریتم (۱)، موجب جابه‌جایی یک عنصر از یک کارپذیر به کارپذیر دیگر می‌شود. زمانی متغیر stop به مقدار یک تنظیم می‌شود که خروجی exchange به A’ منتسب شود و با این انتساب، تابع هدف، مقدار کمتری داشته باشد. در این روش، حل مسئلهٔ بهینه‌سازی به مسئلهٔ برش کمینهٔ گراف نگاشت شده است. قبل از اجرای الگوریتم (۱)، گرافی با دو نوع گره ساخته می‌شود: گره‌های پایانی و گره‌های میانی. گره‌های پایانی به کارپذیران اشاره دارند و گره‌های میانی به تمام عناصر موجود در B. در مرحلهٔ اول الگوریتم (۱)، هر عنصر موجود در B به‌طور دلخواه به یکی از کارپذیران تخصیص داده می‌شود. اما از آنجا که این الگوریتم، یک الگوریتم حریصانه است، به‌شدت به تخصیص تصادفی اولیه وابسته است.

در روش پیشنهادی در جهت کاهش احتمال مشکل بهینه‌سازی محلی الگوریتم (۱)، مواردی شناسایی می‌شود که جابه‌جایی یک عنصر از یک کارپذیر به کارپذیر دیگر، مقدار تابع هدف را به میزان ناچیزی افزایش می‌دهد. ممکن است این جابه‌جایی به‌صورت توأم با جابه‌جایی یک یا چند عنصر دیگر، باعث بهبود تابع هدف به مقدار چشمگیر شود. بدین منظور در الگوریتم پیشنهادی برای کاهش این مشکل (بهینهٔ محلی)، از تابع شانس مجدد استفاده می‌کنیم. نمایشی از اصلاح مورد نظر در الگوریتم (۲) آورده شده است.

در الگوریتم (۲) تابع شانس مجدد (Pr(w)) به این صورت عمل می‌کند که اگر $|G(A') - G(A)|$ کمتر از یک حد آستانه باشد با احتمال Pr (که یک عدد بین صفر تا یک است) جابه‌جایی فعلی حفظ می‌شود و الگوریتم با وضعیت فعلی ادامه پیدا می‌کند. اما در صورتی که این جابه‌جایی، بهبودی در نتایج جابه‌جایی‌های بعدی در مقدار تابع هدف نداشته باشد، این وضعیت در تکرار بعدی به‌صورت خودکار نادیده گرفته می‌شود.

```
Input: matrix B; set of servers S; empty
assignment A; objective function G
Output: assignment A; G(A) is minimal
Initialization: A←an arbitrary assignment;
do
 stop←0;
 for all pairs of servers {s, t} ∈S do
  A’=exchange (s, t, A);
  if G(A’) <G(A) then
   A= A’;
   stop←1;
  elseif |G(A’)- G(A)|<thr_d and
      ω < Pr(ω ~ N (0,1)) then
   A= A’;
  end
 end
while stop == 1;
return A;
```

الگوریتم (۲): یافتن قطعه‌بندی بهینهٔ پیشنهادی

۳. ارزیابی روش پیشنهادی

در این بخش، جزئیات مربوط به نتایج شبیه‌سازی روش پیشنهادی بر اساس الگوریتم (۲) آورده شده است. نتایج به‌دست‌آمده نشان می‌دهد که روش پیشنهادی در دستهٔ برون‌سپاری داده‌ها تا چه میزان می‌تواند نسبت به روش‌های دیگر در ارتقای امنیت داده‌ها مفید عمل کند. منظور از ارتقای امنیت این است که روش پیشنهادی بتواند به طور قابل توجهی افشای اطلاعات، زمان پاسخ، سربارهای محاسباتی و ارتباطی و ذخیره‌سازی را کاهش دهد. بدین منظور در کنار روش پیشنهادی، روش‌های به‌روز دیگری، از منابع و مراجع معتبر انتخاب شده‌اند تا از بابت هریک از معیارهای مطرح‌شده با روش پیشنهادی مورد مقایسه قرار گیرند.

۱.۳. شرایط پیاده‌سازی و اجرای روش‌ها

مجموعه دادهٔ انتخابی برای انجام شبیه‌سازی و ارزیابی نتایج، مربوط به حوزهٔ داده‌های سلامت الکترونیک (EHR) است. به

این ترتیب فرض می‌شود سوابق پزشکی بیمار با عنوان EHR در کارپذیران موجود در نهادهای مختلف ذخیره می‌شوند تا همیشه و همه‌جا در دسترس قرار گیرند.

