مجله نخبگان علوم و مهندسی

Journal of Science and Engineering Elites

جلد ۳- شماره ۲- سال ۱۳۹۷

www.ElitesJour

بررسی و تحلیل امنیت ترافیک شبکه با استفاده از فناوری داده های بزرگ

سید ابوالفضل حسینی[1*] ، محمدرضا خوانساری[1]

۱- دانشجوی کارشناس ارشد مهندسی فناوری و اطلاعات- شبکه های کامپیوتری دانشگاه علوم و فناوری آریان، بابل

۲- کارشناس ارشد مهندسی کامپیوتر – هوش مصنوعی دانشگاه علوم و فناوری آریان، بابل

* Mr.abolfazl96@yahoo.com

ارسال: بهمن ۹۶ پذیرش: اردیبهشت ۹۷

چکیده

با توسعه سریع فناوری اطلاعات در عصر حاضر، کاربردهای داده های بزرگ به بخش مهمی از سبک زندگی، کاری و بسیاری از مناطق مانند تجارت الکترونیک، بهداشت و درمان و صنعت، تبدیل شده است. از سوی دیگر، اختیار داشتن اطلاعات بیشتر و ارزشمند، اساس حالت تحقیق را در جامعه اطلاعاتی تغییر داده است.چیزی که توجه زیادی را در این برهه از زمان، از سوی دانشگاه و صنعت به خود جلب کرده، تجزیه و تحلیل موفقیت آمیز امنیت داده ها بر اساس فناوری داده های بزرگ است. پروسه تحقیق بروی داده های بزرگ جهت کشف دانش و الگوهای مخفی وهمچنین تحلیل اطلاعات با استفاده از تکنیک های هوش مصنوعی والبته داده کاوی برای تشخیص ناهنجاری ها در مشکلات و نگرانی های امنیتی کمک می کند.سپس یک مدل امنیتی هوشمند مبتنی بر تجزیه و تحلیل داده های بزرگ ارائه می گردد. در این مقاله، ما یک طرح جامع از تکنیک های مربوط به داده های بزرگ را در تجزیه و تحلیل امنیت شبکه ارائه می دهیم.کارهای تحقیقاتی موجود را دسته بندی می کنیم، سپس مسائل فنی، رویکرد و مقایسه آنها وهمچنین مزایا و معایب آنها را توضیح می دهیم. در نهایت، پتانسیل ها و مسیرهای تحقیق در آینده را بررسی می کنیم.

کلمات کلیدی: تجزیه و تحلیل امنیت شبکه، داده های بزرگ، تشخیص ناهنجاری، سیستم فایل توزیع هدوپ ، تهدید مستمر پیشرفته.

۱- مقدمه

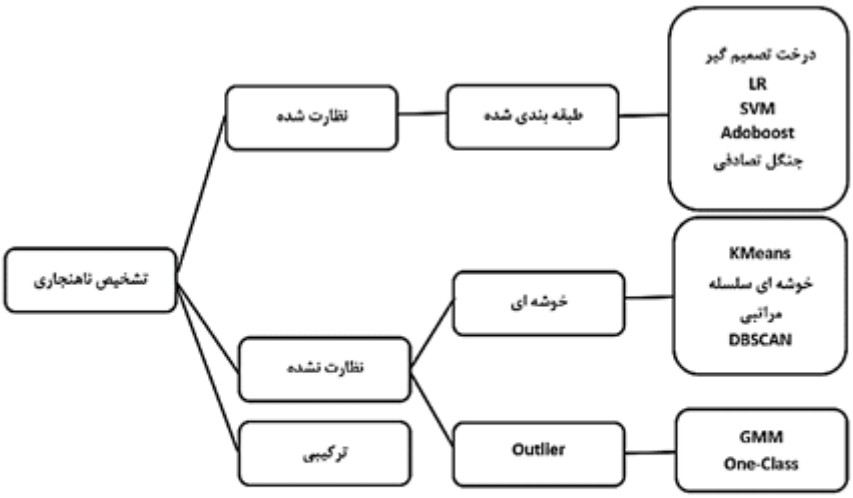
امروزه جامعه اطلاعاتی به عصر داده های بزرگ[۱،۲] وارد شده است. ظهور داده های بزرگ نه تنها سبک زندگی و کار را تغییر می دهد،بلکه حالت تحقیق را اساسا تغییر می دهد. ویژگی های داده های بزرگ به ترتیب در V۳ خلاصه می شوند: حجم، تنوع و سرعت.به طور کلی، داده های بزرگ را نمی توان با روش های سنتی در یک دوره معین حل کرد بنابراین روش های جدید باید برای مقابله با داده های بزرگ مانند داده کاوی و تکنیک های یادگیری ماشین ارائه شود.

با افزایش مقیاس و پیچیدگی شبکه، امنیت شبکه با چالش های بزرگ مواجه است. تشخیص سوء استفاده[۱] و تشخیص ناهنجاری[۲]، دو نوع اصلی روش در امنیت شبکه هستند.تشخیص سوء استفاده نیز به نام روش های مبتنی بر امضا[۳] نامیده می شود که با تطبیق الگوی ناهنجاری ها تشخیص داده می شود. با این حال، با پیچیدگی ترافیک شبکه، الگوهای حمله بسیار زیاد و متنوع هستند، ما نمی

[1] Email: M_khansari@outlook.com

توانیم تمام الگوهای حملات را توصیف کنیم.تشخیص آنومالی بر این فرض استوار است که فعالیت هایی که انحراف از حالت عادی دارند، به عنوان ناهنجاری ها در نظر گرفته می شود که نیازی به الگو ندارد. در این مقاله، تشخیص ناهنجاری را بیشتر توضیح خواهیم داد.

