# روشهای برقراری امنیت در داده های بزرگ

**ایده آذری***

**عضو هیات علمی دانشگاه آزاد اسلامی واحد مبارکه**

Email:i.azari@mau.ac.ir

## چکیده

یکی از چالشهای کلیدی در داده‌های بزرگ مسئله امنیت است. امروزه داده‌های بزرگ به راحتی از طریق زیرساخت ابری در اختیار همگان قرار می‌گیرند. از طرفی زیرساخت‌های نرم‌افزاری این امکان را برای توسعه‌دهندگان ایجاد می‌کنند تا بر روی هزاران گره به صورت موازی پردازش انجام دهند، درنتیجه با ترکیب این حوزه‌ها چالشهای جدید امنیتی برخاسته از ترکیب زیرساختهای متفاوت، میان‌افزارهای پردازشی و محیطهای ذخیره‌سازی ایجاد شده است.

در این مقاله ابتدا چالشها و مشکلات رایج در امنیت و حریم خصوصی مطرح شده است. و سپس زمینه‌های مختلفی از حفظ حریم خصوصی و محافظت از داده‌هادر مقابل حملات مختلف بیان شده و اقدامات لازم برای رسیدن به محرمانگی و تمامیت در ابر و چند نمونه کاربردهای داده‌های بزرگ در امنیت سیستمها مطرح گردیده است. و در نهایت روشهای امنیتی موجودو یک مدل امنیتی پیشنهاد داده شده که تا حدودی روشهای موجود را بهبود بخشده و گامی در جهت ارتقا امنیت و حفظ و نگهداری از این داده ها برداشته است.

**کلمات کلیدی: امنیت داده های بزرگ ، حفظ حریم خصوصی، کنترل دسترسی،تحلیل داده های بزرگ**

## ۱–مقدمه

مسئله امنیت و حریم خصوصی در داده‌های بزرگ با توجه به مقیاس حجیم داده‌ها چالش‌های بسیاری به همراه دارد، مانند الگوریتم‌های کارآمد رمزگذاری و رمزگشایی، بازیابی اطلاعات رمز شده، رمزگذاری مبتنی بر ویژگیها، قابلیت اطمینان و یکپارچگی داده‌های بزرگ. برنامه‌های داده‌های بزرگ، اشتراک‌گذاری و مبادله داده‌ها را در طیف گسترده و حجم بالایی از اطلاعات در فرمتهای مختلف می‌توانند انجام دهند. رویکرد امنیتی برای کنترل این که چگونه، چطور و چه هنگامی دست-اندرکاران به برنامه داده‌های بزرگ دسترسی داشته باشند، مهم است. باید مطمئن بود که سیاستهای امنیت محلی در سیستم‌های تشکیل دهنده با یکدیگر سازگار هستند تا به یک خط مشی امنیتی قابل استفاده و سازگار دست یابند.

## ۲– چالشهای امنیت و حفظ حریم خصوصی

حفظ حریم خصوصی تا حدی به محدودیت‌های تکنولوژی در توانایی استخراج، تجزیه‌وتحلیل روی مجموعه داده‌های حساس مرتبط است. پیشرفت در تحلیل داده‌های بزرگ ابزارهایی را برای استخراج و استفاده از این داده‌ها ارائه می‌کند که نقض حریم خصوصی را آسانتر می‌کند. درنتیجه با گسترش ابزارهای داده‌های بزرگ، لازم است تا یک محافظ امنیتی برای جلوگیری از سو استفاده‌ها ایجاد شود. علاوه بر حریم خصوصی، داده‌های مورد استفاده در تحلیل‌ها، ممکن است شامل اطلاعات تنظیم کننده و یا اطلاعات با مالکیت معنوی باشد، معماران این سیستم‌ها باید اطمینان دهند که داده‌ها محافظت می‌شوند و تنها با توجه به مقررات استفاده می‌شوند. مشکلات مربوط به حریم خصوصی و امنیت درنتیجه سرعت، حجم، تنوع داده‌ها در زیرساخت‌های مقیاس بزرگ ابر، با تنوع منابع و اشکال داده‌ها و حجم بالای مهاجرت درون ابری رشد می-کنند. استفاده از زیرساخت‌های ابر احتمال حمله به کل سیستم را افزایش می‌دهند؛ و مکانیسم‌های سنتی برای ایجاد امنیت در داده‌های کوچک، دیگر مناسب نیستند. در این بخش چند مورد از چالشهای مربوط به امنیت و حریم خصوصی مطرح شود.

