بسم اللهالرحمن الرحيم

دانشکده ریاضی و رایانه
بخش علوم رایانه

پایان نامه تحصیلی برای دریافت درجه کارشناسی ارشد رشته علوم رایانه گرایش سیستم‌های هوشمند

ارائه روشی جدید به منظور تشخیص نفوذ به شبکه با داده‌کاوی

مؤلف:
زهرا اصغری ورزنه

استاد راهنما:
دکتر مرجان کوچکی رفسنجانی

استاد مشاور:
دکتر محمد مسعود جاویدی

شهریور ۹۲

این پایان نامه به‌عنوان یکی از شرایط احراز کارشناسی ارشد به

بخش کامپیوتر- دانشکده ریاضی و رایانه
دانشگاه شهید باهنر کرمان

تسلیم شده است و هیچگونه مدرکی به‌عنوان فراغت از تحصیل دوره مزبور شناخته نمی‌شود.

دانشجو: زهرا اصغری ورزنه

استاد راهنما: دکتر مرجان کوچکی رفسنجانی

استاد مشاور: دکتر محمد مسعود جاویدی

داور۱:

داور۲:

نماینده تحصیلات تکمیلی دانشکده:

معاون آموزشی و پژوهشی دانشکده:

تقدیم به پدر و مادرم

آنانکه وجودم برایشان همه رنج است

و وجودشان برایم همه مهر

و تقدیم به خواهر و برادران عزیزم که همواره موجب شادی و دلگرمی من بوده اند.

باسپاس از سه وجود مقدس:

آنانکه ناتوان شدند تا ما به توانایی برسیم . . .

موهایشان سپید شد تا ما روسفید شویم . . .

و عاشقانه سوختند تا گرمابخش وجود ما و روشنگر راهمان باشند . . .

پدرانمان

مادرانمان

و استادانمان

بطور خاص از اساتید گرانقدر سرکار خانم دکتر کوچکی و جناب آقای دکتر جاویدی که با تشویق ها و راهنمایی های خود راه را برایم هموار نمودند تشکر می نمایم .

چکیده

امروزه با گسترش روزافزون شبکه‌های کامپیوتری، امنیت شبکه از اولویت زیادی برخوردار می-شود. سیستم‌های تشخیص نفوذ، سخت‌افزار و یا نرم‌افزاری است که کار نظارت بر شبکه کامپیوتری را در مورد فعالیت‌های مخرب و یا نقض سیاست‌های مدیریتی و امنیتی انجام می‌دهد و گزارش‌های حاصل را به بخش مدیریت شبکه ارائه می‌دهد. رویکردهای تشخیص نفوذ به دو دسته کلی تشخیص موارد سوءاستفاده و تشخیص موارد غیرمتعارف تقسیم می‌شوند. سیستم‌های تشخیص موارد سوء-استفاده تلاش می‌کنند حمله‌ها را با استفاده از کشف الگوهای نفوذ که توسط خبرگان تشخیص داده شده و گزارش می‌شوند شناسایی کنند. رویکرد تشخیص موارد غیر متعارف در واقع توسعه رویکرد قبلی است با این توضیح که در این رویکرد، الگوهایی از رفتارهای نرمال در شبکه از قبل، تشخیص داده شده است و نفوذ می‌تواند مبتنی بر مقداری انحراف از رفتارهای نرمال در شبکه تعریف شود. در سیستم‌های تشخیص نفوذ از روش‌های مختلفی استفاده می‌شود که یکی از این روش‌ها داده‌کاوی است. داده‌کاوی از تکنیک‌های مختلفی مانند تکنیک‌های آماری و یادگیری ماشین بهره می‌برد. در این پایان نامه تکنیک‌های یادگیری ماشین بررسی شده و برای پیاده‌سازی سیستم تشخیص نفوذ از روش دسته‌بندی مبتنی بر قانون فازی استفاده شده است و این طرح برای یافتن وزن‌های بهینه قوانین از الگوریتم ژنتیک بهره می‌برد. روش پیشنهادی بر روی پایگاه داده KDD99 و با استفاده از نرم‌افزار متلب پیاده‌سازی شده است که در مقایسه با سایر روش‌ها (الگوریتم‌های ماشین بردار پشتیبان(SVM)[1]، K- نزدیکترین همسایه (K-NN)[2]، Naïve Bayes، مدل پنهان مارکوف (HMM)[3]، C4.5، K-mean، Y-mean و برخی از روش‌های ترکیبی) نتایج بهتری حاصل شده است.