تهدیدات در یک شبکه، چالش هایی را برای قابلیت یکپارچگی، عملیاتی و هزینه های مربوط به کارکرد و نگهداری آن ایجاد می کند. امنیت شبکه [۳] شامل اقداماتی برای شناسایی، جلوگیری، پیشگیری و اصلاح موارد نقض امنیتی شامل انتقال اطلاعات می باشد. سه مفهومی که اهداف اساسی امنیت را شامل می شوند عبارتند از: محرمانه بودن، یکپارچگی و قابلیت دسترس(CIA) . اگر CIA حفظ و نگهداری شود، این شبکه امن است . برای اطمینان از اینکه شبکه معیارهای CIA را حفظ می کند، ترافیک شبکه باید به طور مداوم در برابر حملات مختلف با استفاده از تشخیص مبتنی بر امضا و نیز روش های تشخیص مبتنی بر آنومالی نظارت شود. برای تشخیص / جلوگیری از حملات، تجزیه و تحلیل ارتباطات در شبکه، داده ها بر روی شبکه و نوع درخواست های انجام شده است.اگر شبکه از اندازه های کوچک باشد، کار دشواری برای نظارت و تجزیه و تحلیل شبکه نیست، اما در مورد شبکه های بزرگ، انجام تجزیه و تحلیل معیارهای مربوط به اتصالات، درخواست ها و نوع داده ای که ارسال می شود به منظور محافظت از شبکه از حملات بسیار دشوار خواهد بود. در جهان، جایی که در حال حاضر میلیاردها دستگاه با هم مرتبط شده اند و دستگاه های جدیدی که هر روز به آن متصل می شوند، تجزیه و تحلیل ارتباطات شبکه، انتقال داده ها و حفاظت از آن با تکنیک های تشخیص و تجزیه و تحلیل متعارف موجود بسیار مشکل است. برای برخی از شبکه ها، اندازه داده های ورود به سیستم به اندازه سلول های TB افزایش می یابد که نیاز به روش محاسباتی قوی برای انجام تجزیه و تحلیل شبکه دارد. در این مقاله تکنیک تجزیه و تحلیل داده های ترافیک شبکه و تشخیص اتصالات غیرمستقیم به شبکه با استفاده از داده های بزرگ و سیستم فایل توزیع شده Hadoop [۵، ۴] ارائه شده است.



1: Misuse detection

2: Anomaly detection

3: Signature based

شکل ۱- انواع تشخیص ناهنجاری

۲- روش های تشخیص نا هنجاری (تهدیدات) در یک شبکه

در کارهای پژوهشی موجود، روش های تشخیص آنومالی [۶] را می توان به سه دسته، روش های تحت نظارت، روش های نظارت نشده و روش های ترکیبی طبقه بندی کرد (همانطور که در شکل ۱ نشان داده شده است).

۲-۱- روش های نظارت شده:

روش های نظارت شده عمدتا الگوریتم های طبقه بندی، مانند درخت تصمیم گیری، LR، SVM، Adaboost و جنگل های تصادفی است. روش های بر پایه نظارت شامل طبقه بندی دوتایی و طبقه بندی چند طبقه می باشد. طبقه بندی باینری به سادگی داده ها را به عنوان نرمال یا غیر نرمال ارزیابی می کند،در حالی که روش های چندگانه طبقه بندی می توانند فعالیت ها را به انواع خاصی تقسیم کند. از درخت تصمیم گیری یا جنگل تصادفی برای تشخیص شبکه بی نظیر استفاده می کنند. ویژگی ها و نمونه ها در هر درخت به طور تصادفی از کل مجموعه داده نمونه برداری می شود تا زمان اجرا و میزان هشدار خطا را با کاهش تعداد ویژگی های تصادفی کاهش یابد.

Adaboost یک الگوریتم تقویتی با ترکیبی از طبقه بندی ضعیف که در میان محققان بسیار محبوب است.محققان دریافتند که Adaboost در مقایسه با SOM، SVM و ANN کمبود محاسباتی کمتری دارد،در شرایطی که Adaboost دارای میزان تشخیص قابل قبول و میزان هشدار خطا است.محققان در [۷] یک مدل طبقه بندی براساس SVM و متد تصادفی ساختند.آنها نشان میدهند که SVM دقیقتر و زمانبرتر از جنگل تصادفی است.روش های [۸] بر اساس این مدل می تواند انواع خاصی برای ناهنجاری ها را پیش بینی کند.

محققان سعی در استفاده الگوریتم های مختلف برای تشخیص ناهنجاری در ترافیک دارند.از دیدگاه یک الگوریتم ، درخت تصمیم گیری و جنگل تصادفی برای داده های قطعی بیشتر مفید است. در حالی که LR و SVM اولویت داده های مستمر را دارند.برای طبقه بندی باینری، الگوریتم گروه بندی یا تقویت عملکرد بهتر از یک طبقه بندی واحد است.با این حال، طبقه بندی واحد بیشتر احتمال دارد که این مسایل در ابعاد چند کلاسه استفاده شود.علاوه بر این، قابل توجه است که تمام روش های نظارت بر داده های برچسب گذاری شده، که بیشترین کمبود این نوع روش ها را دارند، بسیار گران است.خوشبختانه روش غیرقابل کنترل می تواند با این مشکل حل شود. ما در مورد روش های نظارت نشده به شرح زیر ارائه خواهیم داد.

۲-۲- روش های نظارت نشده:

روش های غیرمتمرکز به روش خوشه ای و غیرفعال طبقه بندی می شوند. در مقاله [۹] ابتدا یک الگوریتم مبتنی بر خوشه بندی برای تشخیص آنومالی ارائه می دهد.در اینجا فرض می شود که فعالیت های عادی بسیار بیشتر از ناهنجاری ها است و دو نوع فعالیت متفاوت هستند.کلمه فعالیت هایی که در این کلاستر وجود دارد، می تواند به عنوان یکی از مهمترین ویژگی های فعالیت در دنیای عمومی باشد و نتیجه خوب زمانی حاصل می شود که نسبت فعالیت های عادی ۹۸.۵٪ باشد.علاوه بر این، عملکرد سه الگوریتم، خوشه ای، K نزدیکترین همسایه و SVM یک کلاس مورد مقایسه ماست.تحقیقات در [۱۰] الگوریتم های مختلف خوشه بندی را در تشخیص آنومالی مقایسه می کند. نتایج نشان می دهد که تشخیص خروجی مبتنی بر فاصله عملکرد بهترتری نسبت به سایر الگوریتم ها بر روی دقت و نرخ مثبت کاذب دارد.