### ۱–۲– محاسبات امن در چارچوبهای برنامه نویسی توزیع شده

چارچوب‌های برنامه‌نویسی توزیع‌شده از موازی سازی در محاسبه و ذخیره‌سازی برای پردازش حجم‌های وسیعی از داده‌ها استفاده می‌کند. یک نمونه محبوب چارچوب Map Reduce است که یک فایل ورودی داده را به چند تکه تقسیم می‌کند. در اولین مرحله از Map Reduce ، یک نقشه ساز برای هر تکه، داده‌ها را می‌خواند، برخی محاسبات را انجام می‌دهد و فهرستی از جفتهای کلید/ مقدار را ارائه می‌دهد. در مرحله دوم یک کاهش دهنده ارزش‌های هر کلید را مجزا محاسبه کرده و نتایج را ارائه می‌دهد. دو ابزار عمده برای جلوگیری ازحملات وجود دارد:امنیت نقشه ساز و امنیت داده در حضور یک نقشه ساز نامطمئن.

### ۲-۲- بهترین عملیات امنیتی برای ذخایر داده غیر رابطه ای:

ذخایر داده غیر رابطه‌ای که توسط پایگاه داده NoSQL رواج یافتند از نظر زیرساخت امنیتی در حال تکامل هستند. توسعه-دهندگان استفاده از پایگاه‌های داده‌های NoSQL معمولاً امنیت را در میان افزار جاسازی می‌کنند. جنبه دسته‌بندی پایگاه-های داده‌هایNoSQL چالش‌های مازادی را متوجه قدرت چنین عملیات‌های امنیتی می‌کنند.

### ۲-۳- ذخیره سازی امن داده ها و ثبتهای تراکنشی

داده‌ها و لاگ های تراکنشی در رسانه‌های ذخیره‌سازی چندلایه ذخیره می‌شوند. انتقال دستی اطلاعات بین لایه‌های مختلف امکان کنترل مستقیم را در رابطه با نوع و زمان دقیق اطلاعات جابه‌جا شده، برای مدیر IT فراهم می‌کند. مقیاس پذیری و در دسترس بودن نیاز به لایه بندی خودکاربرای مدیریت ذخیره‌سازی داده‌های بزرگ مطرح می‌کند. لایه بندی خودکار مکان ذخیره‌سازی داده‌ها را مشخص نمی‌کند و چالش‌های جدیدی را متوجه ذخیره‌سازی امن داده‌ها می کند. روش‌های جدید باید دسترسی‌های غیرمجاز را محدود کرده و قابلیت دسترسی را حفظ کند.

### ۲-۴- اعتبارسنجی ورودی از نقاط پایانی/ فیلترینگ

استفاده از داده‌های بزرگ در محیط‌های شرکتی نیازمند جمع‌آوری داده‌ها از منابع متعددی هستند. یک چالش کلیدی در فرایند گردآوری داده،اعتبارسنجی ورودی است. چگونه می‌توان به این داده‌ها اعتماد کرد؟ چگونه می‌توان دشمن نبودن یک منبع ورودی را تعیین کرد؟ چگونه می‌توان از ورود اندیشه‌های بدخواه به مجموعه خود جلوگیری کرد؟

### ۲-۵- امنیت زمان واقعی /نظارت بر رعایت اصول

مونیتورینگ امنیت زمان واقعی با تولید تعداد زیادی هشدار که از دستگاه‌های امنیتی تولید شده همواره با چالش همراه بوده است. این هشدارها به یقین‌های اشتباه منجر می‌شود که اکثراً نادیده گرفته شده یا به راحتی کنار گذاشته می‌شوند. این مشکل برای داده‌های بزرگ با فرض حجم و سرعت جریان‌های داده افزایش می‌یابد. تکنولوژی داده‌های بزرگ امکان پردازش سریع و تحلیل انواع مختلفی از داده‌ها را فراهم می‌کند که می‌تواند برای تشخیص زمان واقعی آنومالی بر اساس تجزیه و تحلیل امنیتی مقیاس پذیر مورداستفاده قرار گیرد.

