طرح جستجوی کلیدواژه فازی قابل تصدیق بر روی پایگاه داده رمز شده در رایانش ابری مبتنی بر خوشه بندی کلمات

یحیی دهقانیان[۱]، مجید غیوری ثالث[۲]، علیرضا رحیمی[۳]
[۱]دانشجوی دکتری دانشگاه امام حسین(ع) ydehghaniyan@ihu.ac.ir
[۲و۳]استادیار دانشگاه امام حسین(ع)

چکیده

در سال‌های اخیر رویکرد برون‌سپاری پایگاه داده و انتقال زیرساخت‌ها در رایانش ابری از طرف سازمان‌ها و کاربران مورد توجه فراوان قرار گرفته است. مالکان پایگاه داده به ارائه دهندگان سرویس و تامین کنندگان زیرساخت از نظر صحت نگه-داری و دسترس‌پذیری، اطمینان دارند و دغدغه اصلی آن‌ها حفظ محرمانگی اطلاعات حساس است، به همین منظور پایگاه داده را به صورت رمزشده در سرورهای ابری ذخیره می‌کنند. این نوع ذخیره‌سازی، مالکان داده را با یک چالش اصلی هنگام جستجو و بازیابی اطلاعات از پایگاه داده رمزشده روبرو می‌کند. گرچه طرح‌های رمزگذاری قابل جستجو موجود به کاربر اجازه می‌دهد تا داده‌های رمزنگاری شده را با قابلیت اطمینان بالا جستجو و بازیابی کنند، اما اغلب این راه‌حل‌ها به تنهایی قابل اعتماد نیستند و نیاز به تصدیق نتیجه جستجو دارند؛ زیرا ممکن است سرور ابری برای حفظ قابلیت محاسباتی و یا صرفه‌جویی در پهنای باند خود، بخشی از جستجو را انجام داده و نتایج جستجو را کامل و صحیح در اختیار کاربر قرار ندهد. به همین منظور، طرح‌های مختلف رمزنگاری قابل جستجو که نتایج بازگشتی را بررسی و تائید می‌کنند، ارائه شده است. این طرح‌ها به طور کلی علاوه بر الگوریتم‌های مورد نیاز محرمانگی، از یک الگوریتم مستقل برای تصدیق نتایج استفاده می‌کنند که باعث افزایش حجم پردازش می‌گردد. در این مقاله ما یک طرح جستجوی کلیدواژه قابل تصدیق بر روی پایگاه داده رمزشده ارائه می‌کنیم که رمزنگاری و تصدیق نتایج آن با استفاده از روش الگوریتم رمزنگاری توام با احراز اصالت، انجام می‌گیرد. این طرح علاوه بر جستجوی کلیدواژه صحیح، امکان جستجو کلیدواژه با تحمل خطا به صورت فازی را دارد. آزمایشات صورت گرفته بر روی مجموعه داده‌های مختلف نشان می‌دهد که طرح پیشنهادی علاوه بر تصدیق نتایج جستجو، حجم فراداده‌های ذخیره شده را به نصف کاهش داده، و سرعت جستجوی فازی را حداقل به دو برابر افزایش می‌دهد. هم‌چنین طرح پیشنهادی در مقابل تهدیدات در نظر گرفته امن بوده و نتایج را به صورت صحیح و کارآمد بازیابی می‌کند.

کلیدواژه

رایانش ابری، برون‌سپاری پایگاه داده، رمزنگاری قابل جستجو، جستجوی فازی، رمزنگاری توام با احراز اصالت

تاریخ دریافت: ۱۳۹۹/۰۲/۱۴ تاریخ پذیرش: ۱۳۹۹/۰۷/۲۷

مقدمه

با رایج شدن رایانش ابری، برون‌سپاری پایگاه داده به علت کاهش هزینه به صورت گسترده مورد استفاده قرار گرفته است. با برون‌سپاری، مالک پایگاه داده سرویس‌هایی مانند ذخیره، پشتیبان‌گیری و عملیات جستجو را به سرور ابری واگذار می‌کند. بدلیل اینکه پایگاه داده برون‌سپاری شده با چند کاربر به اشتراک گذاشته می‌شود و کاربران به طور جداگانه نیاز به دسترسی به بخش‌های مختلفی از پایگاه داده دارند. از آنجا که سرور ابری در دامنه‌های مورد اعتماد متفاوت با مالک داده قرار دارد، بنابراین داده‌ها از کنترل فیزیکی مالک داده خارج می‌شوند. مالکان پایگاه داده باید قادر به بررسی صحت ذخیره‌سازی در ابر باشند[۳-۱]. علاوه بر این، داده‌های ذخیره شده ممکن است توسط سرور ابر بداندیش[۱] و یا کاربران غیرمجاز مورد سوء استفاده قرار گیرند، که منجر به افشای داده‌های شخصی مالک پایگاه داده می‌شود.

یکی از راهکارها برای حفظ محرمانگی پایگاه داده در رایانش ابری، ذخیره به صورت رمزشده است که در این صورت جستجوی اطلاعات برای کاربران با استفاده از روش‌های جستجوی سنتی، قابل انجام نیست. یک راه‌حل ساده برای این مساله دریافت تمامی داده‌های رمزشده از سرور ابری و رمزگشایی به صورت محلی می‌باشد. این روش به پهنای باند، فضای ذخیره‌سازی و سربار محاسباتی زیادی نیاز دارد و در عمل قابل اجرا نیست. راه‌حل دیگر این است، که کلیدهای رمزگشایی در اختیار سرور ابری قرار گیرد و سرور ابری داده‌ها را رمزگشایی کرده و سپس عملیات جستجو را انجام دهد. بدیهی است که در این روش سرور ابری غیرقابل اعتماد به اصل داده‌ها دسترسی خواهد داشت و محرمانگی اطلاعات که هدف اصلی رمزگذاری بوده، به مخاطره می‌افتد. برای حل این مسئله، طرح‌های مختلف رمزنگاری قابل جستجو (SE[۲]) ارائه شده است. این طرح‌ها به کاربران اجازه می‌دهد تا به طور انتخابی اطلاعات رمزشده روی سرور ابری را با جستجوی کلمه کلیدی مورد نظر بازیابی کنند. روش‌های رمزنگاری قابل جستجو را می‌توان به دو دسته کلی تقسیم کرد؛ دسته اول جستجو روی خود داده‌های رمز شده است که می‌توان به طرح‌هایی مانند روش سانگ و همکاران[۴]، رمزنگاری هم-ریخت[۵]، رمزنگاری مبتنی بر حفظ ترتیب[۳] [۶] و رمزنگاری مبتنی بر تسهیم‌راز[۴] [۷] اشاره کرد. دسته دوم عملیات جستجو را با ساختن دریچه‌های[۵] مرتبط با داده‌های اصلی انجام می‌دهد[۹-۸]. با توجه به ملاحظات رایانش ابری از نظر حجم پردازش، پهنای باند، نوع کاربرد و نوع داده، دسته دوم برای برون‌سپاری امن پایگاه داده مناسب می‌باشد و طرح‌های بیشتری در این زمینه ارائه شده است. در این روش‌ها در زمان برون‌سپاری، کلمات کلیدی استخراج شده و بعد از شاخص-دهی به همراه داده‌های اصلی به صورت رمز شده، برون‌سپاری می‌شود. این داده‌های رمزشده به عنوان فراداده[۶] در عملیات جستجو و بازیابی مورد استفاده قرار می‌گیرد. بدین صورت که در هنگام بازیابی، ابتدا جستجو در داخل فراداده انجام گرفته و بعد از انطباق الگوی جستجو مطابق شاخص‌دهی انجام گرفته، اطلاعات رمزشده استخراج و به کاربر ارسال می‌شود. در اینجا هر چه تعداد کلمات کلیدی از پیش تعریف شده بیشتر باشد، احتمال موفقیت پرس‌وجو بالا می‌رود؛ زیرا شاخص‌ها به صورت رمز شده ذخیره می‌شوند و کاربر بایستی کلیدواژه را عیناً وارد کند و در صورت خطا، نتیجه پرس‌وجو موفق نخواهد بود، و برای جبران خطای ورودی کاربر، پرس‌وجو روی تمامی حالت‌های یک کلمه باید صورت بگیرد که این روال سرعت پرس‌وجو را به شدت کاهش می‌دهد. یک روش بهبود و افزایش سرعت، جستجوی بر مبنای شباهت[۷] کلمات و یا جستجوی فازی[۸] می‌باشد[۱۰] در این روش هم نیاز است که مجموعه کلیدواژه فازی کلمات، به صورت رمزشده برون‌سپاری شود. که در مقیاس بالا، این روش هم باعث افزایش حجم فراداده ذخیره شده و کاهش سرعت جستجو می‌گردد. یکی از روش‌ها برای کاهش فراداده، استفاده از روش خوشه‌بندی کلمات کلیدی است[۱۱]. به این معنا که به جای ذخیره مجموعه کلیدواژه فازی تمامی کلمات، ابتدا آن‌ها را خوشه-بندی کرده و سپس مجموعه کلیدواژه فازی مراکز خوشه‌ها و شاخص‌ها را به صورت رمزشده برون‌سپاری کنیم.

طرح‌های مختلفی که به صورت امن و کارا و بدون رمزگشایی عملیات جستجو را انجام می‌دهد وجود دارد؛ در آنها فرض می‌شود که سرور امین ولی کنجکاو[۹] است. این بدین معنی که سرور ابر از پروتکل تبعیت می‌کند ولی تلاش می‌کند که اطلاعات مخفی را از دارایی‌های که در اختیار دارد جمع‌آوری کند. از طرفی ممکن است سرور ابری، جستجو را کامل انجام ندهد و یا به جهت صرفه‌جویی در پهنای باند نتایج را درست و

[1] Malicious
[2] Searchable Encryption
[3] Order Preserving Encryption Scheme
[4] Secret Sharing

5 Trarpdoors
[6] Metadata
[7] Like
[8] Fuzzy
[9] honest-but-curious

کامل برنگرداند، به همین علت نیاز به بررسی و تصدیق[1] نتایج جستجو داریم. در سال‌های اخیر طرح‌های مختلف رمزنگاری قابل جستجو با قابلیت تصدیق نتایج، ارائه شده است. این طرح‌ها معمولاً از یک الگوریتم جداگانه‌ای علاوه بر الگوریتم‌های رمزنگاری مانند توابع MAC[2] استفاده می‌کنند، که باعث افزایش حجم پردازش می‌گردد[۱۲].

در این مقاله، هدف ارائه یک طرح رمزنگاری قابل جستجو فازی با قابلیت تصدیق نتایج است. بدین منظور از الگوریتم رمزنگاری توام با احراز اصالت جهت حفظ محرمانگی داده و تصدیق نتایج، از الگوریتم "نشانه عام[3]" برای تولید مجموعه کلیدواژه فازی[۱۰] و از طرح خوشه‌بندی کلمات کلیدی[۱۱] برای کاهش حجم فراداده، استفاده شده است. در اینجا با توجه به بکارگیری یک الگوریتم برای هر دو منظور، علاوه بر تصدیق نتایج پرس‌وجو، سرعت جستجو افزایش پیدا می‌کند که در کاهش هزینه‌ها هم موثر است.

در ادامه مقاله، در بخش دوم کارهای پیشین مرتبط را بیان می‌کنیم. برخی مفاهیم پایه و مقدمات پیش‌نیاز را در بخش سوم بررسی خواهیم کرد. در بخش چهارم، توضیح دقیق پروتکل و الگوریتم‌های پیشنهادی و هچنین مراحل آن‌ها به طور کامل ارائه خواهد شد. در بخش پنجم، ضمن تحلیل امنیتی طرح پیشنهادی، آن را با سایر طرح‌های مشابه مورد مقایسه قرار می‌دهیم و در نهایت در بخش آخر نتیجه و جمع‌بندی ارائه خواهد شد.

