پردیس بین الملل

پایان نامه کارشناسی ارشد

استفاده از تکنیک‌های داده کاوی در سیستم‌های تشخیص نفوذ

از

مهدی مکرمی سفیداب

استاد راهنما:

دکتر رضا ابراهیمی آتانی

اسفند ماه ۱۳۹۱


بسم الله الرحمن الرحيم

پردیس بین الملل
گروه مهندسی فناوری اطلاعات
گرایش تجارت الکترونیکی

استفاده از تکنیک‌های داده کاوی در سیستم‌های تشخیص نفوذ

از
مهدی مکرمی سفیداب

استاد راهنما:
دکتر رضا ابراهیمی آتانی

تقدیم به:

به پاس قدردانی از قلبی آکنده از عشق و معرفت که محیطی سرشار از سلامت، امنیـت، آرامـش و آسایش برایم فراهم آورده است. همدلی که با واژه ی نجیب و مغرور تلاش، آشـنایی دارد و تـلاش راستین را می‌شناسد. این پایان نامه تقدیم **همسر مهربانم** می گردد.

تشکر و قدردانی:

ذلک فضل من الله و کفی بالله علیما (آیه ۷۰ از سوره نساء)

چنین فضل ازسوی یکتا خداست که داناییش بس همه خلق راست

شکر بی پایان نثار ایزد منان که توفیق را رفیق راهم ساخت تا این پایان نامه را به سرانجام رسانم.
پس از حمد و سنای کردگار عالم، بر خود فرض می دانم از تمامی عزیزانی که در طی انجام این پژوهش از راهنمایی و مساعدتشان بهره مند گشته ام تشکر و قدردانی نمایم و برایشان از درگاه پروردگار مهربان آرزوی سعادت و پیروزی نمایم.
در آغاز صمیمانه ترین تقدیرها را تقدیم خانواده عزیز و مهربانم می نمایم که همواره حامی و مشوقم بوده اند و پیمودن روزهای سخت و آسان زندگی ام بدون دعای خیر، و برکت وجودشان غیرممکن بود.
از استاد فرزانه و ارجمندم جناب آقای دکتر رضا ابراهیمی آتانی که با سعه صدر و صبوری مرا راهنمایی نموده و با ارائه نظرات سازنده و رهنمودهای بی دریغشان در پیشبرد این پایان نامه سعی تمام مبذول داشتند کمال تشکر را دارم. همچنین از استاد گرانقدرم جنای آقای دکتر مهرگان مهدوی که در آغاز این تحقیق با رهنمودها و تشویق های خود مرا مورد لطف خویش قرار دادند سپاسگزارم.
از اساتید گرامی جناب آقای دکتر اسدالله شاه بهرامی و دکتر هاشم صابری که زحمت بازخوانی و داوری این مجموعه را به عهده داشتند صمیمانه تشکر و قدردانی می‌نمایم.
از کلیه اساتید گرانقدر گروه که در دوران تحصیل از محضرشان کسب فیض نمودم تشکر می‌نمایم.
و در نهایت از تمامی دوستان و هم کلاسیهای عزیزم که در طول این مدت افتخار آشنایی و مصاحبت با آنها را داشتم، به پاس محبت های بی دریغشان سپاسگزارم.