برای افزایش دقت خوشه بندی، در [۱۱] یک رویکرد خوشه بند محدود را ارائه می دهد. این روش شامل برخی از دانش های پس زمینه به الگوریتم خوشه بندی است و اطلاعات پس زمینه از داده های ترافیک مشاهده شده است. این روشها عموما ناهنجاری ها را در یک مکان مشخص زمانی شناسایی می کنند و فعالیت های خوشه ای را برای نتایج چندین بار در زمان های مختلف می یابند. در نهایت، تمام فعالیت ها را با نتایج خوشه خوشه ای خود طبقه بندی می کند و در نهایت ناهنجاری ها را تشخیص می دهد. . مقاله در [۱۲] نشان می دهد که برنامه ای برای کاهش زمان تأخیر با به کارگیری نرمال سازی و در نهایت دستیابی به شناسایی در زمان واقعی بدون هیچ گونه امضا، دانش یا داده های برچسب گذاری شده است. تشخیص بیرونی فقط یک طرح کلی از داده های طبیعی را می گیرد و داده های ناقص ممکن است در یک خوشه قرار نگیرند. در [۱۳] چند طبقه بندی یک طبقه را برای بدست آوردن آستانه اعمال می کند، و در بعضی موارد روش متداول است. از این تحقیقات بالا، روش های نا منظم در تشخیص حملات جدید دارای

اولویت است.علاوه بر این، آنها به اطلاعات نشانه گذاری نیاز ندارند، که باعث کاهش کارهای دستی می شود. با این وجود، این روش ها عملکرد بدتری نسبت به روش های نظارت دارند.

۲-۳- روش های ترکیبی:

برخی از روش های متفاوتی وجود دارد که متشکل از روش های تحت نظارت و بدون نظارت است.روش ها در [۱۴، ۱۵] یکپارچه سازی روش آماری و طبقه بندی است که به طور تکراری اطلاعات را از داده های بدون برچسب استخراج می کند و آنها را با همبستگی طبقه بندی می کند. مقاله در [۱۶] درخت تصمیم گیری را برای تشخیص ناهنجاری ها اعمال می کند و همچنین SVM یک کلاس را برای ایجاد یک پروفایل برای داده های عادی فراهم می کند. این نه تنها می تواند به سرعت تشخیص خوب برسد، بلکه می تواند حملات جدید را نیز تشخیص دهد. روش ترکیبی در مقاله [۱۷] ترکیبی از خوشه بندی فازی c-means و الگوریتم شبکه عصبی مصنوعی است که باعث بهبود دقت تشخیص، به ویژه برای حملات نادر است. طرح در [۱۸] ترکیبی از تشخیص سوء استفاده و تشخیص آنومالی است. این آموزش یادآوری تصادفی را براساس الگوهای امضا شده استخراج کرده و سپس می تواند الگوها را به صورت خودکار تولید کند.تمام روشهای ترکیبی که در بالا ذکر شد بیشترین استفاده در جهت مزایا و عملکرد بالا را داراست و معایب را از بین می برند، به طوری که آنها همیشه می توانند بهتر از روش های تکی عمل کنند.

۳- چالشها و تهدیدات امنیتی شبکه

۳-۱- انواع حملات:

به اقداماتی که جهت محافظت شبکه در مقابل حملات داخلی و خارجی [۱۹] انجام می شود امنیت شبکه گفته می شود. این اقدامات دفاعی در لایه های متعددی از لبه شبکه تا داخل شبکه انجام می شود. هر لایه امنیتی شبکه،دارای سیاست های کنترلی جهت مدیریت دسترسی کاربران مجاز و مسدود کردن دسترسی هکرها به منابع شبکه می باشد.

- **Port Scanner:**

نرم‌افزاری است که درخواست‌های پیاپی از یک کلاینت به سرور را جهت شناسایی پورت‌های فعال ارسال میکند.البته هکرها با استفاده از این ابزار قادر به شناسایی سرویس های ارائه‌شده توسط این سرور با توجه به پورتهای باز میشوند و براساس این اطلاعات فرایند حمله خود را طراحی می‌کنند.

- **Man in the Middle:**

حمله مردمیانی (MITM) به گونه ای است که مهاجم با استفاده از روشهایی مانند Arp Poisoning، در بین دو طرف ارتباط قرار می‌گیرد و بدون اینکه طرفین ارتباط متوجه شوند شروع به شنود، دست‌کاری و جمع اوری اطلاعات می کند.

- **Arp Poisoning or Arp Spoofing:**

همانطور که میدانید وظیفه پروتکل Arp تبدیل Ip به Mac می‌باشد. هکرها با استفاده از این پروتکل و ایجاد بسته GArp جعلی و معرفی Ip Address گیت وی شبکه با Mac خود حمله را انجام می‌دهند و سیستم های شبکه براساس این بسته Arp Table خود را به‌روزرسانی می‌کنند و در نتیجه از این پس ترافیک مربوط به خارج شبکه را تحویل مهاجم می‌دهند و مهاجم بعد از شنود و جمع‌آوری اطلاعات (MITM)، ترافیک را به گیت وی اصلی ارسال میکند تا حمله توسط کاربران و مدیران کشف نشود.

- **Denial-of-Service attack:**

به مجموعه اقداماتی که جهت قطع موقت یا دائمی و یا تعلیق خدمات یک میزبان متصل به شبکه انجام می‌شود حملات منع سرویس یا DOS گفته می‌شود. بانک ها، کارت های اعتباری و حتی سرورهای ریشه را هدف اصلی این حملات هستند.در این نوع حمله

هکر با استفاده از روشهای زیادی مانند سرازیر کردن درخواست‌ها و استفاده بیش از حد از منابع (پردازنده، پایگاه داده، پهنای باند، حافظه و...) باعث کاهش سرعت سرور می‌شود و ممکن است حتی این کارسبب ازکارافتادن سرور شود.هدف از این حملات جلوگیری یا متوقف کردن کارکرد عادی یک وب سایت، سرور یا سایر منابع شبکه طراحی‌شده است.