### ۲-۶- مقیاس پذیری و حفظ حریم خصوصی تجزیه و تحلیل ها و داده کاوی ها

داده‌های بزرگ می‌تواند باعث ایجاد حمله به حریم خصوصی، کاهش آزادیهای شهروندی و افزایش کنترل شرکت‌ها شود. بررسی جدیدی درباره چگونگی نفوذ شرکت‌ها در تحلیل داده‌ها برای بازاریابی محصولات شان انجام شده که چطور شرکتها از تمایلات و خواستهای مردم اطلاع پیدا می‌کنند. به همین صورت، ناشناس ماندن داده‌ها در تحلیل‌ها به منظور حفظ حریم خصوصی کاربر کافی نمی‌باشد. بنابراین تعیین دستورالعمل‌ها و توصیه‌ها برای جلوگیری از افشای غیرآگاهانه اطلاعات حریم خصوصی افراد حائز اهمیت است.

### ۲-۷- کنترل دسترسی رمزگذاری شده و ارتباطات امن

برای اینکه مطمئن شویم داده‌های حساس کاملاً امن بوده و فقط در دسترس نهادهای مجاز قراردارند، باید داده‌ها را بر اساس خط مشی‌های کنترل دسترسی پنهان کرد. برای تضمین احراز هویت، توافق وشفافیت در بین نهادهای پراکنده، باید چارچوبی جهت ارتباط امن و رمزگذاری شده ایجاد نمود.

### ۲-۸- کنترل دسترسی ریزدانه ای:

ویژگی امنیتی که از دیدگاه کنترل دسترسی مهم است جلوگیری از دسترسی به داده‌هایی است که افراد نباید دسترسی داشته باشند. شکلی که در خصوص مکانیسم‌های کنترل دسترسی ریزدانه‌ای وجود دارد، این است که داده‌هایی که قبلاً به طریق دیگری قابل اشتراک گذاری بودند، اغلب در دسته‌های محدودکننده‌تری قرار داده می‌شوند تا امنیت مطمئن‌تری ایجاد کنند. کنترل دسترسی ریزدانه‌ای به جای شمشیر، چاقوی جراحی به دست مدیران می‌دهد تا داده‌ها را بدون توجه به در خطر انداختن محرمانگی به اشتراک بگذارند.

### ۲-۹- ثبت عملیاتی ریزدانه ای

با نظارت امنیتی زمان واقعی تلاش می‌شود تا از وقوع حمله آگاه شوند. برای دسترسی به یک حمله انجام شده، نیاز به ثبت دقیق اطلاعات است. نه به این دلیل که می‌خواهیم از اتفاقاتی که رخداده و اشتباهاتی که انجام شده باخبر شویم، بلکه به دلایل رعایت مقررات قانونی و حفاظتی. دراین ارتباط، ثبت وقایع اتفاق جدیدی نیست و تنهاوسعت و ریزدانه‌ای بودن آن متفاوت است.

### ۲-۱۰- منبع داده

منبع فراداده‌ها به واسطه منشأ گراف‌های بزرگی که ناشی از محیط‌های برنامه‌نویسی موجود در کاربردهای داده‌های بزرگ است، پیچیده‌تر می‌شوند. تحلیل چنین مقادیری از گرافهای منبع، برای تشخیص وابستگی فراداده‌ها به امنیت/محرمانگی برنامه‌های کاربردی بر این محاسبات متمرکز است.

