کنترل دسترسی مبتنی بر محتوا

نویسنده :

دبلیو . زنگ

مترجمان :

دکتر محمدعلی ترکمانی

فریبا سرهنگ نیا

انتشارات ارسطو

(چاپ و نشر ایران)

۱۴۰۱

سرشناسه : دزنگ، ونرونگ Zeng, Wenrong
عنوان و نام پدیدآور : کنترل دسترسی مبتنی بر محتوا/ نویسنده دبلیو زنگ؛ مترجمان محمدعلی ترکمانی، فریبا سرهنگ‌نیا.
مشخصات نشر : ارسطو (سامانه اطلاع رسانی چاپ و نشر ایران) ، ۱۴۰۱.
مشخصات ظاهری : ۱۷۲ ص.: مصور، جدول، نمودار (رنگی).
شابک : ۶-۸۸۷-۴۳۲-۶۰۰-۹۷۸ : ۷۰۰۰۰۰ریال
وضعیت فهرست نویسی : فیپا
یادداشت : عنوان اصلی: Content-based access control, ۲۰۱۵.
یادداشت : کتابنامه: ص. [۱۴۵] - ۱۵۷.
موضوع : پایگاه‌های اطلاعاتی -- کنترل دستیابی
Databases -- Access control
مدارک پزشکی -- کنترل و دستیابی
Medical records -- Access control
شناسه افزوده : ترکمانی، محمدعلی، ۱۳۵۴-، مترجم
شناسه افزوده : سرهنگ‌نیا، فریبا، ۱۳۶۸-، مترجم
رده بندی کنگره : QA۷۶/۹
رده بندی دیویی : ۰۰۵/۷۴
شماره کتابشناسی ملی : ۸۸۷۵۳۱۴
اطلاعات رکورد کتابشناسی : فیپا

نام کتاب : کنترل دسترسی مبتنی بر محتوا
نویسنده : دبلیو . زنگ
مترجمان : دکتر محمدعلی ترکمانی - فریبا سرهنگ نیا
ناشر : ارسطو (سامانه اطلاع رسانی چاپ و نشر ایران)
صفحه آرایی، تنظیم و طرح جلد : پروانه مهاجر
تیراژ : ۱۰۰۰ جلد
نوبت چاپ : اول - ۱۴۰۱
چاپ : مدیران
قیمت : ۷۰۰۰۰ تومان
فروش نسخه الکترونیکی - کتاب‌رسان :
https://chaponashr.ir/ketabresan
شابک : ۶-۸۸۷-۴۳۲-۶۰۰-۹۷۸
تلفن مرکز پخش : ۰۹۱۲۰۲۳۹۲۵۵
www. chaponashr. ir

انتشارات ارسطو
چاپ و نشر ایران
Chaponashr.ir

تقدیم به:

پدر عزیزم

فریبا سرهنگ‌نیا

فهرست مطالب

فصل چهارم: مدل کنترل دسترسی مبتنی بر محتوا ۵۷

فصل پنجم: اجرای CBAC ۶۹

فصل ششم: استراتژی‌های بهینه سازی CBAC ۹۱

فصل هفتم: بهبود برچسبزنی با یادگیری چند برچسبی (MLL) ۱۱۳

مقدمه مترجم:

مدل های کنترل دسترسی پایگاه داده و همینطور مکانیزم های اجرایی این موضوع را مشخص و ایجاب می کند که «چه کسی می تواند به چه چیز دسترسی داشته باشد».

در مدل های مرسوم کنترل دسترسی به پایگاه داده، مدیران پایگاه داده (DBA ها) و یا کاربران یا صاحبان داده به طور ویژه حقوق دسترسی هر مورد داده ای را برای هریک از نقش ها از طریق «اعطا» یا «لغو» حقوق مشخص برای هر نقش، تعیین می کنند. با این وجود، به دلیل رشد تصاعدی داده ها، خصوصاً در مورد داده‌های محتوا محور ، چنین رویکردهایی ممکن است مناسب و حتی عملی نباشد.