کلید واژه: تشخیص نفوذ[4]، داده‌کاوی[5]، سیستم دسته‌بندی مبتنی بر قانون فازی[6]، الگوریتم ژنتیک[7]

[1] Support Vector Machine (SVM)
[2] K-Nearest Neighbor (KNN)
[3] Hidden Marcov models (HMM)
[4] Intrusion Detection (ID)
[5] Data mining
[6] Fuzzy Rule Base Classification System (FRBCS)
[7] Genetic Algorithm (GA)

فهرست مطالب

عنوان **صفحه**

فصل سوم: داده‌کاوی و تکنیک‌های آن

فصل چهارم: پیشینه داده‌کاوی در تشخیص نفوذ

فصل پنجم: ارائه روش پیشنهادی

فصل ششم: پیاده‌سازی و ارزیابی کارایی

فصل هفتم: نتیجه‌گیری و پیشنهادات

فهرست شکل‌ها

عنوان **صفحه**

فهرست جداول

عنوان **صفحه**

فهرست نمودارها

عنوان **صفحه**

فصل اول:

کلیات

۱-۱- مقدمه

با رشد شبکه‌های کامپیوتری، امنیت شبکه‌های کامپیوتری از اهمیت بالایی برخوردار می‌شود. نفوذ، مجموعه اقدامات غیرقانونی است که صحت، محرمانگی و دسترسی به منبع را به خطر می‌اندازد[۱]. نفوذگران را می‌توان به دو دسته نفوذگران داخلی و خارجی دسته‌بندی کرد. نفوذگران داخلی کسانی هستند که برای دستیابی به سیستم اختیارات محدودی دارند، اما سعی دارند به منابعی که اجازه دسترسی به آن را ندارند دست پیدا کنند و نفوذگران خارجی کسانی هستند که اجازه استفاده از سیستم را ندارند، ولی سعی دارند سیستم را مورد دسترسی قرار دهند[۲].

از آنجایی که از نظر تکنیکی ایجاد سیستم‌های کامپیوتری بدون نقاط ضعف و شکست امنیتی عملاً غیرممکن است؛ تشخیص نفوذ در تحقیقات سیستم‌های کامپیوتری با اهمیت خاصی دنبال می‌شود.سیستم‌های تشخیص نفوذ[1]، می‌توانند انواع نفوذ روی شبکه‌ها را که می‌تواند شامل جمع‌آوری اطلاعات، پویش پورت‌ها، بدست آوردن کنترل کامپیوترها و هک کردن آن‌ها باشد را شناسایی کنند.

۱-۲- بیان مسئله و پیشینه تحقیق

سیستم تشخیص نفوذ یک سیستم محافظتی است که خرابکاری‌های در حال وقوع روی شبکه را شناسایی می‌کند. عموماً سیستم‌های تشخیص نفوذ در کنار دیوارهای آتش[2] و به‌صورت مکمل امنیتی برای آن‌ها مورد استفاده قرار می‌گیرند[۳]. این سیستم‌ها سعی دارند نفوذهای غیر مجاز به شبکه را با توجه به الگوریتم‌های خاص تشخیص دهند، که می‌توان آن‌ها را به دو دسته کلی تشخیص سوء‌استفاده[3] و تشخیص رفتار غیرمتعارف[4] تقسیم کرد. سیستم‌های تشخیص موارد سوءاستفاده تلاش می‌کنند حمله‌ها را با استفاده از کشف الگوهای نفوذ که توسط خبرگان تشخیص داده شده و گزارش می‌شوند شناسایی کنند. رویکرد تشخیص موارد غیر متعارف در واقع توسعه رویکرد قبلی است با این توضیح که در این رویکرد، الگوهایی از رفتارهای نرمال در شبکه از قبل، تشخیص داده شده است و نفوذ می‌تواند مبتنی بر مقداری انحراف از رفتارهای نرمال در شبکه تعریف شود.