## ۳- معرفی روشهای برقراری امنیت برای داده های بزرگ

### ۳-۱- استفاده از الگوریتمهای پراکنده سازی اطلاعات امن:

الگوریتمهای پراکنده‌سازی اطلاعات( IDA ) تاکنون برای ذخیره‌سازی امن و قابل اعتماد و همچنین انتقال در سیستمهای توزیعی به کار رفته است. درواقع IDA متدی است که یک فایل F با اندازه‌ی |F| L را می‌گیرد و به n قطعه‌ی غیر قابل تشخیص تقسیم می‌کند که هرکدام اندازه‌ی l / m می‌باشند به گونه‌ای که فایل اصلی می‌تواند از m قطعه ساخته شود. هیچ سگمنتی از فایل اصلی با تعداد کمتری mاز قطعه قابل بازسازی نباشد. ولی درصورتیکه محرمانگی ضعیف باشد با دسترسی به تعداد کمتر ازm قطعه می‌توان به بخشی از فایل دسترسی پیدا کرد. برای دستیابی به محرمانگی بالا باید شرایط خاصی را در ماتریسهای تولیدکننده در این الگوریتم در نظر گرفت.سربار کاری به دلیل شکستن به قطعات مختلف افزایش خواهد داشت[1].

### ۳-۲- استفاده از رمزنگاری کوانتومی برای حفظ امنیت

رمزنگاری کوانتومی(QC)از تولید کلید به منظور احراز هویت پشتیبانی می‌کند و امنیت را با مدیریت کلید بین سرور احراز هویت و کاربران ذینفع در مراکز داده فراهم می‌کند. احراز هویت به وسیله تائید و اعتبارسنجی موجودیتهایی که در سرورها و کاربران وجود دارد از حریم خصوصی افراد حفاظت می‌کند.

### ۳-۳-استفاده از تحلیل‌های داده‌های بزرگ به منظور ارتقا امنیت

تحلیل‌هایی که بر روی داده‌های بزرگ انجام می‌شود، می‌تواند باعث بهبود امنیت اطلاعات باشد. برای مثال با تحلیل داده-های حاصل از LogFile ها و ترافیک شبکه می‌توان اشکالات را شناسایی کرد.

## ۴- امنیت داده های بزرگ در محاسبات ابری

محاسبات ابری، مخاطرات بسیاری را برای هر داده حساسی که با آن در تماس است ایجاد می‌کند، زمانی که مالکان داده‌ها کنترل داده‌هایشان را به محیط ابر می‌سپارند، نیازمند تضمین‌هایی برای حفاظت از داده‌هایشان هستند. امروزه این تضمین، وعده‌های قانونی است که ارائه کننده خدمات ابر به عنوان توافقنامه سطح خدمات به مالکان داده ارائه می‌کند. رمزنگاری به مالکان داده اجازه می‌دهد تا خودشان فعالانه برای محافظت از داده‌هایشان اقدام کنند به جای اینکه صرفاً بر قراردادهای

قانونی که معمولاً مشکلاتی در نظارت و اجرا دارند تکیه کنند. اقدامات لازم برای رسیدن به محرمانگی و تمامیت در ابر بستگی به سناریوهای زیر دارد:

*ابر غیرقابل اعتماد*: در این سناریو مالکان داده‌ها به ابر یا هر نود ابر برای حفظ محرمانه بودن و تمامیت داده و محاسبات آن اعتمادی ندارند، بنابراین حفاظت باید از طرف سرویس گیرنده برای اطمینان از محرمانه بودن و تمامیت در مواجه با یک ابر غیر قابل اطمینان صورت پذیرد.

*ابر قابل اعتماد*: در این سناریو، ابر در یک محیط Air-Gapped قرار دارد وکاملاً از هر شبکه خارجی و دشمنان مبراست. کاربران می‌توانند اطلاعات خود را در ابر قرار داده و اطمینان داشته باشند که به صورت محرمانه و به دوراز دشمنان خارجی باقی خواهد ماند. با این حال حتی در محیط ایزوله نیز ممکن است بعضی نودها خراب شده باشند. نودهای آسیب دیده می-توانند تمامیت داده و یکپارچگی محاسبات را به خطر اندازند. این سناریو معمولاً به ابر خصوصی اشاره دارد.

*ابر نیمه قابل اعتماد*: در این شرایط لازم نیست که مشتری به طور کامل به ابر اعتماد کند. فرض می‌شود بخشهایی از ابر هر زمانی تحت کنترل دشمن باشد اما تعدادی منابع از دسترس دشمن خارج است.