به عنوان مثال یک فرمانده پلیس پایگاه داده ای از پرونده های بسیار حساس در اختیار دارد. وی پرونده ای را به یکی از ماموران خود می سپارد. طبیعتاً، فرمانده باید دسترسی به تمام پرونده های مشابه را به مامور بدهد. در این حالت، مفهوم «پرونده های مشابه» از طریق تشابه معنایی محتوای رکوردهای ثبت شده تعیین می شود، که این تشابه می تواند از نظر جغرافیایی، زمانی، روشی، و یا در توضیحات متنی رکوردهای پرونده باشد. علاوه بر این، هنگامی که پرونده های جدید به پایگاه داده اضافه می شود، پرونده های که شبیه به پرونده اول هستند باید به طور اتوماتیک و بدون نیاز به دخالت بیش تر فرمانده پلیس در دسترس مامور قرار داده شوند.

امروزه در تمامی کشورها اشتراک گذاری اطلاعات مراقبت های بهداشتی به طوری سختگیرانه اجام می شود. سوابق پزشکی به خوبی توسط ارائه دهندگان مراقبت بهداشتی حفاظت شده و تنها تحت قوانینی سخت گیرانه اشتراک گذاری می شوند. با این وجود پزشکان، پرستاران و محققان معمولاً دارای مجوزهای دسترسی گسترده تری هستند. در اینگونه سیستمها نیز مدل های کنترل دسترسی پایگاه داده در موارد اشتراک داده محتوا محور دچار مشکل می شوند. در چنین مواردی،انتظار می رود یک مدل کنترل دسترسی جدید پدیدار شود تا مطابق با نیازهای ایجاد تصمیمات دسترسی بر اساس شباهت های معنایی محتوایی داده ها باشد.

در این کتاب ما مدل کنترل دسترسی مبتنی بر محتوا و مکانیزم های اعمال آن را ارائه بررسی می‌کنیم، که در آن مجوزهای دسترسی بر اساس شباهت واژگانی[1]میان اعتبارنامه درخواست

1 lexicon similarity

کننده و رکوردهای درخواست شده اعطا می شود. این مدل جدید، بعنوان مکملی برای رویکردهای کنترلی موجود، ابزارهایی کارآمد و بهینه برای کنترل دسترسی ارائه می دهد که از خصوصیات محتوایی در اشتراک گذاری داده های غنی از محتوا بهره برده و منجر به اولین اقدامات جهت رفع مشکلات در مورد کنترل دسترسی پایگاه داده محتوا محور در زمینه داده های کلان می شود. مدل کنترل دسترسی داده محور[1] بررسی شده در این کتاب، مدل کنترل دسترسی مبتنی بر محتوا (CBAC) نام دارد و از محتوای داده ها بهره می برد تا به مفاهیم کنترل دسترسی انعطاف‌پذیر تر و قوی تری در مورد پایگاه داده های محتوا محور در اشتراک گذاری اطلاعات دست یابد. CBAC اولین اقدام جهت ساخت یک مدل کنترل دسترسی است که مفهوم امنیت تقریبی[2] را معرفی نموده و قادر به رسیدگی به شرایطی است که در آن سیاست‌های کنترل دسترسی صریحی در اختیار نیستند. در CBAC، از روش‌های یادگیری ماشین (مانند تکنیک های استخراج متن) برای مدل کردن کنترل دسترسی و اعمال آن استفاده می‌شود. با معرفی این روش ها، قاعده کنترل دسترسی به صورت کاربردهای الگوریتمی تبدیل می شود، و در همین راستا، خواص پویا، خودکارسازی و هوشمندی را با استفاده از تمام این تکنیک ها در مدل های کنترل دسترسی بهبود می‌دهد.