همتم بدرقه ی راه کن ای طایر قدس که دراز است ره مقصد و من نوسفرم

فهرست مطالب

فهرست اشکال

فهرست جداول

چکیده

استفاده از تکنیک‌های داده کاوی در سیستم‌های تشخیص نفوذ

مهدی مکرمی سفیداب

امروزه توسعه روزافزون شبکه های رایانه‌ای و کاربرد وسیع آن در زندگی بشر، لزوم تأمین امنیت این شبکه‌ها را بیش از پیش نمایان ساخته است. جهت تأمین امنیت از ابزار و تجهیزات مختلفی استفاده می‌شود که سیستم تشخیص نفوذ از جمله آنها به شمار می‌رود. سیستم‌های تشخیص نفوذ، اغلب از دو روش تشخیص سوء استفاده و تشخیص ناهنجاری به منظور تشخیص نفوذ استفاده می‌کنند. روش‌های تشخیص سوء استفاده، نرخ تشخیص بالایی دارند اما توانایی شناسایی حملات جدید را ندارند. در مقابل، روش‌های تشخیص ناهنجاری، توانایی شناسایی حملات جدید را دارند اما نرخ هشدار غلط بالایی دارند. تا کنون روش‌های مختلفی به منظور افزایش نرخ تشخیص و کاهش نرخ هشدار غلط در تشخیص ناهنجاری‌ها مورد استفاده و آزمایش قرار گرفته است؛ اما هیچ یک نتوانسته‌اند به موفقیت کامل دست یابند و تحقیقات در این زمینه همچنان ادامه دارد.

در این پایان نامه، مدلی جدید به منظور تشخیص ناهنجاری‌های شبکه، مبتنی بر تکنیک‌های داده کاوی، طراحی، پیاده سازی و ارزیابی شده است. مدل ارائه شده در این پایان نامه شامل سه فاز می‌باشد. فاز اول شامل آماده سازی و پیش پردازش داده است و در این فاز تلاش شده است تا ویژگی‌های موثر و مهم در شناسایی حملات استخراج گردد. فاز دوم نیز به ترتیب شامل سه مرحله تعیین تعداد بهینه خوشه‌ها، خوشه بندی توسط الگوریتم K-means و ارزیابی و اعتبارسنجی خوشه های ایجاد شده می‌باشد. فاز سوم، ساخت مدل توسط سه الگوریتم طبقه بندی C5، شبکه عصبی MLP و رگرسیون لجستیک و در نهایت مقایسه آنها و انتخاب مدل بهینه را شامل می‌شود. نتایج حاصل از آزمایشات انجام شده در این پایان نامه نشان می‌دهد که مدل درخت تصمیم گیری C5 با دقت طبقه بندی ۹۹,۸۴٪، نرخ تشخیص ۹۹,۸۳٪ و نرخ هشدار غلط ۰,۱۴٪ نسبت به سایر مدل‌های مورد مقایسه برتری دارد.

واژه‌های کلیدی: تشخیص نفوذ، سیستم‌های تشخیص نفوذ، تشخیص ناهنجاری، داده کاوی، تکنیک‌های داده کاوی.

Abstract

Using Data Mining Techniques in Intrusion Detection Systems

By: Mahdi Mokarrami Sefidab

Nowadays, with the increasing development and extended application of computer networks in human life, the need for security in these networks is more noticeable than ever before. Different equipments and tools are used to provide the security one of which is the Intrusion detection system. The intrusion detection systems more often apply misuse detection and anomaly detection methods to detect the Intrusion. In misuse detection methods, detection rates are high but they do not have the ability to detect new attacks. On the other hand, anomaly detection methods have the ability to detect new attacks, but with high false alarm rate. So far, several methods have been used and tested to increase the detection rate and reduce the false alarm rate in anomaly detection, but none have been able to achieve complete success and research in this field is still in progress.

In this dissertation, a new model for network anomaly detection based on data mining techniques has been designed, implemented and evaluated. The model presented in this thesis consists of three phases. The first phase includes the preparation and preprocessing of data. Furthermore, the most important and effective features of attack detection have been extracted. The second phase consists of three steps of determining the optimum number of clusters, clustering by K-means algorithm and evaluation and validation of clusters created. The third phase includes construction of the model by C5 classification algorithm, MLP neural network and logistic regression and then compares them and selects the optimal model. The results of the experiments conducted in this thesis show that the C5 decision tree model is superior to the other models, with the classification accuracy of 99.84%, detection rate of 99.83% and false alarm rate of 0.14%.

Keywords: Intrusion Detection, Intrusion Detection System, IDS, Data Mining, Data Mining Techniques, Anomaly Detection.