کارهای پیشین

در زمینه رمزنگاری قابل جستجو مطالعات زیادی انجام گرفته است. سانگ[4] و همکاران[۴] برای اولین بار در سال ۲۰۰۰، طرح رمزگذاری متقارن قابل جستجو را ارائه کردند. در این طرح، یک ساختار رمزنگاری دو لایه ویژه برای رمزگذاری هر کلمه کلیدی ساخته می‌شود و در فاز جستجو سرور ابری تمام اسناد را به صورت ترتیبی پویش می‌کند. بدین‌ترتیب زمان جستجو با اندازه مجموعه اسناد رابطه خطی دارد. در سال ۲۰۰۴، طرح[۱۳] استفاده از فیلترهای بلوم و توابع شبه‌تصادفی را برای ساخت شاخص‌های امن ارائه کرد. ۳ سال بعد کورت‌مولا و همکارانش یک طرح امن و کارآمد، ارائه کردند[۱۴]. این طرح از چند کاربر برای ارسال درخواست جستجو پشتیبانی می‌کند و هزینه جستجو متناسب با تعداد اسناد حاوی کلمات کلیدی مورد نظر است. در زمینه رمزنگاری قابل جستجو مطالعات زیادی انجام گرفته است، البته اغلب طرح‌ها بر روی بهبود کارایی، فرمول‌سازی و امنیت تمرکز دارند و بر اساس کلیدواژه دقیق[5] جستجو می‌کنند که به عنوان نمونه خوب از این طرح‌ها می‌توان به طرح‌های[۱۷-۱۵] اشاره کرد. طرح‌های جستجوی با کلیدواژه ثابت نمی‌تواند نتیجه مورد انتظار را در حالتی که کلمه ورودی دارای خطا باشد، باز گرداند و این موضوع کارایی سیستم را به شدت کاهش می‌دهد. برای حل این مشکل، لی[6] و همکاران[۱۰] طرح جستجوی کلیدواژه فازی را ارائه کردند. این طرح از روش "نشانه عام" برای ساخت مجموعه‌های کلیدواژه فازی، و از معیار فاصله ویرایش[7] برای اندازه‌گیری شباهت دو کلمه استفاده می‌کند. کوزو و همکاران[۱۸] از توابع LSH[8] برای تولید شاخص فایل جهت جستجوی سریع شباهت استفاده کردند. وانگ[9] و همکاران[۱۹] طرح جستجوی کلیدواژه فازی چند کلمه‌ای را براساس فیلتر بلوم[10] و تابع LSH پیشنهاد کردند که در آن کلمات کلیدی بر اساس مجموعه دو-گرم[11] ساخته می‌شود.

روش‌های جستجوی فازی کلمات رمزشده، در فضای ابری به نتایج قابل قبولی دست یافته است، اما کارایی جستجو، هزینه ذخیره‌سازی، دقت، و سازگاری این روش‌ها همچنان به تحقیقات بیشتری نیاز دارد. طرح‌های مختلفی برای بهبود هر کدام از پارامترهای فوق ارائه شده است. لیو و همکاران[۲۰] با استفاده از روش مبتنی بر لغت‌نامه[12]، چوا و همکاران[۲۱] با معرفی یک نمایه ساختار درخت بهبود یافته و لی و همکاران[۲۲] با استفاده از جستجو درخت پیشوندی مبتنی بر نماد، طرح[۱۰] را بهبود بخشیدند. مرجع[۱۱] از روش خوشه‌بندی[13] بر اساس سنجه فاصله ویرایش، ماهاجان و همکاران[۲۳] از روش خوشه‌بندی سلسله‌مرتبی بر اساس حداکثر انتظار[14] و مرجع[۱۲] از روش تخصیص یک بردار به کل مجموعه کلیدواژه فازی، برای کاهش حجم محاسبه و فضای ذخیره‌سازی استفاده کرده‌اند.

با توجه به این‌که ممکن است سرور ابری، جستجو را به‌طور کامل انجام ندهد و یا نتایج را درست و کامل برنگرداند، و یا حتی به صورت کامل ذخیره نکند، طرح‌های مختلفی برای بررسی و تائید نتایج برگشتی ارائه شده است[۲۶-۲۴]. وانگ و همکاران[۲۷] یک طرح جستجوی قابل تایید را با استفاده از زنجیره درهم‌سازی برای تایید نتیجه جستجو پیشنهاد دادند.

[1] Verifiable
[2] Message Authentication Code
[3] Wildcard
[4] Song
[5] Exact Keyword
[6] Li
[7] Edit Distance
[8] Locality Sensitive Hashing
[9] Wang
[10] Bloom Filter
[11] Bi-Gram
[12] Dictionary-based
[13] Clustering
[14] Expectation Maximization

کوروساوا و اوتاکی[۲۸] اولین طرح SSE قابل تایید را براساس طرح[۱۴] پیشنهاد کردند. این طرح می‌تواند حذف و یا تغییر نتیجه جستجو را بررسی و تائید کند، در اینجا هزینه محاسبه تایید با تعداد اسناد دارای رابطه خطی است. سان[۱] و همکاران[۲۹] از کلیدواژه‌های کلیدی ربط دهنده برای تایید ساختار داده استفاده کردند. وانگ[۳۰] یک طرح جستجوی کلیدواژه تایید شده مبتنی بر معکوس فیلتر بلوم بدون فرآیند شمارش پیشنهاد داد. مرجع‌های[۳۱] و [۱۲] از توابع MAC برای تایید نتایج استفاده کردند وانگ و همکاران[۲۴] اولین طرح جستجوی کلیدواژه فازی را پیشنهاد کردند، که نه تنها جستجوی کلیدواژه فازی را بر روی داده‌های رمزنگاری شده فراهم می‌کند، بلکه حفظ حریم‌خصوصی کلمات کلیدی و اثبات پذیری نتیجه جستجو را نیز حفظ می‌کند. در طرح ژو و همکاران[۲۵] با استفاده از سیستم کلید عمومی نتیجه جستجو تایید می‌شود و فرآیند تایید امنیت به دلیل استفاده از RSA زمان‌بر است. علاوه بر این، طرح جستجوی کاربر چندگانه قابل تایید[۳۲]، طرح جستجوی کلیدواژه رتبه‌بندی قابل تایید[۳۳]، طرح جستجوی کلیدواژه قابل تایید چند مالکیتی[۳۴]، طرح جستجوی کلیدواژه معنایی قابل تایید[۳۵] و طرح جستجو کلیدواژه پویا[۳۶] ارائه شده‌اند که هر کدام دارای معایب و مزایا هستند.

مفاهیم پایه

رمزنگاری توام با احراز اصالت

می‌توان گفت که تقریباً هر کجا که محرمانگی پیام مورد نظر باشد، صحت پیام نیز مد نظر است، هرچند احراز اصالت[۲] (یا همان صحت) پیام همواره به معنای محرمانه بودن آن نیست. دلایلی وجود دارد که برای حفظ امنیت اطلاعات و ارتباطات نباید از رمزگذاری بدون احراز اصالت استفاده کرد[۳۷]، بنابراین به ناچار می‌بایست که یک الگوریتم برای رمزکردن پیام و الگوریتم دیگری برای احراز اصالت آن در کنار هم استفاده شوند. این ترکیب عموماً به این صورت بوده است که ابتدا پیام را با الگوریتم رمزگذار، رمز می‌کردند و سپس با استفاده از الگوریتم مخصوص احراز اصالت که اصطلاحاً MAC نامیده می‌شود، برچسب صحت آن را بدست آورده و در کنار متن رمزشده قرار می‌دادند. ثابت شده است که چنین ترکیبی، حتی زمانی که الگوریتم رمزگذاری و MAC هر دو امن باشند، می‌تواند به راحتی منجر به یک سیستم ناامن شود[۳۷].

بنابراین سعی شده است تا ترکیب این دو الگوریتم متفاوت به نحوی انجام شود که از بروز چنین امری جلوگیری شود. این مسئله منجر به پیدایش الگوریتم‌های جدیدی شد که در آنها هر دو عمل مذکور همزمان با یک کلید انجام می‌شود. به عبارت دیگر محرمانگی و احراز اصالت همراه باهم برآورده می‌شوند. طرح‌های رمزنگاری که بتوانند این کار را انجام دهند، طرح‌های رمزگذاری احراز اصالت شده یا به اختصار طرح‌های AE[۳] نامیده می‌شوند. طرح‌های AE مانند سایر طرح‌های رمزنگاری دارای بخش‌های رمزگذاری به همراه تولید برچسب احراز اصالت و رمزگشایی به همراه بررسی برچسب احراز اصالت می‌باشد. از بخش رمزگذاری برای برون‌سپاری و از بخش رمزگشایی برای بازیابی اطلاعات استفاده می‌شود. جزئیات و فرایند انجام کار در ادامه در بخش طرح پیشنهادی به‌طور کامل آمده است.

- **طرح رمزنگاری احراز اصالت شده**

با فرض $k,v,t \geq 1$ اگر $K \in \{0,1\}^k$ کلید محرمانه، $N \in \{0,1\}^v$ تک‌شمار تکرار ناپذیر، $M \in \{0,1\}^*$ پیام، $T \in \{0,1\}^t$ برچسب احراز اصالت و $C \in \{0,1\}^*$ متن رمز باشد، آنگاه یک طرح رمزنگاری احراز اصالت، سه‌تایی $\Pi = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ است به شکلی که رویه $\mathcal{K}$ تولیدکننده کلید تصادفی K، $\mathcal{E}_K(N,M)$ الگوریتم رمزنگاری قطعی و $D_K(N,C,T)$ الگوریتم رمزگشایی است. خروجی $\mathcal{E}$ همیشه دوتایی برچسب-متن رمز (C,T) و خروجی D متن آشکار M، یا در غیر معتبر بودن برچسب احراز اصالت پیام، نماد $\perp$ است[۳۸]:

$$\mathcal{E}: \{0,1\}^k \times \{0,1\}^v \times \{0,1\}^* \to \{0,1\}^* \times \{0,1\}^t$$

$$\mathcal{D}: \{0,1\}^k \times \{0,1\}^v \times \{0,1\}^* \times \{0,1\}^t \to \{0,1\}^* \cup \{\perp\}$$

در بسیاری از کاربردها، علاوه بر این‌که پیام، رمزگذاری و احراز اصالت می‌شود، لازم است داده‌ای مثل H وجود داشته باشد؛ به‌طوری که این داده، احراز اصالت شود؛ اما رمزگذاری نشود. مثلاً در شبکه اینترنت نیاز است قسمتی از داده ورودی که آدرس IP است تنها احراز اصالت شود و نیازی به رمز کردن آن نیست، تا مسیریاب‌ها بتوانند به‌طور مناسب بسته‌ها را تشخیص و تبادل کنند. این نیاز باعث می‌شود تا برخی از طرح‌های AE از قابلیت اضافه کردن داده وابسته به ورودی خود برخوردار باشد که این نوع طرح‌ها، طرح رمزگذاری احراز اصالت شده با داده وابسته (AEAD[۴]) نامیده می‌شود[۳۸].