از صنعتگران، اساتید و دانشجویان عزیز تقاضا دارم نقطه نظرات خود را از طریق ایمیل m.a.torkamani@gmail.com با مترجمان در میان بگذارند تا انشالله در ویرایش‌های بعدی اشکالات یا کاستی‌های احتمالی کتاب مورد تجدید نظر قرار گیرد. در پایان وظیفه خود می‌دانم از از مدیریت انتشارات ارسطو و سامانه اطلاع رسانی چاپ و نشر ایران جناب آقای حسین قنبری به خاطر مساعدت در کار چاپ تشکر و قدردانی نمایم.

محمدعلی ترکمانی

تابستان ۱۴۰۱

۱data-driven access control model

۲approximate security

چکیده

در پایگاه داده عادی، محبوب ترین مدل کنترل دسترسی سیاست‌ها برای هریک از نقش های هر کاربر را صراحتا در قبال هر مورد داده ای به صورت دستی معین می کند.امروزه، در اشتراک گذاری مقیاس بزرگ داده به مرکزیت محتوا[1] ممکن است رویکردهای مرسوم بدلیل رشد تصاعدی داده و همینطور حساسیت اشیاء داده ای[2] غیر عملی باشند. علاوه بر این، سیاست مرسوم کنترل دسترسی به پایگاه داده، هنگامی که محتوای معنایی داده قرار است نقشی در تصمیمات دسترسی داشته باشد، عملی نخواهد بود. کاربران معمولاً امکان دسترسی بیش از حدی دارند، و از حسابرسی مربوط به اصول گذشته جهت شناسایی سوء استفاده از این مجوزهای دسترسی استفاده می شود. متاسفانه، اغلب اوقات جبران خسارت دشوار است، چرا که (مقدار بسیار زیادی از) داده‌ها دیگر تا آن زمان فاش شده اند. در این نوشتار، ما ابتدا کنترل دسترسی مبتنی بر محتوا (CBAC)[3] را معرفی می‌کنیم، که یک مدل کنترل دسترسی نوآورانه برای اشتراک گذاری اطلاعات به مرکزیت محتوا است. به عنوان مکملی برای مدل های مرسوم کنترل دسترسی، مدل CBAC تصمیمات کنترل دسترسی را به صورت اتوماتیک بر اساس تشابه محتوا[4] میان اعتبارنامه های کاربر[5] و محتوای داده[6] اتخاذ می‌کند. در CBAC، هر کاربر به وسیله یک فرا قاعده[7] امکان دسترسی به زیردسته‌ای از اشیاء داده‌ای نشان داده شده مربوط به پایگاه داده مبتنی بر محتوا[8] را دارد، در حالی که مرز این زیردسته به طور دینامیکی توسط محتوای متنی اشیاء داده ای معین می‌شود. سپس ما یک مکانیزم اجرایی برای CBAC ارائه می دهیم که از پایگاه داده ی خصوصی مجازی Oracles (VPD)[9] جهت ایجاد یک کنترل دسترسی خطی گونه[10] و همینطور پیش گیری از سوء استفاده از اشیاء داده ای از طریق مجوز دسترسی غیر

---

1 large-scale content-centric data sharing
2 data objects
3 Content-Based Access Control
4 content similarity
5 user credentials
6 data content
7 Meta rule
8 content-centric database
9 Virtual Private Database
10 row-wise access control