قسمت صدمه دیده به وسیله یک دشمن، تمام محاسبات و پروتکلها را همانند یک بخش درست انجام می‌دهد. ولیکن دشمن سعی در دریافت اطلاعات اضافی از طریق ترکیب مشاهدات از مجموعه قسمتهای آسیب دیده دارد. به طور خاص، یک دشمن خرابکار چندبخشی می‌تواند اطلاعاتی را به دست آورد که هیچ بخش تنهایی نمی‌تواند به صورت جداگانه به دست آورد. بخش آسیب دیده توسط دشمن خرابکار ممکن است خودسرانه از پروتکل‌های تعیین شده منحرف شود. به عنوان مثال: ارسال پیامهای ناقص ،تبانی با دیگر بخشها برای نقض محرمانگی و یکپارچگی. هدف از طراحی برای ایجاد محاسبات ابری امن، حفظ محرمانگی و تمامیت داده‌ها در حضور چنین دشمنانی است. تکنیک‌های رمزنگاری که به خصوص برای دستیابی به تجزیه وتحلیل امن داده‌های بزرگ استفاده می‌شوند عبارتند از:رمزنگاری هومومورفیک (HE )، محاسبات قابل اثبات (VC)، محاسبات چندگانه امن، رمزنگاری کاربردی، رمزنگاری مبتنی بر هویت و رمزنگاری مبتنی بر ویژگی. بااین حال روی تکنیک‌هایی تمرکز می‌شود که معتقدیم مناسبترین روش هستند. سه تکنیک رمزگذاری مطابق ابر قابل اعتماد، ابر غیرقابل اعتماد و ابر نیمه قابل اعتماد می‌باشد.

## ۵– چند نمونه کاربردهای داده های بزرگ در امنیت سیستمها

تجزیه و تحلیل داده‌های بزرگ می‌تواند امنیت را افزایش داده و نفوذپذیری به سیستم را کنترل کنند. بسیاری از راه‌حل‌های امنیتی به جلوگیری از حمله و نفوذ قبل از اینکه اتفاق بیفتد تمرکز دارد. روشهای متداول از پردازش لاگ و رویداد استفاده می‌کنند و با توجه به حجم بزرگ لاگهای شبکه به‌کارگیری فنهای داده‌های بزرگ برای پردازش آنها مناسبتراست. در ادامه چند نمونه از این سیستمها بیان شده است.

**۵–۱–کاربرد در سیستم‌های مدیریت رویدادهای امنیتی:** مدیریت اطلاعات امنیتی و رویدادها( SIEM )، از ترکیب مدیریت اطلاعات امنیتی (SIM) و مدیریت رویدادهای امنیتی(SEM) به وجود آمده است. تکنولوژی SIEM هشدارهای امنیتی که توسط سخت‌افزارها و نرم‌افزارها تولید می‌شوند را به صورت بلادرنگ پردازش می‌کنند. از SIEM برای ثبت داده‌های امنیتی و ایجاد گزارش استفاده می‌شود. SIEM، ترکیبی از قابلیت‌ها ازجمله تجمیع داده، همبستگی، هشدار دادن داشبورد، قبول، نگهداری و تجزیه وتحلیل قانونی است. تجمیع داده به معنای جمع‌آوری لاگ ها و رویدادها از منابع مختلف همانند شبکه، سرورها، پایگاه‌داده‌ها و نرم‌افزارهااست. همبستگی بخش اصلی SIEM را تشکیل می‌دهد و برای یکپارچگی بخش اصلی SIEM را تشکیل می‌دهد و برای یکپارچگی منابع مختلف و استخراج داده‌های مفید بکار می‌رود. سپس رویدادهای خروجی بخش همبستگی، تجزیه وتحلیل شده و هشدار تولید می‌شود. این هشدارها در ابزاری بانام داشبورد نمایش داده می‌شود.