باید به این نکته توجه داشت که حفظ محرمانگی داده‌ها و حفاظت از الگوی دسترسی به‌عنوان یکی از چالش‌های اصلی حوزهٔ سلامت الکترونیک مطرح می‌باشد لذا برآورده کردن شرایط محرمانگی بهتر برای داده‌های EHR بسیار ارزشمند است [۱۱].

پیاده‌سازی روش پیشنهادی به‌صورت آزمایشگاهی و بر اساس استفاده از داده‌های شبیه‌سازی‌شده انجام شده است. این داده‌ها در ایالات متحدهٔ آمریکا تولید شده و یک رابطهٔ پزشکی درمانی با ۵۳ صفت با اندازه‌های فیزیکی متفاوت و شامل تعداد ۶۷۶۳۴۲ رکورد است. پرس‌وجوهای تحلیلی مورد استفاده در این پیاده‌سازی، از بین ۱۹ محور بیان‌شده در SNOMED-CT انتخاب شده‌اند [۱۲]. هدف از انتخاب این پرس‌وجوها، به‌دست‌آوردن اطلاعات آماری و یافته‌های بالینی در حوزهٔ سلامت است.

۲.۳. مقایسهٔ روش‌ها

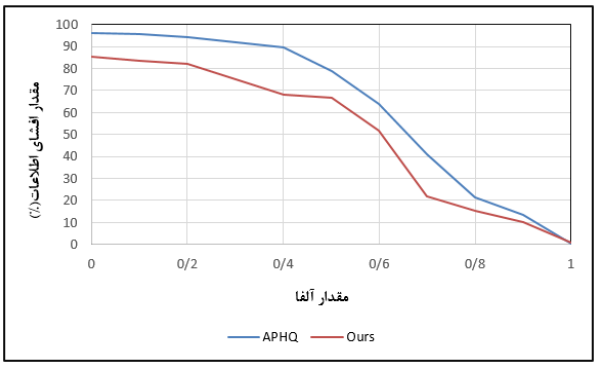
همان طور که در مقدمه عنوان شد، با دسته‌بندی روش‌های فعلی که در زمینهٔ حفاظت از الگوی دسترسی مطرح‌اند، دو دستهٔ عمومی برای این روش‌ها می‌توان در نظر گرفت. دستهٔ اول صرفاً یک توزیع از قبل تعیین‌شده و ایستا برای داده‌ها بر روی چندین کارپذیر ارائه می‌کنند و دستهٔ دوم سعی در ایجاد مکانیزم پویاتری برای توزیع داده‌ها بر اساس شرایط دسترسی به داده‌ها دارند. ازاین‌رو برای مقایسهٔ روش پیشنهادی با آخرین تکنیک‌های ارائه‌شده برای هر دو دسته، روش‌های Hashedbased [۱۳] و Rangedbased [۳] از دستهٔ اول با توزیع ایستای داده‌ها و روش APHQ [۱۰] از دستهٔ دوم با توزیع پویای داده‌ها برای مقایسه انتخاب شده‌اند. در ادامه، مقایسهٔ روش‌های فوق با روش پیشنهادی، بر مبنای معیارهای افشای اطلاعات، زمان پاسخ، تعداد کارپذیران و از طرفی دیگر بر اساس سربارهای محاسباتی، ذخیره‌سازی و ارتباطی آمده است. با توجه به اینکه در تمام روش‌ها، مالک داده و کارخواه معتمد عملاً به‌عنوان یک مؤلفه تلقی می‌شوند، بنابراین در ادامهٔ این بخش، مالک داده‌ها و کارخواه معتمد تحت یک مؤلفه با عنوان کارخواه معتمد معرفی می‌شوند.

سربارها برای انجام آزمون به‌صورت زیر تعریف شده‌اند:

- سربار محاسباتی: به هر پردازش اضافی که از سمت مالک داده‌ها جهت تأمین محرمانگی دسترسی به بخش‌های مختلف مانند مالک داده‌ها، کارخواه و کارپذیر تحمیل گردد، اشاره دارد.
- سربار ذخیره‌سازی: هر فضای اضافی فراتر از ذخیره‌سازی داده‌های واقعی، به‌عنوان سربار ذخیره‌سازی در نظر گرفته می‌شود.
- سربار ارتباطی: هر ارتباط اضافی بین «مالک داده‌ها و کارپذیر» و «کارخواه و کارپذیر» صورت گیرد، به‌عنوان سربار ارتباطی در نظر گرفته می‌شوند.