ضروری، بهره می برد. در راستای بهبود بیش تر عملکرد رویکرد ارائه شده، ما یک مکانیزم بلوک کردن مبتنی بر محتوا[1]معرفی می کنیم تا کارآیی اجرای CBAC را بهبود بخشیم تا در مقایسه با استفاده صرف از اعتبارنامه های کاربر و محتوای داده ای، بخش هایی مرتبط تر از اشیاء داده ای را آشکار سازد. همچنین ما از چندین مکانیزم برچسب زنی[2]جهت تطابق دقیق تر محتوای متنی در مورد تکه های متنی کوتاه (مثلا صفات کوتاه VarChar) بهره بردیم تا برای ارائه محتوای داده به جای رخدادهای کلامی خالص، موضوعات[3]را استخراج نماییم. در این مکانیزم برچسب زنی، تشابه داده بدون وابستگی صرف به رخدادهای کلامی بلکه با موضوعات معنایی تحت محتوای متنی محاسبه شده است. نتایج تجربی نشان می دهد که CBAC تصمیمات کنترل دسترسی دقیقی با میزان خطای کم اتخاذ می کند.

1 content-based blocking mechanism
2 tagging mechanisms
3 topics

فصل اول

مقدمه

۱-۱-مقدمه

به بیانی ساده، مدل های کنترل دسترسی پایگاه داده و همینطور مکانیزم های اجرایی این موضوع را مشخص و ایجاب می کند که «چه کسی می تواند به چه چیز دسترسی داشته باشد». در اینجا، «چه کسی» بیانگر مجموعه‌ای از کاربران یا نقش ها، و «چه چیزی» بیانگر مجموعه‌ای از اشیاء داده ای (مثلا چندتایی ها یا گره های XML، مشخصه های پایگاه داده ی SQL) است. در مدل های مرسوم کنترل دسترسی به پایگاه داده، مدیران پایگاه داده (DBA ها) و یا کاربران یا صاحبان داده به طور ویژه حقوق دسترسی هر مورد داده ای را برای هریک از نقش ها از طریق «اعطا»[1] یا «لغو»[2] حقوق مشخص برای هر نقش، تعیین می کنند. با این وجود، به دلیل رشد تصاعدی داده ها، خصوصاً در مورد داده‌های محتوا محور ، چنین رویکردهایی ممکن است مناسب و حتی عملی نباشد. این موضوع سه دلیل دارد. اولاً، این رویکرد به وسیله مشخصات داده های محتوا محور معین شده است. داده های محتوا محور معمولاً حاوی مقدار زیادی متن آزاد[3] است. برای مثال، رکورد سلامت الکترونیکی (EHR)[4] نوعی داده ی محتوا محور است. در EHR،

1 GRANT
2 REVOKE
3 free text
4 electronic health record

پزشکان علاوه بر فهرست کردن اطلاعات بیماران (مانند نام، جنسیت، سن و غیره) نشانه های بیماری هریک از آن ها را نیز تشریح می کنند. به جای استفاده از کلماتی دقیق جهت توضیح نشانه های بیماری فرد، یک پزشک معمولاً از روش هایی تشریحی تر برای ثبت نشانه های بروز داده شده از طرف بیمار استفاده می کند. به همین دلیل است که داده ی EHR بیش تر داده ای متن آزاد است تا اینکه داده ای قالب بندی[1]شده باشد. متن آزاد، همانطور که مثال نشان داد، قادر به بیان مفهوم معنایی مشابه با توزیع متفاوتی از عبارات است. دوماً، در پایگاه داده ی محتوا محور، از محتوای داده ای انتظار ایفای نقشی در اتخاذ تصمیمات کنترل دسترسی می رود. به ادامه ی مثال EHR بازگردیم. پیش از ارائه تصمیمی نهایی در مورد اینکه مشکل بیمار چیست، پزشکان ممکن است نیاز به مطالعه ی سوابق سایر بیماران با نشانه های بیماری مشابه، خصوصاً در مورد بیماری های غیر عادی، داشته باشند. در مورد چنین شرایطی، ممکن است توصیف صریح حقوق دسترسی برای مقدار زیادی اشیاء داده ای، خصوصاً هنگامی که تصمیمات بر اساس محتوا هستند، دشوار باشد (اینکه از یک مدیر سیستم بخواهیم تمام سوابق موجود در پایگاه داده را به طور دستی بررسی نموده و حقوق دسترسی برای هر کاربر یا نقش اختصاص دهد، کار بسیار زیادی می برد). سوماً، در محیط های پراکنده و پویا، تعیین صریح حقوق دسترسی برای هر کاربر از همتاهای راه دور[2]دشوار است؛ مثلا یک سازمان می تواند به سادگی وظایفی جدید ایجاد نماید بدون اینکه شرکای خود را در جریان بگذارد، که این مورد در اشتراک گذاری اطلاعات زیاد رخ می دهد. در این مورد، تصمیمات کنترل دسترسی می تواند مبتنی بر دانش درخواست کننده ی راه دور باشد که به طوری پویا تمام پرس و جو ها ارسال می‌شود. در همین حین، در اشتراک گذاری توزیع شده اطلاعات، صاحبان داده ممکن است فقط با افرادی قصد اشتراک گذاری داشته باشند که داده های مشابه ارائه می دهند؛ این موضوع می تواند نشانگر این باشد که آن افراد نیز به دلیل حساسیت محتوای داده ای تمایلات داده ای مشابهی دارند، اما صاحبان داده تا زمانی که به محتوای داده های سایرین پی نبرند قادر به تعیین قواعد کنترل دسترسی نخواهند بود. برای درک بیشتر انگیزه این پژوهش، نگاهی به مثال زیر داریم:

1 formatted text
2 remote peers

مثال ۱: یک عامل اجرای قانون (مثلا FBI) پایگاه داده ای از پرونده های بسیار حساس در اختیار دارد. یک مدیر یعنی Bob پرونده ای را به مامور Alice می سپرد. طبیعتاً، این سرپرست باید دسترسی به تمام پرونده های مشابه را به Alice بدهد. در این حالت، مفهوم «پرونده های مشابه» از طریق تشابه معنایی محتوای رکوردهای ثبت شده تعیین می شود، که این تشابه می تواند از نظر جغرافیایی، زمانی، روشی، و یا در توضیحات متنی رکوردهای پرونده باشد. علاوه بر این، هنگامی که پرونده های جدید به پایگاه داده اضافه می شود، پرونده های که شبیه به پرونده ی Alice هستند باید به طور اتوماتیک و بدون نیاز به دخالت بیش تر سرپرست در دسترس Alice قرار داده شوند. برای مثال، پرونده های جدید اضافه شده ی مربوطه ممکن است کلیدی اساسی در مورد پرونده ی تحت بررسی باشند. متاسفانه، در الگوی موجود کنترل دسترسی به پایگاه داده، چنین توصیفی برای کنترل دسترسی پشتیبانی نشده است. در حالی که، بررسی دستی تمام رکوردها توسط سرپرست جهت اعطا یا لغو دسترسی کار بسیار زیادی می برد. در عمل، اغلب مدل امنیتی چند سطحی (MLS)[1] به کار گرفته می شود و هریک از ماموران به تعداد زیادی از رکوردها دسترسی می یابند (هر آنچه با سطح امنیتی وی برابر یا پایین تر از آن باشد). به طور مشابه، بسیاری از شرکت های پردازش محتوا (مثلاً پردازش نظرسنجی و همینطور شرکت های تجارت از راه دور) به دلیل عدم توانایی در اجرای کنترل دسترسی بر اساس محتوای متنی رکوردها، به همه ی کارکنان اجازه دسترسی به تمام رکوردها ی مشتریان در پایگاه داده ی خود (که می توانند حساس باشند) را می دهند. در تمام این سناریو های مشابه، اشتراک گذاری اطلاعات ممکن است یا خیلی محتاطانه باشد و یا بدلیل افشای غیرضروری اطلاعات مورد سوء استفاده قرار گیرد.