فصل ۱:

کلیات تحقیق

۱-۱- مقدمه

امروزه رایانه به یکی از وسایل معمولی و مرسوم در جامعه تبدیل شده است که از آن برای انجام کارهای روزمره مانند انتقال وجوه الکترونیکی و عملیات ماشین‌های خودپرداز بانکی، ذخیره حجم وسیعی از اطلاعات مانند اطلاعات پزشکی، اعتباری و مالی استفاده می‌شود. با این کاربرد وسیع، مسئولیت افرادی که این رایانه‌ها را کنترل می‌کنند مطرح است. هر گونه اشتباهی در اطلاعات می‌تواند به خسارات جبران ناپذیری منجر شود. اشتباهات غیر عمد در یک برنامه و یا در اقلام یک بانک اطلاعاتی، تنها مشکلی نیست که باید راجع به آن نگران بود، بلکه سوء استفاده عمد از این سیستم نیز باید برای ما اهمیت داشته باشد. کارمندان ناراضی ممکن است سعی کنند اطلاعات حساب‌ها را به نفع خود تغییر دهند. شرکت‌ها ممکن است سعی کنند به طرح‌های بازاریابی و رموز تجاری رقیبان خود به منظور رسیدن به فروش بیشتر دسترسی پیدا کنند.

فرآیند ایمن سازی منابع ارزشمند شبکه از حملات مخرب و دسترسی‌های غیر مجاز را امنیت شبکه می‌گویند [۱]. با توجه به مخاطرات موجود، امروزه امنیت شبکه به عنوان یک بخش مهم در فعالیت‌های مرتبط با شبکه مورد توجه قرار گرفته است. به منظور مقابله با حملات، شیوه‌ها و ابزار متعددی وجود دارد که سیستم تشخیص نفوذ[1]، از مهم‌ترین آنها به شمار می‌آید.

سیستم‌های تشخیص نفوذ وظیفه شناسایی و تشخیص هرگونه استفاده غیر مجاز از سیستم، سوء استفاده و یا آسیب رسانی توسط کاربران را بر عهده دارند [۲]. از نظر نوع عملکرد، این سیستم‌ها به دو دسته تشخیص سوء استفاده[2](تشخیص الگو) و تشخیص ناهنجاری[3] تقسیم می‌شوند [۲]. در روش تشخیص الگو، فعالیت‌های شبکه و سیستم به منظور شناسایی حملات و نفوذهای از پیش شناخته شده مورد بررسی قرار می‌گیرد. در روش تشخیص ناهنجاری، فعالیت‌های انجام شده

[1] Intrusion Detection System(IDS)
[2] Signature Detection
[3] Anomaly Detection

داخل شبکه و سیستم با حالت نرمال (رفتار متعارف و معمول) مورد مقایسه قرار می‌گیرد و هر رویداد یا جریان ترافیکی که با حالت نرمال تطابق نداشته باشد به عنوان ناهنجاری شناسایی می‌شود. هرکدام از این روش‌ها نقاط ضعف و قوتی دارند و تاکنون سیستم تشخیص نفوذی که تمامی ویژگی‌های یک سیستم تشخیص نفوذ مناسب همچون پایداری[1]، مقیاس پذیری[2]، گسترش[3]، تطابق[4]، کارایی[5] و قابلیت پیکربندی[6] را داشته باشد ارائه نشده است.

۱-۲- طرح مسئله

همان‌گونه که در مقدمه توضیح داده شد روش‌های تشخیص نفوذ به دو دسته تشخیص الگو و تشخیص ناهنجاری تقسیم می‌شوند. سیستم‌های تشخیص نفوذی که از روش تشخیص الگو استفاده می‌کنند از نرخ هشدار غلط[7] پایین و نرخ تشخیص[8] بالایی برخوردار هستند (هشدار غلط زمانی رخ می‌دهد که یک ترافیک علی رغم مخرب نبودن، مخرب تشخیص داده شود). اما عیب این گونه سیستم‌ها این است که قادر به تشخیص حملات جدید نیستند. برای رفع این مشکل از روش تشخیص ناهنجاری استفاده می‌شود. سیستم‌های تشخیص نفوذ مبتنی بر تشخیص ناهنجاری قادر به تشخیص حملات جدید هستند؛ اما نرخ هشدار غلط بالایی دارند [۱].