- **طرح رمزنگاری احراز اصالت شده همراه با داده وابسته**

[1] Sun
2 Authentication

[3] Authenticated Encryption
[4] Authenticated Encryption with Associated Data

با فرض $k,v,t \geq 1$ اگر $K \in \{0,1\}^k$ کلید محرمانه، $N \in \{0,1\}^v$ تک‌شمار تکرار ناپذیر، $M \in \{0,1\}^*$ پیام، $T \in \{0,1\}^t$ برچسب احراز اصالت، $H \in \{0,1\}^*$ سرآیند (داده وابسته) و $C \in \{0,1\}^*$ متن رمز باشد، آنگاه یک طرح رمزنگاری احراز اصالت به همراه داده وابسته، سه‌تایی $\Pi = (\mathcal{K}, \mathcal{E}, \mathcal{D})$ است به شکلی که رویه $\mathcal{K}$ تولیدکننده کلید تصادفی K، $\mathcal{E}_K(N,H,M)$ الگوریتم رمزنگاری قطعی و $D_K(N,C,T)$ الگوریتم رمزگشایی است. خروجی $\mathcal{E}$همیشه دوتایی برچسب-متن رمز (C,T) خروجی D یا متن آشکار M، یا در غیر معتبر بودن برچسب احراز اصالت پیام، نماد $\perp$ است[۳۸]

$$\mathcal{E}: \{0,1\}^k \times \{0,1\}^v \times \{0,1\}^* \times \{0,1\}^* \to \{0,1\}^* \times \{0,1\}^t$$

$$\mathcal{D}: \{0,1\}^k \times \{0,1\}^v \times \{0,1\}^* \times \{0,1\}^* \times \{0,1\}^t \to \{0,1\}^* \cup \{\perp\}$$

یک راه‌حل برای بدست آوردن طرح AE، روش‌های ترکیبی عام و یا روش‌های سنتی است که از ترکیب مستقیم روش‌های تولید کننده محرمانگی و احراز اصالت پیام بدست می‌آیند. به این معنی که ابتدا یک الگوریتم رمزنگاری برای رمزکردن پیام به کار می‌رود، سپس از یک کد احراز اصالت پیام برای احراز اصالت آن استفاده می‌شود. اگر چه به دلیل جدا بودن الگوریتم‌ها در این روش، بررسی امنیت طرح AE مورد نظر نسبتاً ساده است (چون امنیت الگوریتم رمز و کد احراز اصالت پیام، به طور مستقل از هم بررسی می‌شوند)؛ اما مهمترین مساله در این روش، احتیاج به دو الگوریتم رمزنگاری و احراز اصالت مجزا، با کلیدهای متفاوت و نیز انجام دو سری محاسبات (دو گذر[1]) روی پیام می‌باشد در مقابل روش سنتی، دسته دیگری از طرح‌های AE وجود دارند که برای تامین همزمان محرمانگی و احراز اصالت پیام از یک الگوریتم و یک کلید استفاده کرده و یک سری محاسبات (تک گذر[2]) روی پیام انجام می‌دهند. این نوع طرح‌ها را می‌توان با سه روش کلی زیر به دست آورد:

روش اول، استفاده از یک رمز قالبی در یک مد عمل خاص است. در این روش، رمز قالبی به عنوان یک جعبه سیاه در نظر گرفته شده و بخش مهم طرح، مد عمل مورد نظر می‌باشد. از این نوع مدهای عمل می‌توان به CCM[۳۹] و GCM[۴۰] اشاره کرد که به عنوان استاندارد NIST[3] نیز پذیرفته شده‌اند. در این روش، مشکل احتیاج به دو الگوریتم متفاوت با دو کلید مختلف حل و در مقایسه با روش ترکیبی عام، کارآیی افزایش داده می‌شود؛ اما مشکل دوگذری بودن محاسبات برای برخی مدهای عمل هنوز وجود دارد. مهم‌ترین مساله در استفاده از مدهای عمل (مبتنی بر یک رمز قالبی) AE این است که در هر بار پردازش یک قالب پیام، یک رمز قالبی به طور کامل اجرا می‌شود که حجم پردازش بالایی می‌خواهد.

روش دوم، استفاده از یک رمز دنباله‌ای است. در این روش رشته کلید حاصل از رمز دنباله‌ای به دو بخش تقسیم می‌شود؛ یک بخش برای رمزگذاری و بخش دیگر برای احراز اصالت استفاده می‌شود. به عنوان مثال از این روش، می‌توان به رمز دنباله‌ای Grain-128a[۴۱] اشاره کرد.

روش سوم، طراحی الگوریتم‌های رمزگذاری احراز اصالت شده اختصاصی است. در این روش، یک الگوریتم طوری طراحی می‌شود که با استفاده از یک کلید و با انجام یک سری محاسبات (محاسبات تک‌گذری) روی پیام، متن رمزی و برچسب احراز اصالت را تولید کند. از AEهای اختصاصی جدید می‌توان به طرح AEGIS [۴۲] اشاره کرد.

امنیت و کارایی طرح‌های AE

امنیت در الگوریتم‌های رمزنگاری (با هدف تامین محرمانگی) و کدهای احراز اصالت پیام (با هدف تامین جامعیت و احراز اصالت مبدا پیام) تعریف مشخصی دارد. از طرف دیگر در طرح‌های AE، دو هدف محرمانگی و احراز اصالت پیام در کنار هم قرار دارند؛ بنابراین تعریف امنیت برای طرح‌های AE، ترکیبی از تعریف امنیت برای الگوریتم‌های رمزنگاری و برچسب‌های احراز اصالت پیام خواهد بود.

یکی از روش‌های مهم بررسی امنیت طرح‌هایAE، استفاده از روش‌های ترکیبی عام حاصل می‌باشد[۴۳]. این مفاهیم را می‌توان به طرح‌های AE اختصاصی نیز تعمیم داد. در یک سطح بالا، هدف از محرمانگی برای طرح‌های رمزنگاری متقارن، تمایزناپذیری[4] و شکل‌ناپذیری[5] آنها تعریف می‌شود که هر یک از این موارد را می‌توان تحت مدل حمله متن آشکار انتخابی یا متن رمزی انتخابی (وفقی) در نظر گرفت.

با توجه به این امر می‌توان چهار مفهوم امنیتی را تعریف کرد: تمایزناپذیری تحت حمله متن آشکار انتخابی[6]، تمایزناپذیری تحت حمله متن رمزی انتخابی[7]، شکل‌ناپذیری تحت حمله متن آشکار انتخابی[8] و شکل‌ناپذیری تحت حمله متن رمزی انتخابی[9].

[1] Two-pass
[2] One-pass
[3] National Institute of Standards and Technology

[4] Indistinguishability
[5] Non-Malleability
[6] Indistinguishability under Chosen Plaintext Attack(IND-CPA)
[7] Indistinguishability under Chosen Ciphertext Attack(IND_CCA)
[8] Non-Malleability under Chosen Plaintext Attack(NM_CPA)
[9] Non-Malleability under Chosen Ciphertext Attack(NM_CCA)

همچنین می‌توان دو مفهوم برای جامعیت طرح‌های رمزنگاری متقارن در نظر گرفت. در جامعیت متن آشکار[1] لازم است تا رمزگشایی یک متن رمزی به پیامی که فرستنده آن را رمز نکرده است، عملی نباشد. در جامعیت متن رمزی[2] نیز لازم است تا به‌دست آوردن یک متن رمزی که از قبل توسط فرستنده تولید نشده است، عملی نباشد. در هر دو مفهوم جامعیت متن آشکار و متن رمزی، مهاجم می‌تواند از یک حمله پیام انتخابی استفاده کند.

مهم‌ترین ارتباط میان مفاهیم امنیتی ذکر شده در بالا را می‌توان به این صورت خلاصه کرد که اگر یک طرح AE ویژگی حمله تحت متن آشکار انتخابی و جامعیت متن رمزی را به‌طور همزمان برآورده کند، آنگاه ویژگی حمله تحت متن رمزی و جامعیت متن آشکار را همزمان برآورده می‌کند[۴۳]. این بدان معنی است که برای به دست آوردن یک AE امن طراحان باید روی برآورده شدن ویژگی‌های حمله تحت متن آشکار انتخابی و جامعیت متن رمزی تمرکز کنند.

منظور از کارآیی یک طرح رمزنگاری، سرعت و الزامات حافظه برای پیاده‌سازی آن است. چون در طرح‌های AE، دو هدف محرمانگی و احراز اصالت روی یک پیام می‌بایست به طور همزمان تامین شوند؛ بنابراین مهم‌ترین تعریفی که می‌توان از کارآیی این طرح‌ها داشت، تک گذر یا دو گذربودن آن‌ها است. منظور از یک طرح AE تک‌گذری، طرحی است که برای تامین محرمانگی و احراز اصالت، یک قالب پیام را یک بار مورد پردازش قرار می‌دهد. در مقابل، منظور از یک طرح AE دوگذری، طرحی است که برای رسیدن به اهداف خود، قالب پیام را دو بار (یک بار برای تامین محرمانگی و بار دیگر برای تامین احراز اصالت پیام) مورد پردازش قرار می‌دهد. پشتیبانی از سایر ویژگی‌ها مانند قابلیت موازی‌سازی، برخط بودن، پشتیبانی از داده وابسته و برچسب میانی، باعث کاربردی‌تر شدن الگوریتم‌های رمزنگاری احراز اصالت شده می‌شود.

مسابقه سزار[3]

سزار یک مسابقه و میدان رقابت برای ارائه شیوه‌های رمزنگاری مبتنی بر احراز اصالت امن، کاربردی و نیرومند است. سزار یک مجموعه از رمزهای مبتنی بر احراز هویت را مشخص می‌کند که اولاً مزایای بیشتری نسبت به طرح AES-GCM [۴۰] ارائه می‌دهد و ثانیاً برای استفاده گسترده مناسب است. طراحان الگوریتم‌های رمزنگاری، طرح‌های خودشان را جهت ارزیابی به کمیته داوری مسابقه سزار ارائه کردند. با شروع مسابقه، ۵۷ طرح به‌عنوان نامزد برنده مسابقه معرفی شده که طرح‌های [۴۴Artemia] و سبک CBA[۴۵] از کشور ایران به این مسابقه معرفی شد. بعد از سه دوره ارزیابی، الگوریتم رمز AEGIS-128 به مرحله نهایی راه پیدا کرد.

طرح رمز توام با احراز اصالت AEGIS-128

این طرح یکی از نامزدهای مرحله نهایی بود و با توجه به معیار امنیت و سرعت در نتایج مسابقه، از نظر کارایی یکی از برترین طرح‌های پیشنهادی می‌باشد. رویکرد این طرح، طراحی یک الگوریتم رمز توام با احراز اصالت اختصاصی و دارای نسخه‌های مختلفی است. در این طرح از خود پیام برای بروزرسانی وضعیت رمزگذاری استفاده می‌شود و احراز هویت پیام تقریباً بدون هزینه حاصل می‌شود[۴۲]. AEGIS از توابع دوری AES بجز آخرین دور ساخته شده است. AEGIS-128L از هشت دور AES دور برای پردازش بلوک پیام ۳۲ بایتی در یک مرحله استفاده می کند. AEGIS-128 یک بلوک پیام ۱۶ بایت را با ۵ عملکرد دور AES پردازش می‌کند. هزینه محاسباتی AEGIS تقریبا نیمی از هزینه AES است و بسیار سریع است. در پردازنده Intel Sandy Bridge Core-i5، سرعت رمزگذاری AEGIS-128L و AEGIS-128 به ترتیب در حدود 0.48 cpb و0.66 cpb است.

در پردازنده Intel Haswell Core-i7، سرعت رمزگذاری AEGIS-128L و AEGIS-128 به ترتیب در حدود 0.37 cpb و 0.60 cpb است. سرعت AEGIS-128L بسیار سریع‌تر از AES در حالت مد عملیاتی شمارنده است و تقریباً ۸ برابر رمزگذاری AES در حالت CBC است. AEGIS دارای امنیت بسیار بالایی است و تا زمانی‌که تک‌شمار مورد استفاده مجدد قرار نگیرد، بازیابی حالت AEGIS و کلید سریع‌تر از جستجوی کامل غیرممکن است (در صورت استفاده از برچسب احراز هویت ۱۲۸ بیتی، حمله جعل با تکرار حمله، موفقیت آمیز نیست). AEGIS برای ارتباطات شبکه مناسب است زیرا AEGIS می تواند سرآیند بسته (داده‌های وابسته) را بدون رمزگذاری، محافظت کند. AEGIS-128 دارای کلید ۱۲۸بیتی، تک‌شمار ۱۲۸ بیتی، برچسب ۱۲۸ بیتی و حالت ۱۰۲۴ بیتی است. مشخصات AEGIS-128L شبیه AEGIS-128 به جز اینکه دارای حالت

[1] Integrity of Plaintext(INT_PTXT)
[2] Integrity of Ciphertext(INT_CTXT)
[3] Competition for Authenticated Encryption: Security, Applicability, and Robustness

۶۴۰ بیتی است. به همین دلیل است که AEGIS-128L سریع‌تر از AEGIS-128 می‌باشد.

ساختار طرح AEGIS-128

طرح AEGIS-128 با استفاده از دو الگوریتم، یک کلید ۱۲۸ بیتی و یک بردار اولیه ۱۲۸ بیتی، یک پیام را رمزگذاری و تصدیق می‌کند. طول داده‌های مرتبط و طول متن ساده کم‌تر از ۲۶۴ بیت و طول برچسب تأیید اعتبار کم‌تر یا برابر با ۱۲۸ بیت است.