۳.۳. افشای اطلاعات

دربارهٔ میزان افشای اطلاعات، روش پیشنهادی با روش APHQ مقایسه شده است. این مقایسه بر اساس متغیر α در محدودهٔ صفر تا یک و با درنظرگرفتن سه کارپذیر انجام گرفته است. از آنجایی که مقدار α بر روش‌های Hashedbased و Rangedbased تأثیری ندارد، این دو روش در این مقایسه آورده نشده‌اند. نتیجهٔ مقایسه در شکل (۳) آورده شده است.



شکل (۳): مقایسه بر اساس معیار افشای اطلاعات

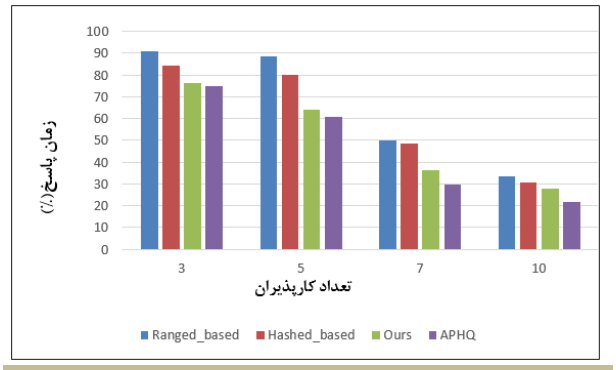
با توجه به نمودار شکل (۳)، روش پیشنهادی مقدار افشای اطلاعات کمتری را به‌ازای هر α داشته است. این بدین معنی

است که روش پیشنهادی تخصیص‌های بهتری نسبت به روش APHQ به کارپذیران انجام داده است.

۴.۳. زمان پاسخ

زمان پاسخ به‌عنوان یکی از معیارهای اصلی در روش پیشنهادی مطرح شده است. در این بخش، چهار روش مذکور از نظر معیار زمان پاسخ مورد مقایسه قرار گرفته‌اند. به منظور مقایسهٔ همهٔ روش‌ها، در این آزمون مقدار α ثابت و برابر ۰/۵ در نظر گرفته شده و زمان پاسخ‌گویی بر اساس تعداد کارپذیران به تفکیک روش‌های مختلف ارائه شده است. این مقایسه در شکل (۴) قابل مشاهده است.

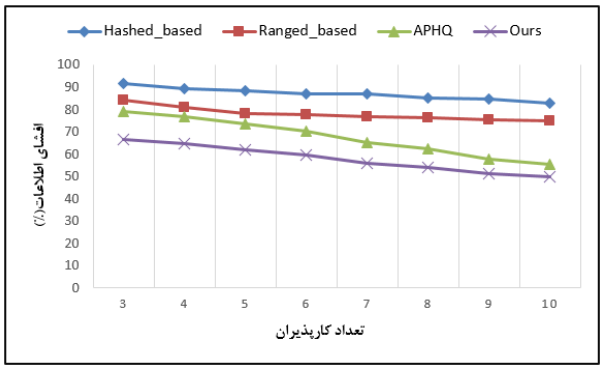
زمان پاسخ در همهٔ روش‌ها با افزایش تعداد کارپذیران کاهش می‌یابد. روش APHQ عملکرد بهتری نسبت به سه روش دیگر در این آزمون داشته است. بعد از این روش، روش پیشنهادی، سپس روش Hashedbased و در نهایت روش Rangedbased مقدار زمان پاسخ کمتری را داشته‌اند. با توجه به اینکه هر پرس‌وجو به‌صورت موازی در چندین کارپذیر اجرا می‌شود، با افزایش تعداد کارپذیران، اجرای موازی بین کارپذیران بیشتر شده و در نتیجه زمان پاسخ کاهش می‌یابد.



شکل (۴): مقایسه بر اساس معیار تعداد کارپذیران- زمان پاسخ

۵.۳. تعداد کارپذیران- افشای اطلاعات

هر چهار روش تحت آزمون، دارای افشای اطلاعات هستند. به‌منظور مقایسهٔ همهٔ روش‌ها، در این آزمون مقدار α ثابت و برابر ۰/۵ در نظر گرفته شده و میزان افشای اطلاعات هریک از روش‌ها به‌ازای تعداد کارپذیران مختلف بررسی شده است. این مقایسه در شکل (۵) قابل مشاهده است.