- **Distributed Denial of Service- DDoS:**

حملات DDoS از مهلک ترین نوع از حملات Dos است. این نوع حمله بسیار شبیه حمله ping flood است اما با این تفاوت که از چندین کامپیوتر استفاده می‌شود. در این روش هکر یک دستگاه آلوده را به عنوان دستگاه اصلی (Master) به کار می برد و حمله را در سایر دستگاهها که zombie نامیده می‌شوند هماهنگ می‌نماید.حملات DDoS عموماً در از کار انداختن سایت های کمپانی های عظیم از حملات DoS مؤثرتر هستند.

۴- روش ها و تکنیک های جلوگیری از حملات شبکه

۴-۱- تکنیک سیستم فایل توزیعی Hadoop:

Hadoop از مفهوم محاسبات موازی استفاده می کند که در آن محاسبات یا برنامه هایی که تجزیه و تحلیل را انجام می دهند به جای انتقال داده ها ، داده به حافظه ای که در آن برنامه اجرا می شود، منتقل می شود. زمان انجام شده برای تکمیل محاسبات یا تجزیه و تحلیل مستقیما بستگی به تعداد ماشینهایی که محاسبات موازی دارند و جمع آوری نتایج حاصل از هر محاسبات می باشد. Hadoop باالهام از جدول بزرگ [۱] که سیستم ذخیره سازی داده های گوگل است ، سیستم فایل گوگل و MapReduce را به وجود می آورد. Hadoop یک چارچوب مبتنی بر جاوا و سکوی متن باز ناهمگون [۲] است. Hadoop جایگزینی برای پایگاه داده ، انبار (warehouse) یا استراتژی (Extract, Transform ,Load) نیست. Hadoop شامل یک سیستم فایل توزیع شده ، تجزیه و تحلیل و سکوی ذخیره سازی داده میباشد و یک لایه ای که محاسبات موازی، گردش کار و مدیریت پیکربندی را اداره میکند. Hadoop برای پردازش رویدادهای مختلط بلادرنگ مثل رشته ها طراحی نشده است.

HDFS [۳] یا سیستم فایل توزیع شده Hadoop ،درمیان گره ها در یک خوشه Hadoop ،اجرا می شود و سیستمهای فایل تعدادی داده ورودی و خروجی را به هم متصل می کند تا آنها را به صورت یک سیستم فایل بزرگ درست کند.تجزیه و تحلیل امنیت شبکه بر اساس داده های مربوط به اتصالات، نوع درخواست ها، داده ها منتقل شده بر روی شبکه با استفاده از ابزارهای ضبط شبکه مثل Wireshark انجام می شود. داده های ورودی فرمت شده و داده های مربوط به اندازه پنجره، پرچم های TCP و غیره از داده های ورودی برای شناسایی نفوذ، داده های مخرب منتقل شده، اتصالات نامناسب و غیره استخراج می شود. داده های استخراج شده به فرمت (.csv) ذخیره می شوند و در محیط HDFS آپلود می شود. سپس داده های استخراج شده به پایگاه داده که یک زبان ساختار یافته پرسوجو به نام زبان پرسه زنبور عسل (HQL) [۲۰، ۲۱] وارد می شود. HQL پیچیدگی برنامه نویسی MapReduce [۲۲، ۲۳، ۲۴] را در HDFS پوشش می دهد. یا به عبارت دیگر این نوع روش از فناوری Hadoop و داده بزرگ برای انجام نمایش داده ها در پایگاه داده بزرگ بسته های شبکه استفاده می کند.این به میزان قابل توجهی زمان لازم برای شناسایی انواع حملات و گزارش حملات را کاهش می دهد.

۴-۲- مکانیزم های پیشنهادی دفاع از حملات :

Kaushik، Pilli و Joshi [۲۵] روشی برای دادرسی های قانونی شبکه ارائه می دهند که بسته های شبکه را جمع آوری می کند و ویژگی ها را استخراج می کند. روش پیشنهادی بسته های مخرب را بر اساس ارتباط بین حملات مختلف شبکه و پارامترهای مربوط به شبکه تحت تأثیر قرار می دهد.یک پایگاه داده از بسته ها و پرس وجوها، برای شناسایی و گزارش حملات ایجاد شد. با این حال، محدودیت این روش زمانی است که برای تجزیه و تحلیل و گزارش حملات زمانی که مقدار زیادی یا بزرگ از بسته های

داده وجود دارد.این مکانیزم این امکان را فراهم می سازد تا به موقع حملات یا ارتباطات غیرمعمول شناسایی شده را در زمان واقعی پاسخ دهند.

Siddiqui و Naahid [۲۶] بر روی تجزیه و تحلیل داده های پایگاه داده Knowledge & Data Mining Cup 1999 با استفاده از الگوریتم خوشه بندی k-means با استفاده از تکنیک Data Mining اوراکل برای ایجاد رابطه بین پروتکل های بهره برداری شده توسط مهاجم و حملات بر روی شبکه کار کرده اند.الگوی انواع حمله در ارتباط با پروتکل های مختلف که توسط هکر مورد استفاده قرار می گیرد ارائه شده است.

Sun، Tao و Faloutsos [۲۷] یک معماری دوجانبه را پیشنهاد دادند تمرکز بر یادگیری بی نظیر است که قادر به تشخیص بی نظیر است.خوشه بندی برای دسته بندی ناهنجاری ها براساس شباهت بین آنها استفاده می شود. رویکرد پیشنهادی در مقاله متفاوت است به این دلیل که روش یادگیری خوشه ای یا بدون نظارت را در بر نمی گیرد.