StateUpdate128 (S, m)

Input: 80 byte state S and
16 byte message block m
Output: state S_{i+1}

1. $S_{i+1,0} = AESRound(S_{i,4}, S_{i,0} \oplus m_i)$
2. $S_{i+1,1} = AESRound(S_{i,0}, S_{i,1})$
3. $S_{i+1,2} = AESRound(S_{i,1}, S_{i,2})$
4. $S_{i+1,3} = AESRound(S_{i,2}, S_{i,3})$
5. $S_{i+1,4} = AESRound(S_{i,3}, S_{i,4}$

الگوریتم ۱. بروز رسانی حالت

مقداردهی اولیه مطابق الگوریتم ۱ و ۲ شامل اجرا کرد تعداد ۱۰ مرحله رمز بعد از پر کردن کلید، IV و مقادیر ثابت است.

Initialization $(S, Key, Iv, Cons)$

Input: 128 bit Key, 128 Bit IV,
32 byte Constant (2*128 Bit)
Output: 640 Bit state (5*128 Bit)

1. $S_{-10,0} = K_{128} \oplus IV_{128}$
2. $S_{-10,1} = const_1$
3. $S_{-10,2} = const_0$
4. $S_{-10,3} = K_{128} \oplus const_0$
5. $S_{-10,4} = K_{128} \oplus const_1$
6. **for** $i \leftarrow -5$ to -1 **do**
7. $m_{2i} = K_{128}$
8. $m_{2i} = K_{128} \oplus IV_{128}$
9. **end for**
10. **for** $i \leftarrow -10$ to -1 **do**
11. S_{i+1} =StateUpdate128 (S_i, m_i)
12. **end for**

الگوریتم ۲.مقداردهی اولیه

الگوریتم ۳ داده‌های وابسته را به الگوریتم اعمال می‌کند.

AssociatedDataUpdate (S, AD, $adlen$)

Input: Associated Data length=$adlen$
Less than $2^{64}Bit$
Output: Update State

1. **for** $i \leftarrow 0$ to $\left\lceil \frac{adlen}{128} \right\rceil - 1$ **do**
2. S_{i+1} =StateUpdate128 (S_i, AD_i)
3. **end for**

الگوریتم ۳. پردازش داده‌های وابسته

پیام رمزشده و برچسب احراز اصالت پیام مطابق الگوریتم ۴ تولید می‌شود.

Enc-Aegis-128 (P, $msglen$, Key, IV, AD, $adlen$)

Input: Plain Message Data, Associated Data
Length Data ($msglen$) and Length
Associated Data ($adlen$), Lengths Less than $2^{64}Bit$
128 bit Key, 128 Bit IV, 32 byte Constant (2*128 Bit)
Output: Cipher Data (C), Tag (T)

1. Initialization $(S, Key, Iv, Cons)$
2. AssociatedDataUpdate (S, AD, $adlen$)
3. $u \leftarrow \left\lceil \frac{adlen}{128} \right\rceil$
4. $v \leftarrow \left\lceil \frac{msglen}{128} \right\rceil$
5. **for** $i \leftarrow 0$ **to** $v - 1$ **do**
6. $C_i = P_i \oplus S_{u+i,1} \oplus S_{u+i,4} \oplus (S_{u+i,2} \& S_{u+i,3})$
7. S_{u+i+1} =StateUpdate128 (S_{u+i}, P_i)
8. **end for**
9. $tmp = S_{u+v,3} \oplus (adlen || msglen)$
10. **for** $i \leftarrow u + v$ **to** $u + v + 6$ **do**
11. S_{i+1} =StateUpdate128 (S_i, tmp)
12. **end for**
13. $T = S_{u+v+7,0}$
14. **for** $i \leftarrow 1$ **to** 4 **do**
15. $T = T \oplus S_{u+v+7,i}$
16. **end for**

الگوریتم۴. رمزگذاری و تولید برچسب احراز اصالت

خروجی الگوریتم ۵ پیام رمزگشایی شده و برچسب احراز اصالت پیام است. اگر برچسب تولید شده با برچسب دریافت شده یکسان باشد، پیام دریافتی مورد تصدیق قرار می‌گیرد.

Dec- Aegis-128 (C, $msglen$, Key, IV, AD, $adlen$)

Input: Cipher Data, Associated Data Length Data ($msglen$) and Length Associated Data ($adlen$), Lengths Less than $2^{64}Bit$ 128 bit Key, 128 Bit IV, 32 byte Constant (2*128 Bit)
Output: Plain Data (P), Tag (T)

1. Initialization $((S, Key, Iv, Cons)$
2. AssociatedDataUpdate (S, AD, $adlen$)
3. $u \leftarrow \left\lceil \frac{adlen}{128} \right\rceil$
4. $v \leftarrow \left\lceil \frac{msglen}{128} \right\rceil$
5. **for** $i \leftarrow 0$ **to** $v - 1$ **do**
6. $P_i = C_i \oplus S_{u+i,1} \oplus S_{u+i,4} \oplus (S_{u+i,2} \& S_{u+i,3})$
7. S_{u+i+1} =StateUpdate128 (S_{u+i}, P_i)
8. **end for**
9. $tmp = S_{u+v,3} \oplus (adlen || msglen)$
10. **for** $i \leftarrow u + v$ **to** $u + v + 6$ **do**
11. S_{i+1} =StateUpdate128 (S_i, tmp)
12. **end for**
13. $T = S_{u+v+7,0}$
14. **for** $i \leftarrow 1$ **to** 4 **do**
15. $T = T \oplus S_{u+v+7,i}$
16. **end for**

الگوریتم ۵. رمزگشایی

طرح جستجوی کلید واژه فازی موثر قابل تصدیق[1]

نمادگذاری

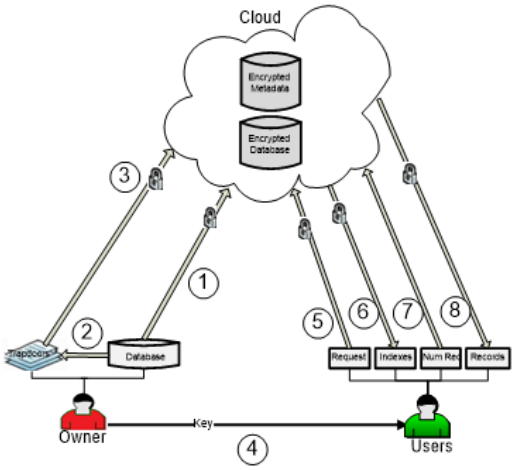
در جدول ۱ نمادهای طرح پیشنهادی به اختصار ارائه شده است.

جدول ۱. توضیح برخی نماد مهم

نماد	توضیح
R_j	*رکورد j‌ام پایگاه داده*
F_i	*فیلد i‌ام پایگاه داده*
N	*تعداد رکوردها*
M	*تعداد فیلدها*
E_{R_j}	*رمزشده رکورد j‌ام پایگاه داده*
T_{R_j}	*برچسب احراز اصالت*
w	*کلمه کلیدی*
$v(w)$	*رکوردهای جاوی کلمه کلیدی* w
$Ev(w)$	*رمز شده رکوردهای جاوی کلمه کلیدی* w
$S_{w_i,d}$	*مجموعه کلیدواژه فازی کلمه کلیدی* w_i *با فاصله ویرایش* d
d	*فاصله ویرایش*
C_i	*خوشه i‌ام*

مدل سیستم

مطابق شکل۲ یک سیستم برون‌سپاری پایگاه داده ابری را که از سه قسمت، مالک پایگاه داده، کاربران و سرور ابری، تشکیل شده است، در نظر می‌گیریم. مالک پایگاه داده، پایگاه داده آشکار $P = \{R_j, 1 \le j \le N\}, R_j = \{F_i, 1 \le i \le M\}$ را به یک پایگاه داده رمزشده $Ep = \{E_{R_j}, T_{R_j}, 1 \le j \le N\}$ تبدیل کرده و برون‌سپاری می‌کند (مرحله ۱). در این‌جا، P به کل پایگاه داده، R_j رکورد شماره j، E_{R_j} محتوی رمزشده رکورد شماره j ، T_{R_j} برچسب احراز اصالت رکورد شماره j، N تعداد رکوردها، F_i مشخصه[2] شماره i و M تعداد مشخصه که در واقع تعداد ستون‌های پایگاه داده می‌باشد، اطلاق شده است. به منظور تحقق جستجوی کارآمد بر روی پایگاه داده رمزگذاری شده، مالک پایگاه داده یک شاخص امن برای کلیه کلمات کلیدی مختلف استخراج شده از فیلدهای پایگاه داده، تولید می‌کند (مرحله ۲). بنابراین مالک پایگاه داده، مجموعه‌ای از کلمات کلیدی مجزا در F_i با عنوان $W_{F_i} = (w_1, w_2, \ldots, w_p)$ را بعد از شاخص‌گذاری، خوشه‌بندی، تولید مجموعه کلیدواژه فازی و رمزگذاری، به عنوان دریچه جستجو G_w در سرور ابری برون‌سپاری می‌کند (مرحله ۳). سرور ابری جستجوی فازی را برای کاربران مجاز از طریق داده‌های رمزگذاری شده انجام می‌دهد. در ادامه، مالک پایگاه داده کلیدهای رمزنگاری مربوط به دریچه‌های جستجو و پایگاه داده را با کاربران مجاز به اشتراک می‌گذارد (مرحله ۴). فرض می‌کنیم که کلیدهای رمزنگاری بین مالک پایگاه داده و کاربران از طریق یک کانال امن توزیع می‌شود. برای جستجو، کاربر مجاز، دریچه مربوط به کلمه کلیدی موردنظر را محاسبه و به سرور ابری می‌فرستد (مرحله ۵). سرور ابری پس از دریافت دریچه از کاربر، عملیات جستجو را روی دریچه‌های ذخیره‌شده انجام داده و در صورت انطباق، شاخص‌های امن به همراه برچسب احراز اصالت مربوطه را به کاربر برمی‌گرداند (مرحله ۶). کاربر پس از رمزگشایی و تائید شاخص‌ها، شماره رکوردهای مورد نظر را از سرور درخواست می‌کند (مرحله ۷). سرور ابری محتوای رکوردهای رمزشده به‌همراه برچسب احراز اصالت را به کاربر ارسال می‌کند (مرحله ۸). در نهایت کاربر پس از رمزگشایی و تائید، نتیجه جستجو را مورد استفاده قرار می‌دهد.



شکل ۲. معماری مدل سیستم

مدل تهدید

در این مدل تهدید، مالک پایگاه داده و کاربران امین فرض شده‌اند. یعنی مالک پایگاه داده، کلیدهای رمزنگاری از طریق کانال امن توزیع می‌کند، صادقانه پایگاه داده را رمزگذاری می‌کند، شاخص‌ها به صورت امن ایجاد می‌کند و برچسب‌های احراز اصالت را درست محاسبه می‌کند. کاربران صادقانه دریچه‌های جستجو را برای کلمه کلیدی ایجاد کرده و به سرور ابر جهت جستجو ارسال می‌کند. ولی سرور ابری "نیمه صادق و کنجکاو" فرض شده، بنابراین یک نهاد غیرقابل اعتماد در نظر گرفته می‌شود. یعنی ممکن است سعی کند اطلاعات ارزشمندی از پایگاه داده رمزگذاری شده، شاخص‌های امن، دریچه‌های جستجو و نتایج جستجو جمع‌آوری کند. علاوه بر این، ممکن است سرور ابری پروتکل توافق شده را به درستی

[1] Efficient Verifiable Fuzzy Keyword Search scheme
[2] Field

دنبال نکند و نتیجه جستجوی نامعتبر و یا حتی ناقص برای صرفه‌جویی در هزینه محاسبه به کاربر برگرداند[۱۲].