در گام بعدی بر اساس هشدارهای مطابق با سیاست‌های امنیتی سیستم، گزارش تهیه می‌شود. تجزیه وتحلیل قانونی مربوط به توانایی جست‌وجو در لاگهای موجود، در گره‌های مختلف و در بازه زمانی خاص انجام می‌شود. برای تسریع در تکنولوژی SIEMاز فن‌های داده‌های بزرگ استفاده شده است BeeHive[۲]. دراین سیستم برای استخراج خودکار دانش از لاگهای شبکه ارائه داده است. چنین سیستمی در نرم‌افزارهای امنیتی کاربرد دارد. سیستم مذکور دارای سه فاز نرمال‌سازی، استخراج ویژگیها و خوشه‌بندی است. ابتدا داده‌ها توسط SEIM جمع‌آوری شده و نرمال می‌شوند. نرمال‌سازی شامل نرمال‌سازی برچسب زمان، نگاشت آدرس IP به نام میزان، تشخیص آدرسهای IP ثابت و میزبانهای اختصاصی است. لاگ ها از سیستم‌های مختلف جمع‌آوری می‌شوند و زمانها و فرمتهای متفاوتی دارند؛ بنابراین بایستی برچسب زمان داده‌های لاگ را به شکل استاندارد تغییر دهد. در فاز استخراج ویژگیها، ارتباط سازمان با شبکه‌های خارج از آن از طریق لاگهای شبکه استخراج می‌شود و شامل ویژگیهای مبتنی بر مقصد، میزبان، سیاست و ترافیک است. ویژگی مبتنی بر مقصد، ارتباط میزبان با مقصدهای جدید و خارج از شبکه بررسی می‌شود. در ویژگی مبتنی بر میزبان نرم‌افزارهایی که کاربر به تازگی نصب کرده است، تجزیه وتحلیل شده و در ویژگیهای مبتنی بر سیاست ارتباط میزان با سایتهای مسدود شده بررسی می‌گردد. از تکنیک خوشه‌بندی برای گروه‌بندی میزبانها استفاده شده و بدین صورت میزبان‌های آلوده شناسایی می‌شوند.

**۵-۲-کاربرد در سیستم‌های پیشگیری و تشخیص نفوذ:** به فرآیند استفاده از شبکه‌های کامپیوتری و داده‌های سیستم برای شناسایی حملات اینترنتی، تشخیص نفوذ گویند. مدیر شبکه بایدهشدار سیستمهای تشخیص نفوذ را تحلیل نماید. برای تسریع در تحلیل هشدارهای سیستمهای تشخیص نفوذ در بسیاری اوقات از روشهای مصورسازی اطلاعات و هشدارها استفاده می‌شود[3]. با توجه به حجم بزرگ داده‌های قابل پردازش در این سیستمها استفاده از فن‌های داده‌های بزرگ علاوه بر بهبودکارایی باعث افزایش قابلیت سیستمهای تشخیص نفوذ در سطح یک سیستمهای جامع با آگاهی وضعیتی می‌شود.

**۵-۳-کاربرد در سیستمهای هشدار زودهنگامEWS :** سیستم‌های EWS در اسرع وقت آسیب‌پذیری و یا حمله ناشناس را ثبت کرده و تجزیه وتحلیل می‌نمایند و اطلاعات حاصل و هشدارهای امنیتی را در اختیار مدیر امنیت شبکه قرار می‌دهند. این سیستم‌ها هشدارهای امنیتی را تولید کرده و الگوهای حمله و ریسک‌های بالقوه را تشخیص می‌دهند و بدین صورت ریسک‌های امنیتی را کاهش می‌دهند [5].

سیستم[5] LarSID با به اشتراک گذاشتن شواهد نفوذ در بین سیستم‌های تشخیص نفوذ، از حملات امنیتی شبکه جلوگیری می‌کند. سیستم مذکور از معماری جدول درهم سازی توزیع شده استفاده می‌کند. در ابتدا شواهد مشکوک از گره-ها در مناطق جغرافیایی مختلف به یکدیگر مرتبط می‌شوند. سرویس تشخیص نفوذ به صورت توزیع شده عمل می‌کند. در مکانیسم تشخیص، مؤلفه‌های DHT مسئول نگهداری شواهد مشکوک به آلودگی در زیرشبکه محلی و ایجاد پیام‌های اخطار در مورد آدرسهای IP مربوط به سیستم‌های درهم سازی است.