۵- تکنیک ها و چشم اندازهای تجزیه و تحلیل چارچوب حفاظتی امنیت شبکه

۵-۱- تجزیه و تحلیل داده های بزرگ:

حفاظت تهدیدات پیشرفته نمی تواند از روش تجزیه و تحلیل داده بزرگ جدا شود، چه از داده های ورودی تولید شده توسط سیستم شبکه خود و چه از اطلاعات ورود به سیستم تولید شده توسط پلت فرم. هر دو می توانند از تکنیک تحلیل داده های بزرگ برای بازآزمایی داده استفاده کنند.آمار مربوط به داده ها، داده کاوی، تجزیه و تحلیل همبستگی، تجزیه و تحلیل روند و غیره می تواند برای کشف آثار حملات مستمر پیشرفته در داده های هیستوریک ثبت شده به منظور جبران کمبود حفاظتی امنیت سنتی مورد استفاده قرار گیرد.بدون شک، تکنیک تجزیه و تحلیل داده های بزرگ نیاز به یک پلت فرم جمع آوری داده ها و توانایی های تجزیه و تحلیل قدرتمند داده ها دارد و باید با طیف گسترده ای از سیستم نظارت واحد و سیستم پاسخ سریع اتوماتیک ترکیب شود تا غلبه ای بر مشکلات ناشی از بررسی و تحلیل ناشی از جزایر اطلاعاتی جدا شده باشد.

۵-۲- تکنیک ردیابی حمله:

تجزیه و تحلیل حمله چند لایه از سیستم امنیتی اطلاعات شبکه شامل تجزیه و تحلیل تماس تلفنی، تجزیه و تحلیل حمله از ساختار اصلی هسته، تجزیه و تحلیل فایل و تجزیه و تحلیل روند، تجزیه و تحلیل جریان شبکه، و غیره است. این روند نیازمند ایجاد یک مدل توصیف حمله یکنواخت چند منظوره دارد و تجزیه و تحلیل همبستگی چندگانه را طبق قوانین مربوطه بر اساس اطلاعات ورودی گسترده ای از مدل حمله چند لایه به منظور شناسایی حملات احتمالی در سیستم و ارائه پایه قابل اعتماد برای ردیابی مسیر نیاز دارد.روش ردیابی داده ها در زمینه پایگاه داده به طور گسترده مورد مطالعه قرار گرفته است.در حال حاضر دو چالش اصلی برای ردیابی داده وجود دارد: (۱) نحوه پردازش منابع ناهمگن داده، مانند پایگاه داده ستونی، پایگاه داده سند، سیستم کلید / ارزش سیستم و پایگاه داده XML . مهم است که از سیستم داده کاوی جهت پردازش منابع مختلف داده ها به طور موثر اطمینان حاصل شود. (۲) نحوه پردازش داده های عظیم. با گذشت زمان و استفاده از اینترنت و اینترنت اشیاء،داده ها به طور نمایی رشد می کنند، به منظور پردازش داده های بزرگ موثر،الگوریتم داده کاوی باید بسیار کارآمد و الاستیک باشد.

۶- نتایج حاصل از تکنیک داده بزرگ

۶-۱- چارچوب سیستم های امنیتی بر اساس تجزیه و تحلیل داده های بزرگ:

امروزه تهدیدات پیچیده[1]، زیرپوستی، پنهانی و بی‌رحم هستند. تهدیدات (سایبری) پیشین اگر موفق به نفوذ به هدف مورد نظر خود نمی‌شدند، آن را رها می کردند، اما یک تهدید پیشرفته مداوماً سعی در نفوذ به هدف مورد نظر خود دارد تا آن را بیابد. به

[1] Big table
[2] Heterogeneous Open Source Platform
[3] Hadoop Distributed File System

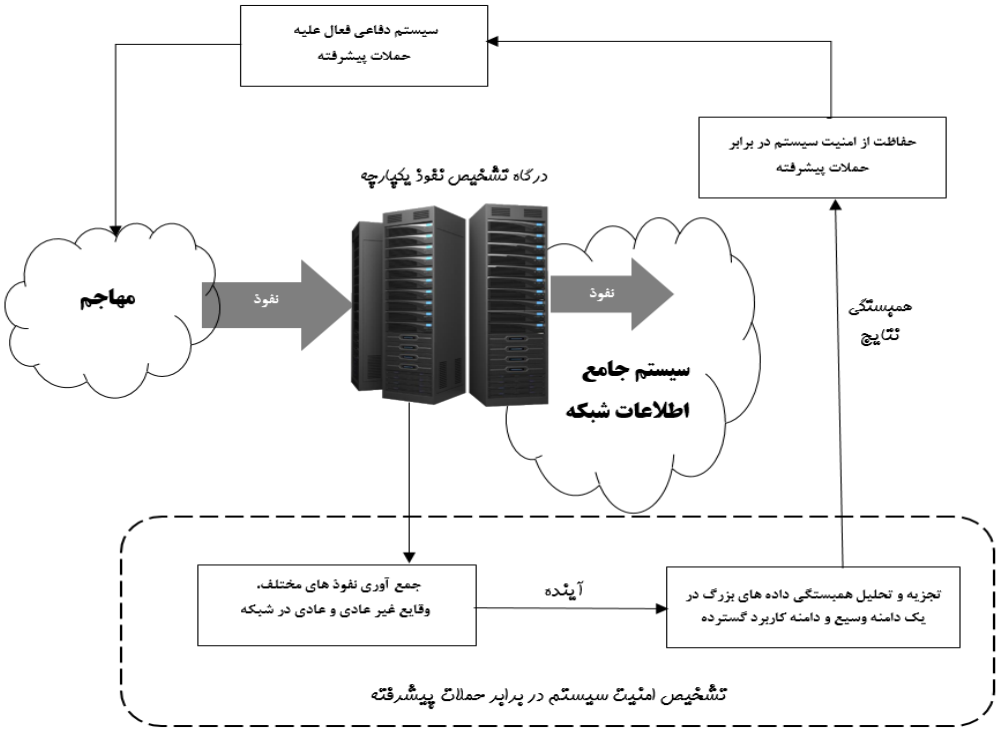
محض اینکه هدفش را یافت، اقدام به پنهان ساختن و یا در صورت نیاز مورف (تبدیل شدن به شئ دیگر) نمودن خود می کند و بدین ترتیب شناسایی و توقف خود را غیرممکن می‌سازد.

به همین دلیل سه چارچوب امنیتی اصلی برای مقابله با تهدیدات پیشرفته وجود دارد: (۱) برای مقابله با چالش های مختلف حملات پیشرفته، نیاز به ایجاد یک سیستم دفاع متقابل چند بعدی برای جلوگیری از این حملات در زوایای مختلف است.(۲) برای مقابله با پنهان سازی بیشتر حملات پیشرفته، باید رفتارهای غیر طبیعی را به عنوان پرونده های منتقل شده از طریق رمزنگاری با جریانهای غیرعادی تشخیص داد. (۳) برای مقابله با پنهان کردن حملات پیشرفتهِ درازمدت، نیاز به نظارت بر وضعیت تمام لایه های ماشین مجازی مشتری برای مدت زمان طولانی است.