اهداف طرح

طرح پیشنهادی بایستی الزامات زیر را برآورده کند:

- **جستجوی کلیدواژه فازی**. این طرح باید قادر به بازیابی همه رکوردهای حاوی کلمه کلیدی مورد جستجو با جبران خطای ورودی با آستانه تعریف شده باشد.
- **کارایی**. این طرح باید به طور موثر نتیجه جستجو را بازگرداند و سربار کمتری از نظر حجم محاسبات و اندازه حافظه برای ذخیره دریچه‌های جستجو را حاصل نماید.
- **حفظ حریم خصوصی**. این طرح نباید هیچ اطلاعات داده‌ای را به سرور ابری مانند متن آشکار کلمات کلیدی و رکوردها خارج از نشت محدود تعریف شده، فاش کند. معمولا، نشت محدود از سه بعد زیر را در نظر می‌گیریم.

۱)**حریم خصوصی پایگاه داده**. نباید محتوی پایگاه داده را بجز شماره رکورد فاش کرد.

۲)**حریم خصوصی شاخص**. سرور ابر نتواند ارتباط بین کلمات کلیدی و رکوردها را از روی شاخص امن بدست آورد

۳)**حریم خصوصی دریچه**. نباید ارتباط بین دریچه و کلمه کلیدی مورد جستجو فاش شود.

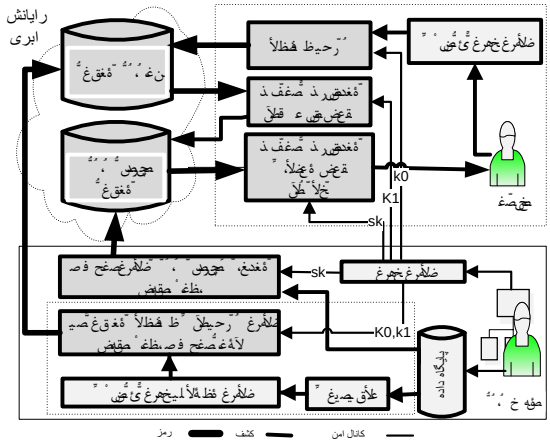
- **تصدیق نتایج پرس‌وجو:** صحت نتایج و کامل بودن جستجو، بررسی و مورد تصدیق قرار گیرد.

توصیف طرح پیشنهادی

مطابق معماری ارائه شده در شکل۳، فرایند انجام کار در دو بخش مالک پایگاه داده و کاربران، نشان داده شده است، در ادامه به بررسی هر یک از اجزای این معماری می‌پردازیم.

مالک پایگاه داده

مالک پایگاه داده به فرد یا سازمانی اطلاق می‌شود که در ابتدا پایگاه داده، دریچه‌ها، شاخص‌ها و برچسب‌های احراز اصالت را به صورت امن برون‌سپاری کرده و کلیدهای رمزنگاری را از طریق یک کانال امن در اختیار کاربران مجاز قرار می‌دهد. توضیح دقیق ریزعملیات‌هایی که باید برای برون‌سپاری انجام شود، مطابق شمل ۳ به شرح زیر است:



شکل ۳. معماری طرح پیشنهادی

- **تولید کلید**[1]

در این مرحله مالک پایگاه داده الگوریتم تولید کلید احتمالی $K \leftarrow Gen(1^k)$ را فراخوانی می‌کند. به‌عنوان مثال می‌توان از توابع تولید کلید شبه‌تصادفی با بذر[2] تصادفی استفاده کرد که خروجی آن غیر قابل تمیز[3] از یک مقدار تصادفی واقعی باشد. خروجی الگوریتم مجموعه کلید مخفی $K = (sk, k_0, k_1)$ می‌باشد که sk کلید رمزنگاری محتوای رکوردها، k_0 کلید رمزنگاری دریچه‌های پرس‌وجو و k_1 کلید رمزنگاری شاخص‌ها می‌باشد. فرض می‌کنیم کلیدهای تولید شده از طریق یک کانال امن توزیع و در اختیار کاربران مجاز قرار می‌گیرد. هم‌چنین با توجه به متقارن بودن الگوریتم‌های رمزنگاری، کلیدهای توزیع شده قابل استفاده در هر دو مرحله رمزگذاری و رمزگشایی می‌باشد.

- **رمزگذاری رکوردهای پایگاه داده**[4]

در این بخش برای تمامی رکوردهای پایگاه داده، متناظر رمزشده به همراه برچسب احراز اصالت با استفاده از بخش رمزگذاری طرح AEGIS تولید و برون‌سپاری می‌شود. در اینجا از شماره رکورد و یا کلید منحصر بفرد رکوردها به‌عنوان داده وابسته و برای جلوگیری از یکسان بودن خروجی رمزشده رکوردها با محتویات یکسان، استفاده شده است که H یک تابع درهم‌ساز خالی از تصادم[5] است. طول پیام بایستی ضریبی از ۱۲۸ بیت باشد و لایه‌گذاری[6] لازم انجام می‌گیرد. روند این مرحله در قالب الگوریتم ۶ ارائه شده است.

[1] Key Generation
[2] Seed
[3] indistinguishable
[4] Encryption Records
[5] Collision
[6] Padding

Enc-DataBase (R, SK)

Input: Records of Database and Key
N Number of Records
$msglen$ Length Records
Output: Encrypted Records and Tags

1. **for** $j \leftarrow 1$ **to** N **do**
2. $IV = H(j)$
3. (C_{R_j}, T_{R_j}) =Enc-Aegis-128 (P_{R_j}, $msglen, sk, IV, j, 128$)
4. **end for**

الگوریتم ۶. رمزگذاری پایگاه داده

- **خوشه‌بندی**

در این مرحله کلمات کلیدی هر فیلد به همراه شماره رکورد آن استخراج شده و با توجه به سنجه تشابه، خوشه‌بندی وخروجی به صورت لیست سه‌تایی که شامل شماره خوشه، نماینده یا مرکز خوشه و شماره رکوردهای مربوط به کلمات داخل خوشه، می‌باشد، استخراج می‌شود. جزئیات این مرحله مطابق الگوریتم خوشه‌بندی مرجع [۱۱] می‌باشد.

- **تولید مجموعه کلیدواژه فازی**

مالک پایگاه داده، مجموعه کلیدواژه فازی را به ازای هر مرکز خوشه تولید می‌کند. فرض می‌کنیم S_{C_i} مجموعه کلیدواژه فازی کلمه کلیدی C_i، با فاصله ویرایش d باشد. مجموعه کلید فازی تولید شده $S_{C_i} = \{C_{i,t}, 1 \le t \le |C_i|\}, 1 \le i \le c$ مجموعه کلیدواژه فازی مراکز خوشه می‌باشد[۱۰].

- **ساخت برچسب، دریچه و شاخص امن**[1]

مالک پایگاه داده، دریچه‌های جستجو را با رمز کردن مجموعه کلیدواژه فازی مراکز خوشه توسط الگوریتم رمزنگاری AEGIS با کلید رمزنگاری k_1، و شاخص امن را با رمز کردن شاخص هر خوشه توسط الگوریتم رمزنگاری AEGIS با کلید رمزنگاری k_2 و استفاده از مرکز خوشه به‌عنوان داده وابسته در تولید شاخص امن، تولید می‌کند. نتایج شامل دریچه‌های رمز شده $ES_{C_{i,t}}$، شاخص‌های رمزشده $Ev(C_i)$ و $Tv(C_i)$ برچسب احراز اصالت شاخص‌ها را در یک لیست سه ستونی برون‌سپاری می‌شود. روند اجرای این مراحل در الگوریتم ۷ آمده است.

Enc-Trapdoors (Clusters,k_1, k_2)

Input: Trapdoors S_c, Indexes $v(C)$, c Number of Cluster, t Number Fuzzy Keywords, $Indlen$ Length index
Output: Encrypted Trapdoors ES_c, Indexes $Ev(C)$ and Tags

1. $IV = Const$
2. **for** $i \leftarrow 1$ **to** c **do**
3. S_{C_i} =Gen_Fuzzy_Set (C_i, d)
4. **for** $j \leftarrow 1$ **to** t **do**
5. $ES_{C_{i,j}}$ =Enc-Aegis-128 ($S_{C_{i,j}}$, 128, k_1, IV, 0,0)
6. **end for**
7. $AD = S_{C_{i,1}}$
$(Ev(C_i), Tv(C_i))$ =Enc-Aegis-128 ($v(C_i)$, $Indlen, k_2, IV, AD$, 128)
8. **end for**

الگوریتم ۷. تولید برچسب، دریچه و شاخص‌های امن

[1] Secure Trapdoor

تعامل کاربر با سرور ابری

- **ساخت دریچه جستجو**

در این مرحله مطابق الگوریتم ۸ ، کاربر کلمه کلیدی مورد نظر را برای جستجو وارد می‌کند، برای جبران خطای ورودی، مجموعه کلیدواژه فازی آن تولید، سپس با استفاده از الگوریتم رمزنگاری AEGIS و کلید رمزنگاری k_1، دریچه‌های جستجو تولید و برای جستجو به سرور ارسال می‌گردد.

Build_Trapdoor (W, d)

Input: W Keywords for Search
d Edit Distance
Output: Fuzzy Keyword Set

1. S_C =Gen_Fuzzy_Set (W, d)
2. $IV = Const$
3. **for** $i \leftarrow 1$ to $|W|$ **do**
4. ES_{C_i} = Enc-Aegis-128 (S_{C_i}, 128, k_1, IV, 0,0)
5. **end for**

return ES_C

الگوریتم ۸. تولید دریچه جستجو

- **پرس‌وجو**

وقتی سرور ابری دریچه‌های جستجو را دریافت می‌کند، مطابق الگوریتم ۹ آنها را با عناصر اولین ستون جدول فراداده (جدول دریچه و شاخص‌ها) مقایسه می‌کند و در صورت برابر بودن هر یک از عناصر، شاخص و برچسب متناظر یعنی $Ev(w')$، $Tv(w')$ و $ES_{c_{c,1}}$ اولین عضو دریچه را به‌عنوان نتیجه به کاربر ارسال کرده و جستجو را خاتمه می‌دهد. علت ادامه ندادن جستجو، تجمیع کلمات کلیدی مشابه در یک خوشه است که عدم وجود دریچه تکراری را تضمین می‌کند.

Search (ET_C)

Input: ET_C Fuzzy Keywords Trapdoor for Search
Output: If find, the authentication label and Record collection
Else Reject

1. **for** $i \leftarrow 1$ to c **do** ; c Cluster Number
2. **for** j$\leftarrow$ 1 to $|ET_C|$ **do**
3. **for** k$\leftarrow$ 1 to $|ES_C|$ **do**
4. **If** ET_{C_j} **Equal** $ES_{C_{i,k}}$ **then**
5. **return** $(Ev(C_i), Tv(C_i), ES_{C_{i,1}})$
6. **end if**
7. **end for**
8. **end for**
9. **end for**
10. **return Reject**

الگوریتم ۹. جستجو

- **رمزگشایی و احراز اصالت نتایج**

در این مرحله، کاربر با رمزگشایی مرکز خوشه رمزشده، برگشتی از سرور توسط الگوریتم رمزنگاری AEGIS با کلید رمزنگاری k_1، از درست بودن جستجو اطمینان حاصل می‌کند (اشتراک مجموعه کلید واژه فازی کلمه کلیدی مورد نظر با

مرکز خوشه دریافتی نباید تهی باشد). در ادامه با رمزگشایی شاخص‌های دریافتی با الگوریتم رمزنگاری AEGIS با کلید رمزنگاری k_2، ضمن اطمینان از صحت و کامل بودن جستجو، رکوردهای حاوی کلمه کلیدی مور نظر بدست آورده و جهت دریافت محتوی رمزشده آنها به سرور ارسال می‌کند. سرور ابری، محتوای رکوردها را از داخل پایگاه داده رمزشده به کاربر بر می‌گرداند. جزئیات در الگوریتم ۱۰ نشان داده شده است.