مثال ۲. در سیستم های اشتراک سنتی، کاربران برای دسترسی به تمام نشریات هزینه پرداخت می کردند. برای نمونه، یک پژوهشگر متمایل به «امنیت اطلاعات» ممکن است در «تراکنش های IEEE در مورد مهندسی دانش و داده» مشترک شود، علی رغم اینکه تنها مایل به بخش کوچکی از مقالات موجود در ژورنال است. رویکردی دیگر می توانست این باشد که هر کاربر در مجموعه‌ای از برچسب ها مشترک شود، و هر مقاله (بعنوان یک رکورد در پایگاه داده) با کلیدواژه

1 Multi-Level Security

ها برچسب زنی شود. از این رو، تصمیمات کنترل دسترسی می توانست از طریق مطابقت دهی برچسب های کاربر و برچسب های مقاله گرفته شود. با این وجود، چنین رویکردی با دو اشکال جدی روبه رو است: (۱) کیفیت برچسب زنی در مورد این رویکرد حیاتی است، اما کنترل کیفیت مسئله ای مبهم است؛ (۲) تعداد مقاله های قابل دسترس ممکن است بسیار زیاد یا خیلی کم باشد، مثلاً مقاله ای دارای یک برچسب، شاید تا حدود کمی مربوط به موضوع آن برچسب باشد. در یک راه حل مطلوب، از مشترک انتظار می رود که تمایلات خود را به صورت یک توصیف متنی ارسال نماید و یا برخی مقالات بذر[1]را تعیین کند (مثلا مقاله خودش را)، و سپس به مقاله هایی با محتوای مشابه دسترسی پیدا کند. در راه حل ایده ال، سیاست‌های کنترل دسترسی اعطا شده بر اساس شباهت محتوا می تواند راندمان کاری کاربران را بر اساس مقاله های انتخابی واجد شرایط بهبود بخشد.

مثال ۳: در سناوریوهای اشتراک گذاری اطلاعات توزیع شده، برخی از صاحبان داده نوشته های خود را تنها با همتا هایی به اشتراک می گذارند که داده هایی مربوط ارائه دهند، از این رو اشتراک گذاری سودمند خواهد بود. برای نمونه، در پروژه ای مشترک با «دپارتمان سرپرستی عمومی» که در مورد مشارکت شهروندی تحقیق می کند، محققان نظرات خود را با دیگرانی که نظرات مشابهی دارند اشتراک می گذارند. در این مورد، نظرات بوسیله یک پاراگراف متنی کوتاه ارائه می شوند.در سایر حوزه های تحقیقاتی علمی، شاهد این هستیم که محققان حتی داده های تحقیقی را نیز (در قالب منبعی اشتراکی و کنترل دسترسی شده) با سایر همکاران خود که داده های مشابه ارائه می کنند، به اشتراک می گذارند. اجازه دهید نگاهی دوباره به مثال ۱ داشته باشیم: هنگامی که FBI با سایر عوامل اعمال قانون همکاری می کند (مثلا CIA)، آن ها فقط «پرونده های مربوط» را اشتراک گذاری می کنند، در حالی که روابط پرونده ها فقط با شباهت معنایی محتوا ارزیابی شده است. حفاظت از حریم خصوصی در حین تطابق تشابه مدارک (Murugesan et al. (2010); Scannapieco et al. (2007)) برای تعیین و اشتراک گذاری مدارک مشابه استفاده شده است. با این وجود، در موردی که FBI قصد دارد پرونده ای را به خاطر شباهت با یکی از پرونده های CIA آشکار سازد، یک راه حل دیگر این است که

1 seed articles

دسترسی به پایگاه داده جهت ممکن ساختن دسترسی CIA به «پرونده های مشابه» به کار گرفته شود.