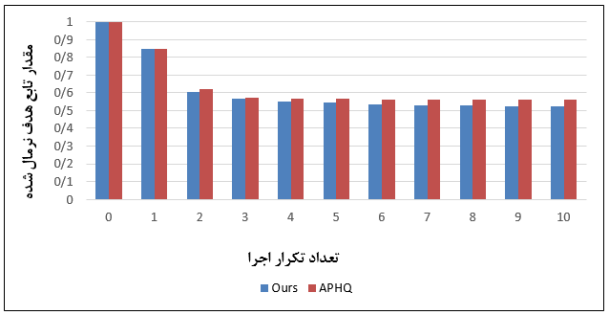
شکل (۵): مقایسه بر اساس تعداد کارپذیران- افشای اطلاعات

همان طور که در شکل (۵) مشاهده می‌شود، افشای اطلاعات همهٔ روش‌ها با افزایش تعداد کارپذیران کاهش می‌یابد و روش پیشنهادی نسبت به سه روش دیگر، عملکرد بهتری داشته است. بعد از روش پیشنهادی، به‌ترتیب روش APHQ، روش Rangedbased و در نهایت روش Hashedbased مقدار افشای اطلاعات کمتری دارند. علت اینکه با افزایش تعداد کارپذیران، افشای اطلاعات کاهش می‌یابد، این است که هریک از کارپذیران بخشی از داده‌ها را در خود نگهداری کرده و الگوی دسترسی کمتری را مشاهده می‌کنند. در روش پیشنهادی و روش APHQ تعداد پرس‌وجوها و رکوردهایی که به‌عنوان پاسخ به آن پرس‌وجوها در نظر گرفته شده‌اند، در تعداد بیشتری کارپذیر توزیع می‌شوند. این امر سبب شده است فرکانس کلمات کلیدی که هر کارپذیر مشاهده می‌کند، نادرست باشد.

۶.۳. تعداد تکرار اجرا

دو روش پیشنهادی و APHQ بر مبنای معیار تکرار اجرا نیز مورد مقایسه قرار گرفته‌اند. این آزمون با مقدار α ثابت و برابر ۰/۵ و برای سه کارپذیر انجام گرفته است. این مقایسه در شکل (۶) قابل مشاهده است.

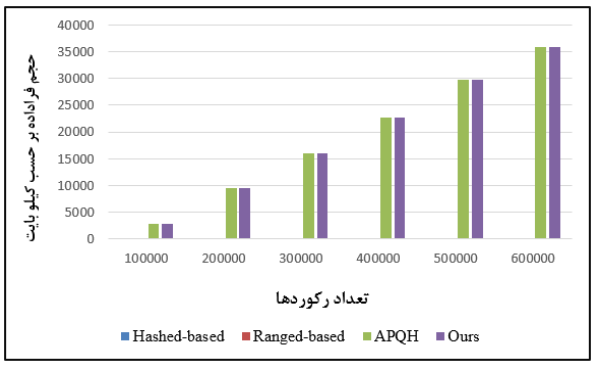
طبق شکل (۶)، روش پیشنهادی سریع‌تر به جواب بهینه نزدیک می‌شود و بعد از چند تکرار روش پیشنهادی عملکرد بهتری نسبت به روش APHQ نشان می‌دهد. در روش پیشنهادی، به‌دلیل استفاده از تابع افشای اطلاعات بهتر نسبت به روش APHQ، مقدار تابع هدف کمتری حاصل شده است.



شکل (۶): مقایسه بر اساس معیار تعداد تکرار اجرا

۷.۳. سربار ذخیره‌سازی

سربار ذخیره‌سازی به دو بخش سربار ذخیره‌سازی در کارپذیران و کارخواه معتمد تقسیم می‌شود. در روش‌هایی که برای مقایسه انتخاب شده‌اند، هیچ سربار ذخیره‌سازی برای کارپذیران وجود ندارد؛ زیرا فقط داده‌های اصلی به‌صورت رمزنگاری‌شده در کارپذیران ذخیره می‌شوند. با توجه به استفاده از رویکرد برون‌سپاری کامل داده‌ها، انتظار می‌رود هیچ سربار ذخیره‌سازی داده در سمت کارخواه معتمد وجود نداشته باشد. اما در برخی از روش‌های مورد مقایسه، به دلایل مختلفی سربار ذخیره‌سازی در کارخواه معتمد وجود دارد. کمینه کردن و به صفر رساندن این سربار یکی از نیازهای اصلی در رویکرد برون‌سپاری کامل داده‌هاست. در شکل (۷)، سربار ذخیره‌سازی داده‌ها روش‌ها در کارخواه معتمد با یکدیگر مقایسه شده‌اند (فقط دو روش پیشنهادی و APQH دارای سربار ذخیره‌سازی هستند).