۶-۲- طراحی سیستم های امنیتی بر اساس تجزیه و تحلیل داده های بزرگ:

تهدیدات پیشرفته با هدف استفاده از سیستم های اطلاعات شبکه، انواع مختلفی از ابزارهای فنی را در برمی گیرد، که دارای اهمیت و استقامت بیشتری است. با توجه به اقدامات احتمالی توسط مهاجمان، لازم است یک سیستم دفاع عمیق و سه بعدی در برابر این حملات ایجاد شود. این سیستم باید قادر به شناسایی هوشمندانه و تجزیه و تحلیل ارتباطی عمیق باشد و حفاظت امنیتی هدفمند و دفاع لحظه ای و کاری دهد. بر اساس نکات فوق، نویسنده [۲۸] پیشنهاد می کند که یک چارچوب امنیتی برای جلوگیری از حملات از طریق یک سیستم اطلاعاتی شبکه شامل تست امنیتی سیستم، حفاظت از امنیت سیستم و دفاع فعال طراحی گردد، به شکل ۲ برای جزئیات مراجعه شود. شکل نشان می دهد، در صورت دسترسی هر شبکه، درگاه تشخیص نفوذ یکپارچه می تواند برخی از حملات را شناسایی و مسدود کند.

در عین حال، سیستم امنیتی شبکه، انواع وقایع شبکه را از مهاجمان، دروازه ها و اینترانت ها جمع آوری می کند و ویژگی های رویداد را استخراج می کند، تجزیه و تحلیل ارتباطی عمیق داده ها را برای کشف تهدیدات پیشرفته که توسط دروازه تشخیص نفوذ یکپارچه شناسایی نشده است، انجام می دهد. سپس سیستم دفاعی، بر اساس نتایج تجزیه و تحلیل ارتباطی از فرایند تهدید پیشرفته در اینترانت ها ، از طریق حفاظت سیستم امنیتی جلوگیری خواهد کرد.



شکل ۲- چارچوب سیستم امنیتی اطلاعات شبکه براساس تجزیه و تحلیل داده های بزرگ

۶-۳- تست سیستم های امنیتی بر اساس تجزیه و تحلیل داده های بزرگ:

تشخیص امنیت سیستم در برابر تهدیدات پیشرفته شامل سه فرآیند به عنوان تشخیص نفوذ یکپارچه، ردیابی، جمع آوری رویداد شبکه و تجزیه و تحلیل ارتباطی عمیق داده های بزرگ است.

اول از همه، تشخیص نفوذ یکپارچه سیستم: از طریق معرفی مکانیزم فریبندگی عبور از سرور هسته (مانند تکنولوژی Honeynet یا Honeypot) از ابتکار عمل برای هرگونه حمله احتمالی ممکن بهره خواهد برد. Honeypot مجازی یک سیستم هوشمند جمع آوری اطلاعات مدافع است، استقرار هدف غیر تجاری از منابع امنیتی به منظور جلب حمله کنندگان به حمله، به منظور گرفتن و شواهد حملات و درک ابزار و روش های حمله است.

پس از کشف حملات، منبع شناسایی خواهد شد، که عمدتا از طریق ردیابی داده ها، استخراج داده ها و غیره انجام می شود. در نهایت، با در نظر گرفتن اطلاعاتی که سیستم دفاع شخصی جمع آوری می کند.نشان دهنده ویژگی های معمولی از داده های بزرگ مانند حجم داده های بزرگ، به روز رسانی سریع داده ها، نوع داده های چندگانه و چگالی ارزش داده کم بوده است. تجزیه و تحلیل داده های بزرگ برای انجام تجزیه و تحلیل همبستگی در اطلاعات رویداد شبکه استفاده می شود و دسترسی به اطلاعات اجرایی در یک دامنه وسیع و دامنه کاربرد گسترده به طوری که برای شناسایی رفتار حمله، حتی درگاه تشخیص نفوذ سیستم اطلاعات نظامی نمی تواند تشخیص دهد.

۶-۳-۱- حفاظت از امنیت سیستم در برابر حملات پیشرفته

نیازهای امنیتی منابع باید بر اساس سطح امنیتی اطلاعات محرمانه طبقه بندی شوند،دامنه مجازی معتبر باید با توجه به سطح محرمانگی کاربر و منابع سطح بالا، بتوان با خیال راحت از هم جدا کرد. در عین حال، داده های مهم باید رمزگذاری و ذخیره شوند و دسترسی رمزنگاری شده برای محدود کردن نفوذ حملات پیشرفته از خارج به داخل انجام شود. علاوه بر این، باید امنیت داده ها را در حالت پویا تضمین کند و مهاجرت ایمن از ماشین های مجازی را فراهم کند و در طول مهاجرت استراتژی های امنیتی سازگار باشد.

۶-۳-۲- سیستم دفاعی فعال علیه حملات پیشرفته

اول از همه باید حملات پیشرفته را به طور کلی در نظر گرفت و توانایی سرعت بخشیدن به مقدار زیادی اطلاعات در شبکه را تشخیص داده و سیگنال های ضعیفی را که می تواند در حملات پیچیده و مخفی شبکه ای دستگیر شود، شناسایی کند و یک سیستم ضد اطلاعاتی ایجاد کند که می تواند شناسایی حمله پیشرفته یا پیچیده را انجام دهد. دوم، سیستم دفاع باید یک لیست حمله را فهرست کند و اجسام حمله مناسب را برای مقابله انتخاب کند و بعد از رسیدن، آن شی را با سطح حفاظت بالاتر کنترل کند. فرایند دفاعی فعال بالا، ممکن است بر روی اشیاء اشتباه کار کند اما برای حفظ امنیت بالای سیستم اطلاعات شبکه، نیاز به تداخل و مقابله با آن وجود دارد. استراتژی دفاعی فعال می تواند به طور موثر برنامه حمله مهاجم را مختل کند و توان حمله کننده را برای دفاع بهتر از حملات پیشرفته محدود کند.