مثال ۴: اشتراک گذاری اطلاعات مراقبت های بهداشتی به طوری سختگیرانه توسط HIPAA اداره می شود. سوابق پزشکی به خوبی توسط ارائه دهندگان مراقبت بهداشتی حفاظت شده و تنها تحت قوانینی سخت گیرانه اشتراک گذاری می شوند. با این وجود، در داخل تشکیلات، کاربران (دکترها، پرستاران، محققان) معمولاً دارای مجوزهای دسترسی گسترده تری هستند، در حالی که حسابرسی های بعدی برای تشخیص و تنبیه سوءاستفاده از مجوزها اعمال می شود (Appari &Johnson (2010); Malin et al. (2007); Boxwala et al. (2011); Rostad & Edsberg (2006)). یک راه حل اعمال شده دیگر مکانیزم «شیشه را بشکن (BTG)»[1] را به کار می برد تا به کاربران امکان دهد قوانین کنترل دسترسی را به صورت کنترل شده و تحت شرایط ویژه زیر پا بگذارند ((Ferreira et al. (2006). حسابرسی های بیش تر هنگامی که یک کاربر از سیاست BTG استفاده کند به کار گرفته خواهد شد.

از نمونه های آورده شده می توان مشاهده نمود که مدل های قطعی فراگیر کنترل دسترسی پایگاه داده در موارد اشتراک داده محتوا محور دچار مشکل می شوند. در چنین مواردی،انتظار می رود یک مدل کنترل دسترسی جدید پدیدار شود تا مطابق با نیازهای ایجاد تصمیمات دسترسی بر اساس شباهت های معنایی محتوایی داده ها باشد. یکی دیگر از قابلیت های مطلوب چنین مدل کنترل دسترسی مبتنی بر محتوا این است که شباهت محتوای معنایی باید به صورت مبانی ذاتی ارائه شده توسط RDBMS سنجیده شده و تنها نیاز به حداقل دخالت از سوی مدیران پایگاه داده (DBA ها) دارد. در این نوشتار، ما اولین اقدامات در راستای این مسیر را انجام می دهیم: ما مدل کنترل دسترسی مبتنی بر محتوا و مکانیزم های اعمال آن را ارائه می دهیم، که در آن مجوزهای دسترسی بر اساس شباهت واژگانی[2] میان اعتبارنامه درخواست کننده و رکوردهای درخواست شده اعطا می شود. این مدل جدید، بعنوان مکملی برای رویکردهای کنترلی موجود، ابزارهایی کارآمد و بهینه برای کنترل دسترسی ارائه می دهد که از خصوصیات محتوایی در اشتراک گذاری داده های غنی از محتوا بهره برده و منجر به اولین اقدامات جهت رفع مشکلات در

1 break the glass
2 lexicon similarity

مورد کنترل دسترسی پایگاه داده ی محتوا محور در زمینه ی داده های کلان[1] می شود. در این نوشتار، ما نیازهای جدید امنیتی و حریم خصوصی در سیستم های اطلاعاتی توزیع شده را بررسی نموده و تصمیم بر این داریم که چنین مسائلی را با طراحی های نوآورانه حل کنیم. از این رو، ما در این نوشتار به صورت رسمی یک مدل کنترل دسترسی داده محور[2] به نام مدل کنترل دسترسی مبتنی بر محتوا (CBAC) را ارائه می دهیم که از محتوای داده ها بهره می برد تا به مفاهیم کنترل دسترسی انعطاف‌پذیر تر و قوی تری در مورد پایگاه داده های محتوا محور در اشتراک گذاری اطلاعات دست یابد. CBAC اولین اقدام جهت ساخت یک مدل کنترل دسترسی است که مفهوم امنیت تقریبی[3] را معرفی نموده و قادر به رسیدگی به شرایطی است که در آن سیاست‌های کنترل دسترسی صریحی در اختیار نیستند. در CBAC، تصمیم گرفتیم از روش های یادگیری ماشین (مانند تکنیک های استخراج متن[4]) برای مدل کردن کنترل دسترسی و اعمال آن استفاده نماییم. با معرفی این روش ها، قاعده کنترل دسترسی به صورت کاربردهای الگوریتمی تبدیل می شود، و در همین راستا، ما قصد داریم خواص پویا، خودکارسازی و هوشمندی را با استفاده از تمام این تکنیک ها در مدل های کنترل دسترسی بهبود بخشیم.