در مسیر افزایش دقت در تشخیص ناهنجاری‌ها و همچنین کاهش نرخ هشدار غلط، مشکلاتی وجود دارد که مهم‌ترین آنها به شرح زیر است:

- نفوذ کننده به منظور عدم شناسایی خود سعی دارد فعالیت‌های غیر مجاز خود را تا حد ممکن به فعالیت‌های متعارف و مجاز نزدیک نماید و این امر، فرآیند تشخیص ناهنجاری را با

---

[1] Robustness
[2] Scalability
[3] Extendibility
[4] Adaptability
[5] Performance
[6] Configurability
[7] False Alarm Rate
[8] Detection Rate

دشواری فراوانی مواجه می‌سازد.

- از آنجایی که اغلب مرز بین رفتار نرمال و غیر نرمال به طور دقیق مشخص نیست، بنابراین یک مشاهده ناهنجاری که در نزدیک مرز تطبیق قرار بگیرد ممکن است نرمال تلقی شود و برعکس.
- برای آموزش و ارزیابی سیستم‌های تشخیص نفوذ از داده های آموزش و آزمون استفاده می‌شود. برچسب اختصاص داده شده به نمونه آموزش تعیین می‌کند که این داده نرمال است یا غیر نرمال. نکته حائز اهمیت این است که بدست آوردن چنین داده برچسب‌گذاری شده‌ای که دقیق و مبین تمام رفتارهای سیستم، اعم از نرمال و غیر نرمال باشد دشوار و بسیار پرهزینه است. عموماً بدست آوردن یک مجموعه برچسب گذاری شده از نمونه های غیر نرمال که همه انواع رفتارهای ناهنجار ممکن در سیستم را پوشش دهد دشوارتر از داشتن برچسب نمونه های مربوط به رفتار نرمال است [۳].
- رفتار ناهنجار، طبیعتی پویاتر نسبت به رفتار نرمال سیستم دارد. به همین علت، ممکن است ناهنجاری‌های جدیدی در سیستم مشاهده شوند که مجموعه آموزش برچسب خورده ای برای آنها وجود نداشته باشد. این مسائل آموزش مدل‌های تشخیص ناهنجاری را با مشکل مواجه می‌کند.

در راستای رفع مشکلات موجود نیز تکنیک‌های متعددی (همچون الگوریتم ژنتیک، شبکه های عصبی، مدل های آماری) مورد استفاده قرار گرفته است که هر یک مزایای خاص خود را دارند؛ اما هیچ کدام نتوانسته‌اند به ساخت یک سیستم تشخیص نفوذ کامل با تمامی ویژگی‌های آن از جمله دقت بالا، انعطاف پذیری و قدرت تشخیص صد در صدی، منجر شوند و تحقیقات در این خصوص همچنان ادامه دارد. داده کاوی نیز با توجه به قدرت و سرعت آن در تحلیل داده های حجیم و کشف دانش از بین این داده‌ها، امروزه به پرکاربردترین و قدرتمندترین روش در این خصوص تبدیل شده است و اکثر تحقیقات در زمینه تشخیص ناهنجاری هم به سمت استفاده از این علم، سوق داده شده است.