## ۶- بررسی انواع روشهای امنیتی موجود

مجموعه کاملی از تکنیکها و روشهای رمزگذاری وجود دارد که میتوان بعضاً کاملاً ایمن باشند؛ رمزنگاری سرقت داده‌ها را به طور قابل ملاحظه‌ای کاهش می‌دهد؛ ولیکن باعث کاهش کارایی می‌شود.بنابراین فقط بر روی داده‌هایی حساس رمزنگاری انجام می‌شود. رمزنگاری هنگام استفاده از سرویسهای ابر اهمیت بیشتری دارد و اگر ابر رمزگذاری شود، آنگاه مورد پذیرش وسیع کاربران قرارمی‌گیرد.

### ۱-۶-روش هوموورفیک:

منظور از رمزنگاری هوموورفیک الگوریتمهایی است که پس از اعمال شدن بر روی داده‌ها، همچنان امکان پردازش داده‌ها را فراهم می‌آورند. در این روش داده‌ها در مبدأ رمزگذاری شده و برای تجزیه و تحلیل به ابر فرستاده می‌شوند، توابع موجود در ابر عملیات موردنظر کاربر را انجام می‌دهند و نتایج را به مبدأ برمی‌گردانند. به عبارت دیگر شرکتهای ابر، می‌توانند در هنگام درخواست کاربر محاسباتی روی داده‌ها انجام دهند، بدون این‌که داده‌های اصلی بر آنها فاش شود.در این روش

محرمانگی داده‌ها حفظ می‌شود، ولی جامعیت داده‌ها تضمین نمی شود.از معایب این روش ناکارآمدی آن است. زیرا ازنظر اندازه فایل محدودیت داشته و اندازه متن رمزی و پیچیدگی عملیات رمزگذاری و رمزگشایی به اندازه فایل بستگی دارد. همچنین اطلاعات در عملیات جستجو به دلیل رمزدار بودن پیدا نخواهند شد. البته دولایه بودن رمزگذاری نیز باعث می‌شود سیستم بسیار کند کار کند[6].

### ۲-۶-محاسبات قابل اثبات:

به منظور اعتماد بیشتر به ابر، سرویس گیرندگان باید بتوانند صحت نتایج حاصله را بررسی کنند. در محاسبات قابل اثبات به روش پینوکیو[7] .هرکس می‌تواند با استفاده از کلید عمومی درستی اثبات را بررسی کند درنتیجه تمامیت داده‌ها بررسی می‌شود. این روش کارایی را بالا برده و باعث کاهش هزینه‌ها خواهد شد. ولی محرمانگی راتضمین نخواهد کرد. این تکنیکها هنوز هم برای اکثر برنامه های کاربردی که روی تجزیه وتحلیل داده‌های بزرگ کار می‌کنند بسیار کند است.

### ۶-۳-روش پیشنهادی جهت ارتقاء امنیت

برای محافظت از داده‌های ذخیره شده‌ی کاربران در برابر مهاجمان می‌توان، سرورهای ذخیره‌سازی در ابر را طبقه‌بندی کرد. سپس از طریق بکارگیری و استقرار موانع قوی در مسیر سرورها و بکارگیری سیستمهای شناسایی تهاجم به منظور شناسایی هرگونه فعالیت ناخواسته از سوی مهاجمان و سرورهای غیرمجاز از داده‌های سیستم محافظت کرد. مطابق با نیاز کاربران، دسترسی مجزا به سرورها می‌تواند از طریق به کارگیری شیوه تفکیک داده‌های ذخیره شده در سرورها انجام شود. دو حالت وجود خواهند داشت :اول، وقتی کاربران می‌خواهند داده‌ها را تنها برای استفاده خودشان در سرور ذخیره کنند و دوم، اگر کاربران مایل باشند داده‌ها را در ابری ذخیره کنند که خود می‌تواند به‌وسیله سایر کاربران مجاز نیز مورد دسترسی قرار گیرد .این دو نوع داده‌هایی کاربران روی سرورهایی مجزا ضبط خواهند شد و این سرورها هیچ ارتباطی باهم نخواهند داشت. وقتی کاربری قصد داشته باشد تا داده‌ها را تنها برای استفاده خودش در ابر ذخیره کند، آنگاه ایجاد اتصال از سمت خارج با سرورهای ابر، مستقیماً با این سرور مربوط به این داده‌ها در ابر برقرار نخواهد شد، باید فایروالی دراین بین قرار گیرد تا از ایجاد اتصالات مستقیم با سرورهای داخلی ابر ممانعت کند، البته فقط سرورهایی که روی آنها داده‌های خصوصی کاربران ذخیره می‌شوند. این کار آدرسهای IP سرورها را پنهان می‌کند و به این ترتیب سرورها می‌توانند در برابر هکرها محافظت شوند. از این طریق ارائه کننده خدمات ابر می‌تواند داده‌های خصوصی کاربر محافظت کند.