1 big data
2 data-driven access control model
3 approximate security
4 text mining

فصل دوم

کارهای مرتبط

فناوری رایانه ای زندگی روزمره شامل تحصیلات، زندگی کاری، و سرگرمی افراد را تغییر داده است. این فناوری یافتن اطلاعات در مورد دانش،پیدا کردن شغل، لذت بردن از فیلم ها و موسیقی های خوب را آسان ساخته است. در عین حال، این تکنولوژی روش اداره شرکت ها را نیز از نظر جست و جو برای تامین کنندگان جهت مقایسه پیشنهادات، گردآوری، ذخیره سازی و ارائه ی اطلاعات مربوط به محصولات، و حفظ رابطه نزدیک با مشتری، دستخوش تغییراتی کرده است. فناوری رایانه ای نه تنها راندمان زندگی روزمره افراد را بهبود بخشیده است، بلکه چگونگی ایجاد ، پردازش، انتقال، ذخیره سازی و پنهان سازی اطلاعات را نیز تغییر داده است. امروزه، یکی از مهم ترین مشکلات امنیتی پیش گیری از دسترسی غیر مجاز به اطلاعات است، که در واقع از دسترسی به اطلاعات اعتباری که شخص اجازه دسترسی به آن را ندارد جلوگیری می کند. خطرات رایج در مورد دسترسی غیر مجاز شامل موارد زیر شده اما به آن ها محدود نمی شود:

- افشای غیرمجاز اطلاعات
- قطع خدمات رایانه ای
- از دست رفتن بهره وری که فعالیت های معمول رایانه ای را در عملیات های حساس به زمان به تاخیر می اندازد
- ضررهای مالی، همچون خدشه دار شدن اطلاعات و قطع خدمات
- پیامدهای قانونی به دلیل دادخواهی از سوی سرمایه گذاران، مشتریان، و یا عموم
- باج خواهان مزاحم که با تهدید سیستم امنیتی از شرکت پول اخاذی می کنند

جهت اجتناب از این خطرات، محققان مدل های کنترل دسترسی مختلفی بر اساس این الگو ایجاد نمودند که "چه کسی" حق دسترسی به "چه چیز" را دارد. در این فصل، ما برخی مدل های کنترل دسترسی رایج را برای معرفی انتخاب می کنیم. شکل ۲-۱ از شکل ۱ (NIST (2009)) بدست آمده تا رابطه میان این مدل ها را نشان دهد. ما فهرست مدل های کنترل دسترسی را دنبال می کنیم ((NIST (2009) و همینطور جزئیات بیش تری در مورد مدل هایی که پایه تعریف ریاضی قوی دارند اضافه می کنیم.

تحقیقات مربوط به کنترل دسترسی پایگاه داده را تقریباً می توان به صورت «مدل های کنترل دسترسی» و «اجرای کنترل دسترسی»[1]دسته بندی نمود. مدل های کنترل دسترسی مرتبط را می توان به صورت های زیر دسته بندی کرد: «کنترل دسترسی اجباری»[2](Jajodia & Sandhu (1991); Sandhu (1993); Sandhu & Chen (1998); Winslett et al. (1994); McCune et al. (2006); Lindqvist (2006); Thuraisingham (2009); Upadhyaya (2011))، «کنترل دسترسی اختیاری» (DAC)[3](Moffett et al. (1990); Thomas et al. (1993); Ahn (2009);Downs et al. (1985))و کنترل دسترسی نقش مبنا[4] (RBAC) (Ferraiolo et al. (2001); Osborn et al. (2000); Sandhu et al. (1996)).

1 access control enforcement
2 Mandatory access control
3 Discretionary access control
4 role-based access control