شکل (۷): مقایسه بر اساس معیار سربار ذخیره‌سازی

حجم جدول بیماران قبل از برون‌سپاری کامل در حدود ۸۵۵۲۱ کیلوبایت است. منظور از فضای ذخیره‌سازی کل، میزان کل حجم جدول جدید برای برون‌سپاری است. همچنین فضای ذخیره‌سازی سمت کارخواه معتمد، میزان داده‌ای است که باید در سمت کارخواه معتمد نگهداری شود. با توجه به اطلاعاتی که در شکل (۷) ارائه شده است، روش‌های Rangedbased و روش Hashedbased سربار ذخیره‌سازی در کارخواه معتمد ندارند. پس این روش‌ها به‌عنوان روش‌های بهینه از نظر معیار سربار ذخیره‌سازی تلقی می‌شوند. روش پیشنهادی و روش APQH این سربار را به میزان کمی دارا هستند.

سربار ذخیره‌سازی روش APQH نیز همانند روش پیشنهادی است که مربوط به فراداده است. این فراداده از وضعیت کارپذیران، ماتریس تعیین سرور و جدول مدیریت کلید جدول بیماران تشکیل شده است.

۸.۳. سربار محاسباتی

به‌دلیل قابل دسترس بودن حجم‌های بالای ذخیره‌سازی به‌ویژه در رویکرد برون‌سپاری، سربارهای ذخیره‌سازی اهمیت زیادی ندارند اما نیاز به کاهش سربارهای محاسباتی در کارخواه متعمد و کارپذیران به‌منظور دستیابی به توان اجرایی بیشتر، اهمیت ویژه‌ای دارد. به‌طور کلی، سربار محاسباتی به دو بخش سربار محاسباتی کارپذیران و کارخواه معتمد تقسیم می‌شود. در این بخش سعی شده است سربار محاسباتی روش پیشنهادی با سه روش دیگر مورد مقایسه قرار گیرد. نتایج این مقایسه در جدول (۱) قابل مشاهده است.

همان طور که در جدول (۱) مشاهده می‌شود، روش پیشنهادی هیچ سربار محاسباتی در سمت کارخواه معتمد و کارپذیران ندارد و از عملکرد بهتری نسبت به سه روش دیگر برخوردار است. در روش APHQ با توجه به وضعیت ذخیره‌شدهٔ هر کارپذیر در سمت کارخواه معتمد، مشخص می‌شود که هر پرس‌وجو برای دریافت پاسخ باید به کدام کارپذیران ارسال شود. اما در روش پیشنهادی، ماتریسی برای تعیین کارپذیر جهت پاسخ‌گویی به پرس‌وجوها ساخته شده است. سطرهای این ماتریس نشان‌دهندهٔ کارپذیران و ستون‌های آن نشان‌دهندهٔ پرس‌وجوهاست. این ماتریس در قسمت فرادادهٔ سمت کارخواه معتمد ذخیره می‌شود. در روش پیشنهادی، طبق

رویکرد بیان‌شده برای انتخاب کارپذیران جهت پاسخ‌گویی به پرس‌وجوها، این سربار ناچیز و قابل چشم‌پوشی است.

نام روش	سربار محاسباتی کارپذیران	سربار محاسباتی کارخواه معتمد
روش *APHQ*	×	✓
روش پیشنهادی	×	×
روش *Range-based*	✓	×
روش *Hashed-based*	✓	×

جدول (۱): مقایسهٔ روش‌ها بر اساس معیار سربار محاسباتی

در روش‌های Rangedbased و Hashedbased سربار محاسباتی بالایی روی کارپذیران وجود دارد؛ زیرا توزیع رکوردها به کارپذیران بر اساس پرس‌وجوها انجام نمی‌گیرد، بلکه با استفاده از شمارهٔ هر رکورد انجام می‌شود. بنابراین در هر دو روش، پرس‌وجوها به تمامی کارپذیران ارسال می‌شود تا پاسخ نهایی دریافت شود. این کار سبب ایجاد سربار محاسباتی برای کارپذیرانی که برای پاسخ به آن پرس‌وجو مورد نیاز نبوده‌اند، خواهد شد.