۷- مدل امنیتی مبتنی بر داده های بزرگ و تجزیه و تحلیل امنیت شبکه

با توجه به خرابی های محیط های دفاعی سنتی به همراه توانایی های حمله کننده ها به نجات از سیستمهای امنیتی سنتی ، لازم است سازمانها ، یک مدل امنیتی هوشمند اتخاذ کنند که دورتر از خطر، متنی و سریع باشد. امنیت هوشمند مبتنی بر تجزیه و تحلیل داده های حجیم است. داده های حجیم ، شامل هر دو مورد وسعت منابع و عمق اطلاعات می باشد که مورد نیاز برنامه ها جهت مشخص کردن خطرات، به طور دقیق و حمایت کردن در مقابل فعالیتهای غیر مجاز و تهدیدات سایبری پیشرفته است. یک مدل امنیتی مبتنی بر داده های حجیم ، دارای مشخصه های زیر میباشد:

- منابع داده داخلی و خارجی در مقدار ثابت ضرب می شود و یک تأثیر آموزشی همکاری کننده ایجاد می کند.
- ابزارهای خودکار انواع داده های مختلف را جمع آوری کرده و آنها را نرمال می کنند.

- مدیریت موتورهای تجزیه و تحلیل جهت پردازش حجم بزرگی از داده هایی که در زمان واقعی به سرعت در حال تغییر هستند.
- سیستمهای نظارت پیشرفته بطور مداوم، سیستم های با ارزش بالا و منابع را تجزیه و تحلیل می کنند و بر مبنای رفتار و مدل های خطا رسیدگی می کنند.
- دارای کنترلهای فعال از قبیل:نیاز تصدیق هویت کاربر سنتی،مسدود کردن انتقال داده ها یا ساده سازی تصمیم گیری تحلیلگران است.
- متمرکز سازی انبار بطوریکه همه داده های وابسته امنیتی برای تحلیلگران امنیتی جهت پرس و جو در دسترس باشند.
- استاندارد سازی دیدگاه ها به نمایشهای سازگار که به شکل قابل خواندن برای ماشین ایجاد شده اند و میتوانند در مقیاس منابع قابل اعتماد ، به اشتراک گذاشته شوند.
- زیرساختهای n-لایه که در تمام جهات مقیاس پذیری را ایجاد میکنند و قادر به پردازش پرس و جوها و جستجوهای بزرگ و پیچیده هستند.
- درجه بالایی از یکپارچگی را از طریق امنیت و ابزارهای مدیریت خطا جهت تسهیل بررسی مفصل مشکلات بالقوه ارایه می دهد.

از طریق نگهداری داده ها در یک مکان [۲۹] ، یک هدف برای حمله کننده ها جهت خرابکاری در سازمان ایجاد می شود. این نیاز دارد که انبارهای داده های حجیم به درستی کنترل شوند. جهت تضمین تصدیق ، یک چارچوب ارتباط امن رمز شده اجرا می شود. کنترل ها باید از اصل امتیاز کاهش استفاده کنند، مخصوصاً برای قوانین دسترسی به جز برای یک مدیر [۱] که اجازه دسترسی فیزیکی به داده ها را دارد. برای کنترلهای دسترسی مؤثر، آنها باید به طور ممتد مشاهده شوند و تغییر داده شوند. مانند تغییر نقشهای کارمندان سازمان. بنابراین کارمندان ، حقوق افراطی که می تواند مورد سوء استفاده قرار گیرد ، جمع نمی کنند. دیگر روشهای امنیتی ، نیاز به ضبط و تجزیه و تحلیل ترافیک شبکه دارند. سازمانها باید سرمایه گذاری در محصولات امنیتی را با استفاده از تکنولوژی های سریع مبتنی بر تجزیه و تحلیل تجهیزات غیر ایستا، تضمین کنند. مشکل دیگر مربوط به سازماندهی منطبق با قوانین حفاظت از داده ها می باشد. سازمانها باید برای ذخیره سازی داده ها ، انشعابات حقوقی را در نظر داشته باشند.به هر جهت ، داده های حجیم ، مزایای امنیتی دارند. زمانیکه سازمانها دانش را طبقه بندی می کنند، آنها داده را بطور خاص از طریق مقررات کنترل می کنند. نظیر تحمیل دوره های ذخیره سازی. این به سازمانها اجازه انتخاب داده ای می دهد که نه مقدار کوچک دارد و نه هیچ نیازی به نگهداری؛ بطوریکه به طور طولانی جهت سرقت در دسترس نیست. مزیت دیگر این است که داده های حجیم می تواند از تهدیدهایی نظیر شواهدی از نرم افزارهای مخرب ، ناهنجاری ها یا دزدی های اینترنتی در امان باشد.

۸- نتیجه‌گیری

در این بازبینی ، یک چارچوب کلی از محتوای داده های حجیم ، حوزه ، نمونه ها ، روشها ، مزایا ، چالشها و بحث های نگرانی های حریم خصوصی و امنیتی مربوط به شبکه مرور شده است .نتایج نشان دادند که حتی اگر داده ها و ابزارها و تکنیکها واقعاً در دسترس باشند ، نکات بسیاری جهت رسیدگی ، بحث ، بهبود ، توسعه ، تجزیه و تحلیل و ... وجود دارند.

در این تحقیق ، فرآیند تشخیص و تست نفوذ را می توان به صورت شکل ۲ بدست آورد. در ابتدا، ما باید قبل از پردازش، داده های اصلی را ایجاد کنیم. این فرآیند شامل استخراج ویژگی، انتخاب ویژگی و بیان ویژگی است. پس از آن، می توانیم داده های آموزشی را به الگوریتم یا روش های ترکیبی قرار داده و مدل تشخیص را بدست آوریم .