بدلیل اهمیت امنیت در شبکه‌های کامپیوتری از روش‌های زیادی برای تشخیص نفوذ استفاده شده

[1] Intrusion Detection System (IDS)
[2] Firewall
[3] Misuse Detection
[4] Anomaly Detection

است، از جمله این روش‌ها استفاده از الگوریتم‌های داده‌کاوی می‌باشد. در سیستم‌های تشخیص نفوذ نرم افزارهای مبتنی بر الگو، حسگرهای ترافیک شبکه را بررسی می‌کنند و الگوهایی را هم ذخیره می کنند و سپس هشدارها[1] را فعال می‌سازند و در صورتی‌که رخدادی در سیستم اتفاق بیفتد، مأمور امنیتی را آگاه می‌سازند. وقتی یک شبکه بسیار بزرگ و پیچیده باشد، تعداد هشدارها زیاد می‌شود و بنابراین حملات هم بیشتر می‌شوند و حسگرها نیز ممکن است الگوها را درست تشخیص ندهند. این می‌تواند دلیلی برای بکارگیری داده‌کاوی باشد[۴و۵]. داده‌کاوی فرآیند[2] کشف مدل‌های مختلف، خلاصه‌ها و مقادیر کسب شده از مجموعه داده می‌باشد. داده‌کاوی یکی از تکنولوژی‌های کاربردی برای تشخیص نفوذ است که به ابداع یک الگوی جدید از داده‌های شبکه‌های عظیم می‌پردازد. در حال حاضر بسیاری از محققان بر روی سیستم‌های تشخیص نفوذ براساس تکنیک‌های داده‌کاوی متمرکز شده‌اند. داده‌کاوی از تکنیک‌های مختلفی مانند تکنیک‌های آماری و یادگیری ماشین استفاده می‌کند.

وجود مسائل پیچیده علمی منجر می‌شود تا به سراغ روش‌های بهینه‌سازی رفته و مسئله مورد نظر را به وسیله آن‌ها حل کرد. با توجه به زمانبر بودن و پیچیدگی روش‌های دقیق از رویکرد بهینه‌سازی هوشمند استفاده می‌شود. بهینه‌سازی، تغییر دادن ورودی‌ها و خصوصیات یک دستگاه، فرآیند ریاضی و یا آزمایش تجربی است به‌طوری که بهترین نتیجه حاصل شود. در تحقیقاتی که در سال‌های اخیر در زمینه تشخیص نفوذ انجام گرفته، سعی شده تا با سرعت و دقت بیشتر نفوذهای انجام شده به شبکه شناسایی شود. از اینرو محققان از مجموعه داده استاندارد KDD [۶]، برای انجام آزمایشات استفاده کرده‌اند. استفاده از الگوریتم‌های ترکیبی، تکنیک‌های فازی، شبکه های عصبی، الگوریتم‌های ژنتیک و غیره از جمله کارهایی است که به منظور تشخیص نفوذ به شبکه انجام شده است.

Wang و همکارانش[۷] از شبکه‌های عصبی مصنوعی[3] و خوشه‌بندی فازی برای تشخیص نفوذ استفاده کردند. در این روش ابتدا از تکنیک‌های خوشه‌بندی فازی برای تولید زیرمجموعه‌های آموزشی مختلف استفاده می‌شود. براساس زیر مجموعه‌های آموزشی مختلف مدل‌های شبکه عصبی برای فرموله کردن مدل‌های مختلف پایه آموزش می‌بینند. در نهایت رویه تجمیع فازی، برای تجمیع این نتایج بکار برده می‌شود.

[1] Alarm
[2] Process
[3] Artificial Neural Networks (ANN)

Muniyandi و همکاران [۸] از ترکیب الگوریتم خوشه‌بندی K-mean و الگوریتم دسته‌بندی C4.5 برای تشخیص رفتار غیرعادی استفاده کردند. در این روش ابتدا از K-mean برای بخش‌بندی نمونه-های آزمایش استفاده شده است و سپس درخت تصمیم در هر خوشه مرزهای تصمیم‌گیری را تصحیح می‌کند.