Dec-Verify-Trapdoors $(d, w', ES_{c_{c,1}}, Ev(w'), len, Tv(w'), k_1, k_2)$

Input: w' Keyword for search, d=Edit Distance
$ES_{w_1'}$ Encrypted Center Cluster
$Ev(w')$ Encrypted Index, len Length index
$Tv(w')$ Tags
Output: Accept (Number or Key Records Results) or Reject

1. $IV = Const$
2. C_c =**Dec-Aegis-128** ($ES_{C_{c,1}}$, 128, k_1, IV, 0,0)
3. **if** (**Gen-Fuzzy-Set** (C_c,d) ∩ **Gen-Fuzzy-Set** (w', d))==∅ **then**
4. **Reject**
5. **else**
6. $AD = H(C_c)$
7. $(v(w'), Tv(w'))$ = **Dec-Aegis-128** ($Ev(w')$, len, k_0, IV, AD, 128)
8. **if** $Tv(w') \neq Tv(C_i)$ **then**
9. **Reject**
10. **else**
11. **Accept**
12. **Return** $v(w')$
13. **end if**

end if

الگوریتم ۱۰. رمزگشایی و تصدیق شاخص‌ها

نهایتاً کاربر محتوای رکوردها را مطابق الگوریتم توسط الگوریتم رمزنگاری AEGIS با کلید رمزنگاری sk مطابق الگوریتم ۱۳ رمزگشایی کرده و در صورت احراز اصالت و تائید شدن، نتایج مورد استفاده قرار می‌دهد.

Dec-Verify-Record $(E_j, T_j, j, msglen, sk)$

Input: E_j Encrypted Record, $msglen$ Length Records
T_j Tag
j Number or Key Record
Output: Accept or Reject Results and Decrypt Record

1. $IV = H(j)$
2. (R_j, T) =**Dec-Aegis-128** ($E_j, msglen, sk, IV, j$, 128)
3. **if** $T_j \neq T$ **then**
4. **Reject**
5. **else**
6. **Accept**
7. **Return** R_j
8. **end if**

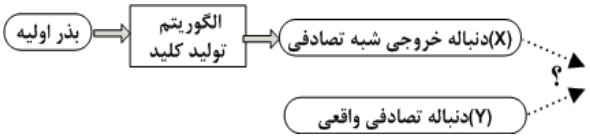
الگوریتم ۱۱. رمزگشایی و تصدیق رکوردها

ملاحظات امنیتی طرح پیشنهادی

تعریف ۱. دو دنباله $X \stackrel{def}{=} \{X_n\}_{n\in\mathbb{N}}$ و $Y \stackrel{def}{=} \{Y_n\}_{n\in N}$ از نظر زمان چندجمله‌ای غیر قابل تمایز هستند، اگر برای هر الگوریتم زمان چندجمله‌ای احتمالی D، چندجمله‌ای مثبت $p(.)$، برای هر n به اندازه کافی بزرگ داشته باشیم[۴۶]:

$$|P_r[D(X_n, 1^n) = 1] - P_r[D(Y_n, 1^n) = 1]| < \frac{1}{p(n)}$$

الف) مجموعه کلیدهای رمزنگاری $K = (sk, k_0, k_1)$ مطابق شکل ۴ و تعریف ۱ توسط الگوریتم تولید کلید رمزنگاری غیر قابل تمایز با دنباله تصادفی واقعی تولید و از طریق کانال امن در اختیار کاربران مجاز قرار می‌گیرد. البته رکوردها، دریچه‌ها و شاخص‌ها با کلیدهای مجزا رمزنگاری می‌شوند که می‌توان برای راحتی توزیع کلید و با مراقبت بیشتر از یک کلید برای هر سه منظور استفاده کرد.



شکل۴. تولید شبه‌تصادفی کلیدهای رمزنگاری

ب) رکوردهای پایگاه داده و شاخص‌ها با IVهای غیر تکراری و اندازه یکسان به صورت رمزشده برون‌سپاری شدند.

در طرح پیشنهادی ما، بعد از تولید مجموعه کلیدواژه فازی مرکز خوشه و ایجاد شاخص‌های مربوط به رکوردها، بدلیل متفاوت بودن طول کلمات، تعداد مجموعه کلیدواژه فازی مرکز خوشه و همچنین تعداد اعضای هر خوشه متفاوت است. بنابراین دشمن و سرور ابری می‌توانند طول کلمه کلیدی را با توجه به تعداد مجموعه کلید واژه‌فازی یاد بگیرد. برای حل این مشکل، ابتدا طولانی‌ترین گره را پیدا می‌کنیم (تعداد مجموعه کلیدواژه فازی و تعداد رکوردهای مرتبط) و سپس بقیه گره‌ها را به اندازه آنها لایه‌گذاری می‌کنیم.

- از شماره رکورد به‌عنوان IV غیر تکراری برای عملیات رمزگذاری و رمزگشایی هر رکورد استفاده می‌شود.
- از مرکز خوشه به‌عنوان IV غیر تکراری برای عملیات رمزگذاری و رمزگشایی شاخص‌های هر خوشه استفاده می‌شود.
- از برچسب احراز اصالت ۱۲۸ بیتی برای بررسی و تایید نتایج استفاده می‌شود.

با توجه به اینکه از طرح *AEGIS* مطابق الگوریتم‌های ۴ و ۵ برای رمزگذاری، رمزگشایی، تولید و تائید برچسب احراز اصالت استفاده شده است، بنابراین امنیت این الگوریتم با رعایت الزامات ذکر شده اثبات شده است [۴۲].

به‌طور کلی طرح پیشنهادی چالش‌های زیر برآورده می‌کند:

۱. اطمینان از صحت ذخیره امن پایگاه داده و فراداده
۲. اطمینان از جستجوی صحیح و کامل
۳. اطمینان از صحت نتایج جستجو

تحلیل امنیتی

تحلیل امنیت از دو جنبه محرمانگی، تحت عنوان قضیه ۱ و تمامیت تحت عنوان قیضه ۲ بررسی می‌کنیم (جنبه دسترس‌پذیری جزء مفروضات طرح می‌باشد).

به طور کلی اطلاعاتی که در اختیار دشمن، مهاجم و یا سرور نا امن قرار گرفته شامل تعداد رکوردها، طول رکوردها، تعدادخوشه‌ها، طول دریچه‌ها، طول شاخص‌ها، طول برچسب احراز اصالت و الگوریتم‌های رمزنگاری و تولید کلید می‌باشد. بقیه اطلاعات با ملاحظات ذکر شده به صورت رمزشده ذخیره شده‌اند.

قضیه ۱: طرح جستجوی کلیدواژه فازی موثر قابل تصدیق، محرمانگی را حفظ می‌کند.

اثبات:

الف) فرض می‌کنیم که در صورت موفقیت آمیز بودن حمله جعل، وضعیت داخلی و کلید سریع‌تر از جستجوی کامل[1] کلید قابل بازیابی نیستند. لذا بدست آوردن کلید از طریق جعل پیام، از نظر محاسباتی کمتر از جستجوی کامل کلیدها نیست، بنابراین رسیدن به کلید از طریق جعل پیام از لحاظ محاسباتی عملی نمی‌باشد[۴۲].

ب) فرض می‌کنیم مهاجم $\mathcal{A}$ با فراخوانی توابع تولید کلید، رمزگشایی رکوردهای پایگاه داده، دریچه‌ها و شاخص‌ها، سعی بر دست‌یابی به متن اصلی رکوردها را دارد.

۱. مهاجم $\mathcal{A}$ الگوریتم KeyGen را برای تولید مجموعه کلید $K = (sk, k_1, k_2)$ فراخوانی می‌کند.

۲.دشمن $\mathcal{A}$ تابع Dec-Verify-Record (E_j, T_j, j, l, sk) را برای رمزگشایی رکوردهای $1 \le j \le N$ فراخوانی می‌کند. N تعداد رکوردها، l طول رکورد، E_j متن رمزشده و T_j برچسب احراز اصالت ذخیره می‌باشد. در اینجا با توجه رعایت ملاحظات امنیتی الگوریتم و تولید شبه‌تصادفی کلید در زمان رمزگذاری، برچسب احراز اصالت تولید شده در حین رمزگشایی با برچسب احراز اصالت ذخیره شده، یکسان نخواهد بود، بنابراین دشمن به متن اصلی نمی‌رسد و محرمانگی حفظ می‌شود.

قضیه۲: طرح جستجوی کلیدواژه فازی موثر قابل تصدیق، صحت و قابلیت اطمینان را برآورده می‌کند.

اثبات: صحت و قابلیت اطمینان این طرح در سه مرحله بررسی می‌شود.

الف) جستجوی فازی: وقتی سرور دریچه جستجو را به صورت رمزشده دریافت می‌کند $T_{w'}$، آن را با تک‌تک عناصر دریچه‌های هر خوشه $ES_{C_{c,t}}$ مقایسه می‌کند. در صورت برابر بودن یکی از عناصر، جستجو خاتمه یافته و سرور $Ev(w')$، $Tv(w')$ و $ES_{C_{c,1}}$ که شامل شاخص‌ها، برچسب شاخص‌ها و اولین عضو دریچه رمزشده است را به‌عنوان نتیجه به کاربر ارسال می‌کند. مطابق الگوریتم ۱۲ کاربر با رمزگشایی $ES_{C_{c,1}}$ به مرکز خوشه می‌رسد C_c و سپس مجموعه کلیدواژه فازی مرکز خوشه و کلمه کلیدی وارد شده w' را تولید می‌کند. اگر اشتراک دو مجموعه مخالف تهی باشد داریم:

$$S_{w',d} \cap S_{C,d} \neq \emptyset \Rightarrow d(C_c, w') \le d$$

یعنی جستجوی فازی با فاصله ویرایش d درست انجام شده است.

ب) تمامیت و صحت جستجو: یعنی اینکه شاخص‌های مرتبط با دریچه‌ها را به صورت کامل در اختیار کاربر قرار گیرد. برای اثبات این موضوع کافی است مطابق الگوریتم ۱۰ عملیات رمزگشایی شاخص‌ها را که بر مبنای الگوریتم AEGIS می‌باشد به صورت زیر انجام دهیم:

Dec-Aegis-128 $(Ev(w'), len, k_2, IV, H(C_c), 128)$

که H یک تابع درهم‌ساز خالی از تصادم است. اگر برچسب احراز اصالت تولید شده با برچسب احراز اصالت دریافت شده یکسان باشد، صحیح و کامل بودن نتیجه جستجو را تائید می‌شود و شماره و یا کلید منحصر بفرد رکوردها در اختیار کاربر قرار می‌گیرد.

ج) تمامیت و صحت رکوردها: یعنی اینکه رکوردهای رمزشده درخواستی به صورت صحیح، کامل و درست در اختیار کاربر قرار گیرد. در اینجا مطابق الگوریتم ۱۱ وقتی کاربر رکوردی را با شماره رکورد و یا کلید مخصوص آن مانند j از سرور درخواست می‌کند، بعد از درخواست، دریافت محتوی رکورد و برچسب احراز اصالت رکورد درخواست شده، تابع رمزگشایی Dec-Verify-Record $(E_j, T_j, j, msglen, sk)$ را فراخوانی می‌کند و از شماره منحصر بفرد رکورد برای تولید IV و داده وابسته استفاده می‌کند. اگر برچسب احراز اصالت تولید شده با برچسب احراز اصالت دریافت شده یکسان باشد، صحیح و کامل بودن نتیجه جستجو را تائید می‌شود.

مقایسه کارایی و ارزیابی نتایج

ارزیابی در دو بخش، یکی مقایسه از نظر کارایی (حجم پردازش و حافظه مورد نیاز) و دیگری مقایسه نتایج تجربی و آزمایش عملی بر روی انواع مجموعه داده[2] صورت گرفته است.

[1] Exhaustive key

[2] Data Set

مقایسه کارایی

در جدول ۲ قابلیت طرح پیشنهادی با طرح‌های جی[۱۲] و وانگ[۲۴]، ژو[۲۵] و کوروساوا[۲۸] مقایسه شده است. طرح کوروساوا نتایج جستجو را تائید می‌کند، اما نمی‌توانند به صورت فازی جستجو کنند. طرح پیشنهادی و بقیه طرح‌ها توانایی جستجوی فازی و ارزیابی نتایج جستجو را دارند.

جدول ۲. مقایسه قابلیت عملکردی

قابلیت جستجو	طرح کوروساوا [28]	طرح وانگ [24]	طرح ژو [25]	طرح جی [12]	طرح پیشنهادی
فازی	–	√	√	√	√
تصدیق	√	√	√	√	√

در جدول ۳ کارایی طرح پیشنهادی با طرح‌های فوق از نظر پیچیدگی زمانی[1] شاخص‌گذاری، تولید دریچه، عملیات جستجو، ایجاد محرمانگی، تصدیق و هزینه حجم حافظه جهت ذخیره دریچه‌های جستجو، انجام گرفته است.