۱-۳- روش تحقیق

روش پژوهش در این پایان نامه بر مطالعه نظری و کتابخانه ای و همچنین استفاده از تکنولوژی داده کاوی در امر تحقیق و آزمایش بر روی داده های موجود در حوزه سیستم‌های تشخیص نفوذ به منظور بررسی و اثبات صحت و درستی طرح پیشنهادی استوار می‌باشد. مراحل انجام این پایان نامه در قالب محورهای پژوهشی زیر قابل بحث و ارائه است:

۱. بررسی مسئله امنیت در شبکه های رایانه‌ای و همچنین ابزار تشخیص و جلوگیری از حملات و نفوذها به آنها.
۲. بررسی ابعاد مختلف سیستم‌های تشخیص نفوذ، همچنین نقاط قوت و ضعف آنها.
۳. بررسی روش‌ها و شیوه های تشخیص نفوذ بکار رفته در سیستم‌های تشخیص نفوذ.
۴. بررسی داده کاوی و نقش آن در سیستم‌های تشخیص نفوذ و پیشنهاد یک مدل برای تشخیص ناهنجاری‌های شبکه با استفاده از تکنیک‌های داده کاوی.
۵. تحلیل و ارزیابی مدل پیشنهادی با استفاده از مجموعه داده های مرجع و همچنین نرم‌افزارهای داده کاوی.
۶. نتیجه گیری، شامل ارائه نتایج و تجربیات حاصل و همچنین طرح محورهای پژوهشی آینده در این راستا.

۱-۴- اهداف

در این پایان نامه در خصوص تشخیص ناهنجاری‌های شبکه صحبت می‌شود. هدف اصلی، ارائه مدلی برای تشخیص ناهنجاری شبکه با استفاده از تکنیک‌های داده کاوی است. از آنجایی که از تکنیک‌های داده کاوی در بخش‌های مختلفی از تشخیص نفوذ استفاده شده است (از مرحله جمع‌آوری و آماده سازی داده‌ها تا مرحله تحلیل آنها) و هر کدام از این بخش‌ها به نوبه خود اهمیت ویژه ای در موضوع تشخیص نفوذ دارند و همچنین در اکثر تحقیقات، از تکنیک‌های خاصی از داده کاوی استفاده شده و تنها بر بخش خاصی از تشخیص نفوذ متمرکز شده‌اند، لذا در این پایان نامه با استفاده از چند تکنیک داده کاوی و پیاده سازی آنها، مدلی جهت تشخیص

ناهنجاری‌های شبکه ارائه شده است که دقت تشخیص را به میزان قابل توجهی افزایش داده و نرخ هشدار غلط را کاهش می‌دهد. مهم‌ترین اهداف این مدل عبارتست از:

- کاهش ابعاد داده‌ها و انتخاب ویژگی‌های مهم و موثر در شناسایی نفوذها.
- افزایش دقت در تشخیص نفوذ، به واسطه حذف داده‌های اضافی و نامربوط و انتخاب ویژگی‌های موثر در هر حمله.
- افزایش سرعت و به حداقل رساندن زمان شناسایی حملات و نفوذها.
- کاهش نرخ هشدار غلط در روش تشخیص ناهنجاری‌های شبکه.
- تأمین امنیت اطلاعات و سیستم‌ها در برابر نفوذها.

۱-۵- ساختار پایان نامه

این پایان نامه به شکل زیر سازماندهی شده است:

فصل ۲ در خصوص مفاهیم، تاریخچه و نقاط ضعف و قوت سیستم‌های تشخیص نفوذ بحث کرده است.

فصل ۳ فرآیند آنالیز داده‌ها را تشریح نموده و تکنیک‌های تشخیص ناهنجاری را بر می‌شمرد و در ادامه به معرفی داده کاوی و نقش آن در تشخیص نفوذ و ناهنجاری می‌پردازد و در پایان نیز مروری بر تحقیقات انجام شده در این حوزه دارد.

فصل ۴ در خصوص مدل پیشنهاد شده صحبت می‌کند.

فصل ۵ به تحلیل و ارزیابی مدل پیشنهادی می‌پردازد.

فصل ۶ به نتیجه گیری و ارائه پیشنهاداتی برای کارهای آینده تعلق دارد.

فصل ۲:

مروری بر سیستم‌های تشخیص نفوذ