۹.۳. سربار ارتباطی

از آنجایی که در هر چهار روش مذکور به‌ازای هر پرس‌وجو فقط یک ارتباط بین کارخواه معتمد و کارپذیران ایجاد می‌شود، هیچ سربار ارتباطی بین آن‌ها وجود ندارد. با توجه به اینکه چندین کارپذیر در این سناریو برای ذخیره‌سازی داده‌ها در نظر گرفته شده‌اند، اجرای هر پرس‌وجو به‌صورت موازی در کارپذیران انجام می‌گیرد که هیچ سربار ارتباطی نخواهد داشت.

۴. جمع‌بندی و نتیجه‌گیری

برون‌سپاری داده‌ها در کارپذیر خارجی، نیازمندی‌های جدیدی برای تأمین امنیت اطلاعات دارد که محرمانگی داده‌ها و حفاظت از الگوی دسترسی از مهم‌ترین آن‌ها هستند. در این مقاله، جدیدترین روش های موجود برای تأمین محرمانگی داده‌ها و حفاظت از الگوی دسترسی معرفی شدند. برای بهبود عملکرد این روش‌ها در حفاظت از ویژگی‌های امنیتی مدنظر، به‌کارگیری مکانیزم‌هایی برای حفاظت از الگوی دسترسی مورد نیاز است. مکانیزم‌های قدیمی فقط توزیع ایستایی از داده‌ها را روی چند کارپذیر پیشنهاد می‌کنند اما در این مقاله رویکرد پویایی معرفی شد که مکانیزم شانس مجدد را در توزیع داده‌ها و پرس‌وجوها روی چند کارپذیر به کار می‌گیرد تا بهینه‌سازی بهتری ارائه کند. روش ارائه‌شده دراین مقاله با بهبود نحوهٔ توزیع داده‌ها بر روی کارپذیران خارجی، موجب کاهش افشای اطلاعات و ارتقای محرمانگی داده‌ها بدون ایجاد سربارهای اضافی می‌شود. البته هنوز امکان افشای اطلاعات دیده می‌شود که هدف بعدی کار در این زمینه، کاهش افشای اطلاعات به حداقل میزان ممکن است.

مراجع

[1] Salinas, D., Basnight, M. and Morris, D., "*Providing a database as a service in a multi-tenant environment*", U.S. Patent Application No 16/113,591, 2019.

[2] Wei, X., Minghao, Z. and Qiuliang, X., "*Efficient and secure outsourced approximate pattern matching protocol*", Soft Computing, Vol. 22, No. 4, 2018.

[3] Kanwal, T., Abdul Jabbar, A., Anjum, A. and Malik, S., "*Privacy-aware relationship semantics–based XACML access control model for electronic health records in hybrid cloud*", International Journal of Distributed Sensor Networks, Vol.15, No. 6, 2019.

[4] Amalarethinam, G. and Rajakumari, S., "*A Survey on Security Challenges in Cloud Computing*", Journal of Physical Sciences, Vol. 24, No. 1, 2019.

[5] Kellaris, G., "*Generic attacks on secure outsourced databases*", Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, ACM, 2016.

[6] Tang, J., "*Ensuring security and privacy preservation for cloud data services*", ACM Computing Surveys, Vol. 49, No. 1, 2016.

[7] Xiao, Yinhao, *"Edge Computing Security: State of the Art and Challenge*", Proceedings of the IEEE, 2019.

[8] Aggarwal, G., "*Two can keep a secret: A distributed architecture for secure database services*", CIDR, 2005.

[9] Di, V. and Sabrina, D., "*Distributed shuffling for preserving access confidentiality*", European Symposium on Research in Computer Security, Berlin, 2013.

[10] Dou, Y. and Chan, H.C., "*Access Pattern Hidden Query*

over Encrypted Data through Multi-Clouds", In IEEE Global Communications Conference (pp. 1-6), 2017.

[11] Kumekawa, J.K., "*Health information privacy protection: crisis or common sense*", Online Journal of Issues in Nursing, Vol. 6, No. 3, 2001.

[12] Bodenreider, O., "*Comparing SNOMED CT and the NCI Thesaurus through Semantic Web Technologies*", 3rd international conference on Knowledge Representation in Medicine, 2008.

[13] Rahim, N., Ahmad, J., Muhammad, K., Sangaiah, A.K. and Baik, S.W., "*Privacy-preserving image retrieval for mobile devices with deep features on the cloud*", Computer Communications, Vol. 127, pp. 75-85, 2018.