در اینجا N به‌عنوان تعداد رکورد یا اسناد، n به عنوان تعداد کلمات کلیدی، l طول کلمات کلیدی، M اندازه مجموعه کلیدواژه فازی، m تعداد کل کلمات فازی، h عمق شاخص-دهی، C تعداد خوشه و M' تعداد کلمات کلیدی فازی برای جستجو است. تفاوت اساسی که طرح پیشنهادی با بقیه طرح-ها دارد این است که در طرح پیشنهادی محرمانگی و تائید اصالت با استفاده از یک الگوریتم کارآمد AE به صورت هم-زمان تامین می‌شود[۴۲] ولی در طرح‌های دیگر از دو الگوریتم جداگانه، یعنی از یک تابع برای محرمانگی (PRP[2]) و از یک تابع دیگر (PRF[3] یا MAC) برای تصدیق نتایج جستجو استفاده شده است. به همین علت الگوریتم پیشنهادی در دو شاخص پیچیدگی زمانی تولید دریچه و جستجو نسبت به طرحهای قبلی عملکرد بهتری دارد.

جدول ۳. مقایسه کارایی

طرح / هزینه	کوروساوا [28]	وانگ [24]	ژو [25]	جی [12]	پیشنهادی
پیچیدگی زمانی تولید شاخص	$O(N)$	$O(nM)$	$O(m)$	$O(n)$	$O(n)$
پیچیدگی زمانی تولید دریچه	$O(N)$	$O(M)$	$O(M)$	O(1)	O(C)
پیچیدگی زمانی جستجو	$O(N)$	$O(M'h)$	$O(M'h)$	$O(m)$	$O(C)$
پیچیدگی زمانی تصدیق	$O(N)$	$O(1)$	$O(1)$	$O(1)$	$O(1)$
هزینه حجم شاخص و دریچه	$O(mlM)$	$O(mlM)$	$O(mlM)$	$O(mlM)$	$O(MlC)$

- **پیچیدگی زمانی تولید شاخص**

این مرحله شامل استخراج کلمات کلیدی، برچسب‌گذاری، حذف ایست واژه‌ها و تشکیل ماتریس ارتباط می‌باشد و زمان مورد نیاز در هر ستون به تعداد کلمات وابسته است. بنابراین از نظر پیچیدگی زمانی مرتبه‌ای از تعداد کلمات استخراج شده یعنی $O(n)$ می‌باشد.

- **پیچیدگی زمانی تولید دریچه**

این مرحله شامل تولید مجموعه فازی کلمات، رمزگذاری دریچه‌ها، رمزگذاری شاخص‌ها و تولید برچسب احراز اصالت برای هر کلمه می‌باشد در طرح پیشنهادی زمان مورد نیاز کل وابسته به تعداد خوشه است یا به‌عبارتی در داخل حلقه به تعداد خوشه $O(c)$ عملیات تولید مجموعه کلیدواژه فازی و تولید دریچه، شاخص و برچسب انجام می‌گیرد ولی در طرح-های مشابه متناسب با تعداد کلمات اصلی یعنی $O(n)$ می‌باشد.

- **پیچیدگی زمانی عملیات جستجو**

پیچیدگی زمانی جستجو شامل کل زمان دستیابی کاربر به شاخص‌ها و یا شماره رکوردهای حاوی کلمه می‌باشد که مراحل تولید دریچه جستجو (تولید مجموعه کلیدواژه فازی و رمزنگاری آن)، جستجو توسط سرور، رمزگشایی به همراه تائید و اطمینان از کامل بودن می‌باشد. در طرح پیشنهادی این زمان متناسب با تعداد خوشه می‌باشد، زیرا بخش تولید دریچه جستجو و رمزگشایی و احراز اصالت آنها یک زمان ثابتی است و سرور متناسب با تعداد خوشه زمان برای پیدا کردن دریچه زمان نیاز دارد $o(c)$. در سایر طرحها این زمان را درجه‌ای از تعداد کل کلمات فازی$o(m)$ در نظر گرفته‌اند.

- **پیچیدگی زمانی تصدیق نتایج**

پیچیدگی زمانی این بخش مربوط به رمزگشایی و احراز اصالت و کامل بودن نتایج جستجو می‌باشد. در طر پیشنهادی این زمان تنها برای اجرای الگوریتم رمزگشایی به همراه مقایسه برچسب احراز اصالت می‌باشد. در سایر طرحها هم این زمان دارای درجه ثابت است.

- **اندازه حجم حافظه**

اندازه حافظه مورد نیاز برای ذخیره دریچه‌ها، شاخص‌ها و برچسب‌های احراز اصالت در طرح پیشنهادی متناسب با تعداد خوشه، طول کلمات و تعداد کلمات مجموعه کلید واژه فازی می‌باشد$o(Mlc)$، و این حجم در طرح مشابه متناسب با تعداد کلمات، طول کلمات و تعداد کلمات مجموعه کلید واژه فازی می‌باشد$o(mlc)$. با توجه به اینکه تعداد خوشه‌ها همواره از تعداد کلمات خیلی کمتر است، بنابراین حافظه مورد نیاز برای ذخیره خیلی کمتر از طرح‌های مشابه می‌باشد.

[1] time complexity
[2] Pseudo-Random Permutation functions
[3] Pseudo-Random Functions

ارزیابی عملی

در این بخش برای نشان دادن کارآیی طرح پیشنهادی، آزمایش‌ها روی مجموعه داده‌های واقعی مانند اسامی و لغات با حجم و اندازه مختلف، انجام می‌شود. پارامترهای که برای ارزیابی مد نظر قرار گرفته شامل تعداد کلمات کلیدی، حجم حافظه مورد نیاز برای ذخیره دریچه و شاخص‌ها، سرعت جستجو و تصدیق جستجو می‌باشد. با توجه به اینکه طرح جی[۱۲] نسبت به بقیه طرح‌ها بهینه‌تر است، لذا در ارزیابی عملی طرح پیشنهادی با این طرح مقایسه گردیده است.

برای ارزیابی از یک سیستم کامپیوتری با پردازنده اصلی اینتل Core i3-2350M CPU 4.0GB RAM با 2.30 GHz, استفاده‌شده است. کد (بر اساس الگوریتم) به زبان ویژوال $C++$ در یک سیستم‌عامل ۶۴ بیتی ویندوز ۸، نوشته‌شده است. تمامی الگوریتم‌های طرح پیشنهادی و طرح‌های مشابه در شرایط یکسان پیاده‌سازی و با داده‌های آزمایشی یکسان، مورد ارزیابی قرارگرفته‌اند. روال اجرای طرح پیشنهادی با بقیه طرح‌ها به‌جز الگوریتم‌های خوشه‌بندی یکسان است. برای طرح جی الگوریتم رمز AES با مد عملکردی CBC برای محرمانگی و تولید دریچه و شاخص امن و از SHA-256 برای تولید برچسب ۱۲۸ بیتی استفاده شده است. برای محرمانگی و تولید دریچه و شاخص امن و احراز اصالت طرح پیشنهادی از طرح AEGIS-128 استفاده شده است. برای خوشه‌بندی، حداکثر طول کلمات کلیدی ۳۰ کاراکتر و فاصله ویرایش (آستانه تحمل خطا) عدد یک فرض شده است.

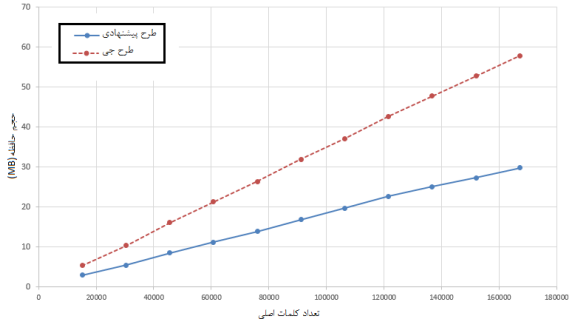
آزمایش عملی طرح پیشنهادی بر روی بیش از ۱،۰۰۰،۰۰۰کلمه در قالب ۸ مجموعه داده اجرا گردیده است که در اینجا ما به گزارش سه نمونه که شامل کلمات کلیدی ذیل هستند اکتفا می‌کنیم.

الف) ۱۶۷۳۵۰ کلمه ۱ تا ۳۰ کاراکتری موجود در لغت‌نامه جیبی تزاروس[1]

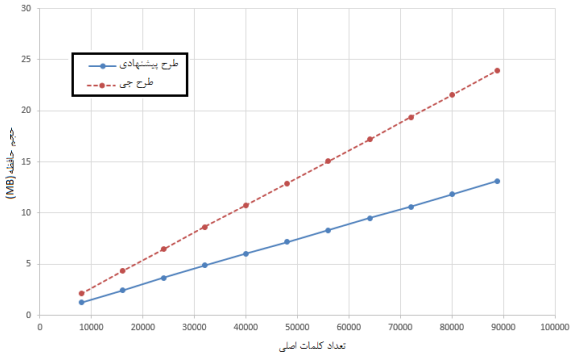
ب) ۸۹۱۰۰ لغات ۱ تا ۲۰ کاراکتری بکار برده شده در مجموعه مقالات کنفرانس نیپس[2] به زبان لاتین

ج) ۱۳۲۰۰ نام خانوادگی ۵ تا ۱۷ کاراکتری به زبان لاتین

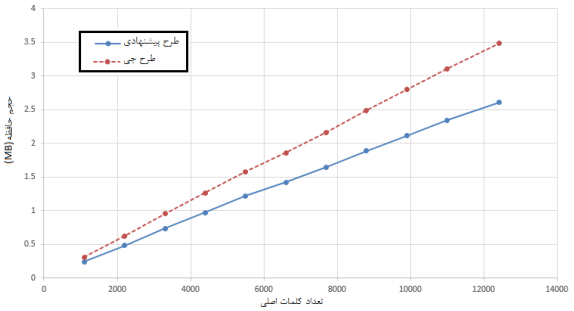
شکل‌های ۵-الف، ۵-ب و ۵-ج مقایسه حجم حافظه مورد نیاز برای ذخیره دریچه‌های جستجو بر اساس تعداد کلمات اصلی را نشان می‌دهد.



شکل ۵-الف. خوشه‌بندی کلمات کلیدی

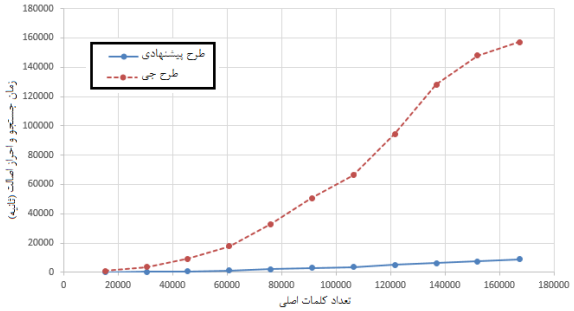


شکل ۵-ب. خوشه‌بندی کلمات کلیدی



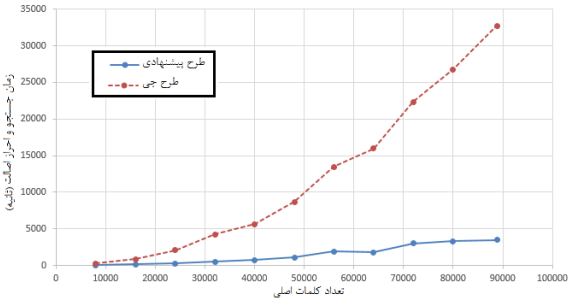
شکل ۵ -ج. خوشه‌بندی کلمات کلیدی

شکل‌های ۶-الف، ۶-ب و ۶-ج زمان جستجو و احراز اصالت را بین طرح پیشنهادی و طرح جی در سه مجموعه داده را نشان می‌دهد.