حملات مستمر پیشرفته چالش های بزرگی را برای امنیت اطلاعات شبکه ایجاد کرده است و تکنولوژی های دفاعی سنتی برای حفاظت از اطلاعات مهم و دارایی های اطلاعاتی موثر بسیار سخت است. خوشبختانه، برخی از فناوری های دفاعی برای مقابله با این حملات، توسعه یافته اند. با این حال، اکثر محصولات دفاعی فقط روی برخی از جنبه های زنجیره حملات پیشرفته جهت حفاظت

تمرکز می کنند. بنابراین، لازم است که یک معماری سیستم دفاعی مبتنی بر زنجیره حملات که تمام جنبه های حملات پیشرفته و پیچیده را پوشش می دهد و دستیابی به دفاع کامل در مقابل این حملات از طریق ترکیبی از مدیریت و ابزارهای فنی، طراحی گردد. گذشته از این ، موضوع فهم حفظ حریم خصوصی و امنیت داده های حجیم ، نیز پیامد بزرگی است که در آینده باید بیشتر مورد بحث قرار گیرد.

۹- مراجع

1. https://en.wikipedia.org/wiki/Big_data , last access : 01.01.2018

2. N. Miloslavskaya and A. Makhmudova, "Survey of Big Data Information Security," 2016 IEEE 4th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW), Vienna, 2016, pp. 133-138.

3. W. Stallings and W. Stallings. Cryptography and Network Security, 4/E. Pearson Education India, 2006.

4. J. Shafer, S. Rixner and A. L. Cox. "The hadoop distributed filesystem: Balancing portability and performance." Performance Analysis of Systems & Software (ISPASS), 2010 IEEE International Symposium on. IEEE, 2010.

5. K. Shvachko, et al. "The hadoop distributed file system." Mass Storage Systems and Technologies (MSST), 2010 IEEE 26th Symposium on. IEEE, 2010.

6. Caiyou zhang, Xiaojun Shen, Xubin Pei, Yiyang Yao . "Applying Big Data Analytics Into Network Security:Challenges, Techniques and Outlooks" Information & Telecommunication Branch,State Grid Zhejiang Electric Power Company, Hangzhou 310007, China.

7. M. A. M. Hasan, M. Nasser, B. Pal, S. Ahmad, Support vector machine and random forest modeling for intrusion detection system (IDS), Journal of Intelligent Learning Systems and Applications 2014 (2014) 45–52.

8. Y. Wang, A multinomial logistic regression modeling approach for anomaly intrusion detection, Computers & Security 24 (8) (2005) 662–674.

9. Portnoy L, Eskin E, Stolfo S. Intrusion detection with unlabeled data using clustering[C]//In Proceedings of ACM CSS Workshop on Data Mining Applied to Security (DMSA-2001. 2001.

10. Syarif I, Prugel-Bennett A, Wills G. Unsupervised clustering approach for network anomaly detection[C]//International Conference on Networked Digital Technologies. Springer Berlin Heidelberg, 2012: 135-145.

11. Wang Y, Xiang Y, Zhang J, et al. Internet traffic classification using constrained clustering[J]. IEEE Transactions on Parallel and Distributed Systems, 2014, 25(11): 2932-2943.

12. Owezarski P. A Near Real-Time Algorithm for Autonomous Identification and Characterization of Honeypot Attacks[C]//Proceedings of the 10th ACM Symposium on Information, Computer and Communications Security. ACM, 2015: 531-542.

13. G. Giacinto, R. Perdisci, M. Del Rio, F. Roli, Intrusion detection in computer networks by a modular ensemble of one-class classifiers, Information Fusion 9 (1) (2008) 69–82

14. Zhang J, Chen C, Xiang Y, Zhou W. Robust network traffic identification with unknown applications. In Proceedings of the 8th ACM SIGSAC symposium on Information, computer and communications security, Hangzhou, China, ACM, 2013; 405– 414.

15. Zhang J, Chen X, Xiang Y, Zhou W, Wu J. Robust network traffic classification. Networking 2015; 23(3):1257–1270.

16. G. Kim, S. Lee, S. Kim, A novel hybrid intrusion detection method integrating anomaly detection with misuse detection, Expert Systems with Applications 41 (4) (2014) 1690–1700.

17. Pandeeswari N, Kumar G. Anomaly detection system in cloud environment using fuzzy clustering based ANN[J]. Mobile Networks and Applications, 2015: 1-12.

18. Garcia-Teodoro P, Diaz-Verdejo J E, Tapiador J E, et al. Automatic generation of HTTP intrusion signatures by selective identification of anomalies[J]. Computers & Security, 2015, 55: 159-174.329

19. https://geekboy.pro/methods-of-creating-security-in-computer-networks/ , معرفی حملات و روش های ایجاد امنیت در شبکه های کامپیوتری

20. A. Thusoo, et al. "Hive: a warehousing solution over a map-reduce framework." Proceedings of the VLDB Endowment 2.2 (2009): 1626-1629.

21. A. Thusoo, et al. "Hive-a petabyte scale data warehouse using hadoop." Data Engineering (ICDE), 2010 IEEE 26th International Conference on. IEEE, 2010.

22. J. Dean and S. Ghemawat. "MapReduce: simplified data processing on large clusters." Communications of the ACM 51.1 (2008): 107-113.

23. R. Lämmel. "Google's MapReduce programming model— Revisited."Science of computer programming 70.1 (2008): 1-30.

24. J. Dean and S. Ghemawat. "MapReduce: a flexible data processing tool." Communications of the ACM 53.1 (2010): 72-77.

25. A. K. Kaushik, E. S. Pilli, and R. C. Joshi. "Network Forensic Analysis by Correlation of Attacks with Network Attributes." Information and ommunication Technologies. Springer Berlin Heidelberg, 2010.

26. M. K. Siddiqui and S. Naahid. "Analysis of KDD CUP 99 dataset using Clustering based Data Mining." International Journal of Database Theory and Application 6.5 (2013): 23-34.

27. J. Sun, D. Tao, and C. Faloutsos, "Beyond streams and graphs: dynamic tensor analysis," in Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, 2006, pp. 374-383.

28. W. Jia, "Study on Network Information Security Based on Big Data," 2017 9th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), Changsha, 2017, pp. 408-409.

29. https://academic.itpro.ir/articles/31554/big-data-چه-هستند-حجیم-های-داده-یا