Fores و همکاران [۹] ، از رویکرد داده‌کاوی برای استخراج الگوهایی که رفتار نرمال را برای تشخیص نفوذ نشان می‌دهد، استفاده کردند. آن‌ها مجموعه‌ای از قوانین پیوندی فازی را بکار بردند که از داده بازبینی شبکه به‌عنوان مدل‌های رفتار نرمال استخراج می‌شوند. همچنین برای کشف رفتارهای غیر نرمال قوانین پیوندی را با هم ترکیب کردند و شباهت آن را با مجموعه‌های استخراج شده از داده نرمال محاسبه کردند.

Jian-hua و همکارش [۱۰] از الگوریتم ژنتیک برای بهبود وزن‌ها و حدآستانه شبکه‌های عصبی استفاده کردند. آن‌ها از شبکه‌های عصبی چندلایه BP برای دسته‌بندی داده‌های نرمال و حمله در زمینه تشخیص نفوذ استفاده کردند. وزن‌های اولیه و حدآستانه در این شبکه‌ها بصورت تصادفی انتخاب می-شوند. آن‌ها برای انتخاب بهترین آستانه و وزن برای بهبود نتایج از الگوریتم ژنتیک استفاده کردند. این روش مدل شبکه عصبی BP را سرعت می‌بخشد و نرخ کشف نفوذ را بالا می‌برد.

۱-۳- چالش‌ها و اهداف تحقیق

با توجه به اینکه در سیستم‌های تشخیص نفوذ دو معیار دقت در تشخیص نفوذ به شبکه‌های کامپیوتری و همچنین کاهش تعداد هشدارهای نادرست از اهمیت زیادی برخوردار است، در این تحقیق سعی شده است تا با ارائه روش جدید مبتنی بر دسته بندی و الگوریتم‌های تکاملی نرخ تشخیص نفوذ را افزایش دهیم و کمترین میزان هشدارهای نادرست را داشته باشیم.

هدف از این تحقیق آشنایی با روش‌های تشخیص نفوذ، بررسی سیستم‌های تشخیص نفوذ موجود، بررسی روش‌های مختلف داده‌کاوی در تشخیص نفوذ و در نهایت هدف کلی ارائه روشی به منظور تشخیص نفوذ به شبکه با استفاده از تکنیک‌های داده‌کاوی جهت افزایش دقت و صحت تشخیص می-باشد. در طرح پیشنهادی برای پیاده‌سازی سیستم تشخیص نفوذ از روش دسته‌بندی مبتنی بر قانون فازی استفاده شده است. علاوه بر آن، این طرح برای یافتن وزن‌های بهینه قوانین از الگوریتم ژنتیک بهره می‌برد.

۱-۴- مروری بر فصول پایان نامه

این پایان نامه مشتمل بر شش فصل می‌باشد. در فصل دوم سیستم‌های تشخیص نفوذ توضیح داده شده و روش‌های استفاده شده برای تشخیص نفوذ ذکر شده است. همچنین برخی از سیستم‌های تشخیص نفوذ موجود معرفی شده است.

در فصل سوم مفاهیم داده‌کاوی و تکنیک‌ها و الگوریتم‌های مختلف آن بررسی شده است. در این فصل سعی شده است تمامی موارد مربوط به داده‌کاوی به‌طور خلاصه پوشش داده شود.

در فصل چهارم ارتباط داده‌کاوی و سیستم‌های تشخیص نفوذ بررسی شده است و سعی شده است گام‌های انجام یک فرآیند داده‌کاوی به منظور تشخیص نفوذ توضیح داده شود. در ادامه این فصل مجموعه داده استاندارد استفاده شده به همراه ویژگی‌های آن معرفی شده است.همچنین تکنیک‌های داده‌کاوی مورد استفاده در تشخیص نفوذ بررسی شده و کارهای انجام گرفته در این زمینه معرفی شده است.

در فصل پنجم ابتدا مدل دسته‌بندی مبتنی بر قانون فازی توضیح داده شده و در ادامه روش پیشنهادی به‌منظور بهبود کارایی قوانین ارائه شده است.

در فصل ششم گام‌های پیاده‌سازی روش پیشنهادی تشریح شده و در ادامه نتایج بدست آمده با کارهای قبلی مقایسه شده است.

در فصل هفتم نتیجه‌گیری و پیشنهادات برای کارهای آینده ارائه شده است.

فصل دوم:

مفاهیم کلی درباره سیستم‌های تشخیص نفوذ