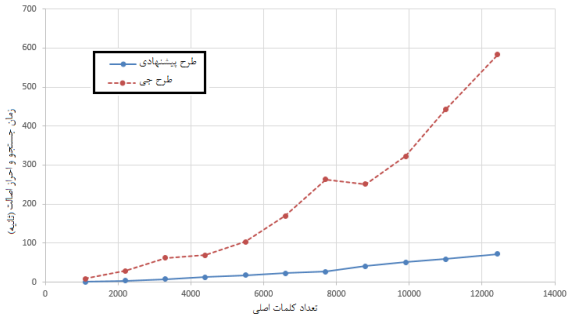


شکل۶-الف. زمان جستجو و احراز اصالت

[1] Thesaurus
[2] Nips



شکل۶-ب. زمان جستجو و احراز اصالت



شکل۶-ج. زمان جستجو و احراز اصالت

در طرح پیشنهادی با توجه به این‌که تولید برچسب احراز اصالت همزمان با رمزگشایی (تک‌گذر) انجام می‌گیرد، لذا زمان مورد نیاز برای بررسی و تائید تقریباً قابل چشم‌پوشی است.

نتیجه گیری

در این مقاله، ما یک طرح جستجوی کلمات کلیدی فازی قابل تصدیق کارا بر روی پایگاه داده برون‌سپاری شده در رایانش ابری ارائه دادیم. برای کاهش حجم حافظه مورد نیاز برای ذخیره فراداده، از روش خوشه‌بندی کلمات کلیدی استفاده کردیم تا با کاهش حجم فراداده، سرعت جستجو را هم افزایش دهیم، همچنین از روش تولید مجموعه کلیدواژه فازی برای جبران خطای کاربران استفاده کردیم. در طرح پیشنهادی برای احراز اصالت نتایج جستجو، تمامیت جستجو و اطمینان از صحت ذخیره‌سازی، با استفاده از روش رمزگذاری توام با احراز اصالت برای محرمانگی و تمامیت نتایج استفاده کردیم.

از طریق مقایسه با طرح‌های مشابه و انجام آزمایش بر روی مجموعه داده‌ها نشان داده شد که طرح پیشنهادی نسبت به طرح‌های مشابه، از نظر حجم دریچه‌های جستجو و سرعت جستجو در وضعیت بهتری قرار دارد. همچنین با لحاظ کردن ملاحظات امنیتی و استفاده طرح رمز توام با احراز اصالت AEGIS، طرح پیشنهادی از امنیت و قابلیت مقیاس‌پذیری[1] لازم برخوردار است.

[1] Scaling factor

مراجع

[1] J. Yu, K. Ren, C. Wang, and V. Varadharajan, "Enabling cloud storage auditing with key-exposure resistance," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 6, pp. 1167_1179, Jun. 2015.

[2] Y. Zhang, J. Yu, R. Hao, C. Wang, and K. Ren, "Enabling efficient user revocation in identity-based cloud storage auditing for shared big data," *IEEE Trans. Depend. Sec. Comput.*, to be published, doi: 10.1109/TDSC.2018.2829880.

[3] J. Yu, K. Ren, and C. Wang, "Enabling cloud storage auditing with verifiable outsourcing of key updates," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 6, pp. 1362_1375, Jun. 2016.

[4] X. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," In: *Proceedings of the 2000 IEEE Symposium on Security and Privacy*, pp. 44–55. IEEE, Berkeley, California, USA (2000).

[5] J. Domingo-Ferrer, "A new privacy homomorphism and applications,"*Information Processing Letters*. 60(5): 277–82, Dec 1996.

[6] R. Agrawal, J. Kieman, R. Srikant, and Y. Xu; "Order-preserving encryption for numeric data," In *Proc. of the ACM SIGMOD 2004 Conference*. Paris, France, June 2004.

[7] R. Brinkman, J. M. Doumen, P. H. Hartel, and W. Jonker, "using secret sharing for searching in encrypted data," In *Secure Data Management VLDB 2004 workshop*, volume LNCS 3178, pages 18–27, Toronto, Canada, August 2004. Springer-Verlag, Berlin.

[8] H. Hacıgumus, R. Iyer, and S. Mehrotra: "Executing SQL over encrypted data in the database service provider model," In *SIGMOD Conference*, 2002.

[9] H. Hacıg, B. Iyer, S. Mehrotra, "Efficient execution of aggregation queries over encrypted relational databases," *In ACM SIGMOD*, 2002 June 46, Madison, Wisconsin, USA Copyright 2002 ACM 1581134975/02/06.

[10] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren and W. Lou, "Fuzzy keyword search over Encrypted data in cloud computing," In *Proceedings of the 29th IEEE International Conference on Computer Communications*, 2010.

[11] Y.Dehghanian, M.Ghayouri Sales, A.Rahimi,
"Fuzzy keyword search scheme on the encrypted database in cloud computing using

word clustering," *Journal of Electronical & Cyber Defence*, 2020, in Persian.

[12] X. Ge, j. Yu, H. Zhangr and R. Hao: "Enabling Efficient Verifiable Fuzzy Keyword Search over Encrypted Data in Cloud Computing", *IEEE Access*, August 17, 2018.

[13] E. J. Goh, "Secure indexes," In *Cryptology ePrint Archive*, Report 2003/216, 2003.

[14] R. Curtmola, J. Gary, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: Improved definitions and ef_cient constructions," *in Proc.ACM Conf. Comput. Commun*. Secur. 2006, pp. 79_88.

[15] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-preserving multi-keyword ranked search over encrypted cloud data," in *Proc.IEEE INFOCOM*, 2011.

[16] Z. Xia, X. Wang, X. Sun, and Q. Wang, "A secure and dynamic multikeyword ranked search scheme over encrypted cloud data," *IEEE Trans.Parallel Distrib, Syst*, 2016.

[17] Z. Fu, K. Ren, J. Shu, X. Sun, and F. Huang, "Enabling personalized search over encrypted outsourced data with efficiency improvement," *IEEE Trans. Parallel Distrib. Syst*, vol. 27, no. 9, 2016.

[18] M. Kuzu, M. S. Islam, and M. Kantarcioglu, "Efficient similarity search over encrypted data," *28th International Conference on Data Engineering*, pp. 1156–1167, 2012.

[19] B. Wang, S. Yu, W. Lou, and Y. T. Hou, "Privacy-preserving multi keyword fuzzy search over encrypted data in the cloud", in *Proc. IEEE INFOCOM*, 2014.

[20] C. Liu, L. Zhu, L. Li, and Y. Tan, "Fuzzy keyword search on encrypted cloud storage data with small index," *ICCCIS 2011*, pp. 269–273, 2011.

[21] M. Chuah and W. Hu, "Privacy-aware bedtree based solution for fuzzy multi-keyword search over encrypted data," *ICDCSW 2011*, pp. 273–281, 2011.

[22] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W.Lou, "Enabling Efficient Fuzzy Keyword Search over Encrypted Data in Cloud Computing", 2010.

[23] N. Mahajan, V. Barkade, "Clustering Based Efficient Privacy Preserving Multi Keyword Search over Encrypted Data", *IEEE Trans*, 2018.

[24] J. Wang, H. Ma, Q. Tang, j.Li, H.Zhu, S.Ma, X.Chen: "Efficient Verifiable Fuzzy Keyword over Encrypted Data in Cloud Computing", Com SIS Vol.10, No, 2 Special Issue 2013.

[25] X.Zhu, Q. Liu, and G. Wang, "A novel veri_able and dynamic fuzzy keyword search scheme over encrypted data in cloud computing," in Proc. IEEE rustcom/BigDataSE/ISPA, Aug. 2017, pp. 845_851.

[26] Q. Chai and G. Gong, "Verifiable symmetric searchable encryption for semi-honest-but-curious cloud servers," *in Proc. IEEE Int. Conf. Commun*., Jun. 2012, pp. 917_922.

[27] C. Wang, N. Cao, K. Ren, and W. Lou, "Enabling secure and eficient ranked keyword search over outsourced cloud data, " *IEEE Trans. Parallel Distrib. Syst.*, vol. 23, no. 8, pp. 1467_1479, Aug. 2012.

[28] K. Kurosawa and Y. Ohtaki, "UC-secure searchable symmetric encryption," *in Proc. Int. Conf. Financial Cryptogr. Data Secur*. 2012, pp. 285_298.

[29] W. Sun, X. Liu, W. Lou, Y. T. Hou, and H. Li, "Catch you if you lie to me: Ef_cient veri_able conjunctive keyword search over large dynamic encrypted cloud data," in Proc. IEEE Conf. Comput. Commun. (INFOCOM), Apr. /May 2015, pp. 2110_2118.

[30] J.Wang, X. Chen, J. Li, J. Zhao, and J. Shen, "Towards achieving _exible and veri_able search for outsourced database in cloud computing," Future Gener. Comput. Syst., vol. 67, pp. 266_275, Feb. 2017.

[31] X. Jiang, J.Yu, J.Yan, and R. Hao, "Enabling efficient and verifiable multikeyword ranked search over encrypted cloud data," *Inf. Sci.*, vols. 403_404, pp. 22_41, Sep. 2017.

[32] L. Chen and N. Zhang, "Efficient verifiable multi-user searchable symmetric encryption for encrypted data in the cloud," *in Proc. Int. Conf. Secur. Privacy New Comput*. Environ. 2016, pp. 173_183.

[33] W. Zhang, Y. Lin, and Q. GU, "Catch you if you misbehave: Ranked keyword search results veri_cation in cloud computing," *IEEE Trans. Cloud Comput.*, vol. 6, no. 1, pp. 74_86, Jan. /Mar. 2018.

[34] Y. Miao, J. Ma, X. Liu, J. Zhang, and Z. Liu, "VKSE-MO: Verifiable keyword search over encrypted data in multi-owner settings," *Sci. China Inf. Sci.*, vol. 60, no. 12, p. 122105, 2017.

[35] Z. Fu, J. Shu, X. Sun, and N. Linge, "Smart cloud search services: Verifiable keyword-based semantic search over encrypted cloud data," *IEEE Trans. Consum. Electron.* vol. 60, no. 4, pp. 762_770, Nov. 2014.

[36] J. Wang, X. Chen, and J. Li, "Verifiable search for dynamic outsourced database in cloud computing," *in Proc. Int. Conf. Broadband Wireless Comput.*, Nov. 2015, pp. 568_571.

[37] S. Bellovin, "Problem areas for the IP security protocols", *Proceedings of the Sixth USENIX Security Symposium*, pp. 1–16, (1996).

[38] F. Abed, C. Forler, and S. Lucks, "Genral Overview of the Firsy-Round CAESAR Candidates for Authenticated Encryption", February 25, 2015.

[39] National Institute of Standards and Technology, "Recommendations for Block Cipher Modes of Operation: The CCM Mode for Authentication and Confidentiality", *NIST special publication 800-38C*, (2004).

[40] National Institute of Standards and Technology, "Recommendations for Block Cipher Modes of Operation: Galois/Counter Mode (GCM) and GMAC", *NIST special publication* 800-38D, (2007).

[41] M. Agren, M. Hell, T. Johansson, W. Meier, "Grain-128a: A New Version of Grain-128 with Optional Authentication", *International Journal of Wireless and MobileComputing*, Vol 5, No 1, pp. 48-59, (2011).

[42] H. Wu1, B. Preneel, "AEGIS: A Fast Authenticated Encryption Algorithm (v1.1)", *In Selected Areas in Cryptography*, volume 8282 of Lecture Notes in Computer Science, pages 185-201.

[43] M.Bellare, C.Namprempre, "Authenticated Encryption: Relations among notions and analysis of the generic composition paradigm", *Advances in Cryptology - Asiacrypt* 2000, pp. 531-545, (2000).

[44] J. Alizadeh, M. Aref, B. Bagheri, "Artemia: A Family of Provably Secure Authenticated Encryption Schemes", *ISeCure*, Jul2014, Vol. 6 Issue 2, p125-139. 15p.

[45] H. Hpsseni, M.R. Aref, S.Khazaei, "CBA Mode, A SUBMISSION TO CAESAR COMPETITION FOR AUTHENTICATED ENCRYPTION", 2014.

[46] O. Goldreich, "Foundations of Cryptography Basic Tools", *Weizmann Institute of Science*, 2004