## ۶-نتیجه‌گیری

در این مقاله چالش‌های مختلف فنآوری داده‌های بزرگ از دید امنیت و حفاظت از حریم خصوصی مورد بررسی قرار گرفت و فعالیتهای مختلفی را که در این حوزه توسط محققین انجام شده بود، معرفی شد. به عنوان نمونه می‌توان روشهای استفاده از تحلیل داده‌های بزرگ برای تشخیص مشکلات، پراکنده‌سازی اطلاعات و یا استفاده از روشهای مبتنی بر برقراری امنیت در ابر و متدهای مختلف رمزگذاری همانند رمزگذاری ریزدانه ای، ...را نام برد. هرچند این روشها تا حدودی برخی از مشکلات امنیتی در این حوزه را برطرف کرده ولی هرکدام از دیدگاه خود و در سطحی خاص به این امر پرداخته‌اند ولی وجود این حجم داده‌ها و تنوع و بی ساختار بودن آنها همواره مانعی برای ارائه یک شیوه کامل و جامع بوده است؛ و هرکدام دارای مزایای و معایبی هستند و تلاش می‌شود تا با ارائه راهکارهایی برخی از این معایب برطرف گردند.

## ۷-مراجع


1. M. Li, "On the Confidentiality of Information Dispersal Algorithms and Their Erasure Codes," in Proc. CoRR, pp. 1-4abs/1206.4123, 2012
2. T.-F. Yen, A. Oprea, K. Onarlioglu, T. Leetham, W. Robertson, A. Juels and E. Kirda, "Beehive: Large-Scale Log Analysis for Detecting Suspicious Activity in Enterprise Networks," in ACSAC, Proceedings of the 29th Annual Computer Security Applications Conference, New Orleans, Louisiana, USA, pp. 199-208, 2013

3. رضا عزمی و همکاران" بررسی چالشهای امنیتی داده‌های عظیم و چند نمونه کاربردهای امنیتی"**همایش داده‌های عظیم، تهران** ۱۳۹۳ صفحات۱۹۴ - ۱۸۴

4. P.-S. Huang, C.-H. Yang and T. Ahn, "Design and Implementation of A Distributed Early Warning System Combined with Intrusion Detection System andHoneypot," in International Conference on Hybrid Information Technology, pp. 232-238, 2009.
5. C. Vincent Zhou, S. Karunasekera and C. Leckie, "Evaluation of a Decentralized Architecture for Large Scale Collaborative Intrusion Detection,"International Symposium on Integrated Network Management (IM '07), 10th IFIP/IEEE, pp. 80-89, 2007.
6. Z. Brakerski, “Fully homomorphic encryption without modulus switching from classical GapSVP,” in CRYPTO, pp. 1-19, 2012.
7. B. Parno, J. Howell, C. Gentry, and M. Raykova, “Pinocchio: Nearly practical verifiable computation,” in Proceedings of the 2013 IEEE Symposium on Security and Privacy, pp. 238-252, 2013.
8. Y. Li, K. Gai, L. Qiu, M. Qiu, and H. Zhao, “Intelligent cryptography approach for secure distributed big data storage in cloud computing,” Inform. Sci., vol. 387, pp. 103–115, May 2017.