تشخیص نفوذ در شبکه‌های رایانه‌ای با استفاده از مدل مخفی مارکوف تکاملی

محمد درویشی*[1] و مجید غیوری ثالث[2]

[1]دانشجوی کارشناسی ارشد، دانشگاه جامع امام حسین(ع)، تهران، ایران
Mhdarvishi@ihu.ac.ir

[2]دانشیار، دانشگاه جامع امام حسین(ع)، تهران، ایران
Ghayoori@ihu.acir

چکیده

سامانه‌های تشخیص نفوذ، وظیفه شناسایی و تشخیص هرگونه ورود غیرمجاز به سیستم، سوء استفاده و یا آسیب‌رسانی را بر عهده دارند، که با استفاده از تحلیل بسته‌های شبکه، قادر به پیش‌گیری از حملات سایبری است. در حال حاضر یکی از چالش‌های عمده در استفاده از این ابزار کمبود الگوهای آموزشی حملات در بخش موتور تحلیل است، که باعث عدم آموزش کامل موتور تحلیل و درنتیجه تولید حجم بالایی از هشدارهای غلط خواهد شد. از طرفی بالابودن زمان آموزش سامانه‌های تشخیص نفوذ، موجب تأخیر قابل توجهی در بخش آموزش سامانه به همراه خواهد داشت. پژوهش پیش رو نیز تلاشی است برای ارائه یک راه‌کار تشخیص نفوذ مبتنی بر امضا با محوریت مدل مخفی مارکوف تکاملی با نام EHMM که در راستای غلبه بر چالش‌های مطرح‌شده ارائه شده است. مهم‌ترین بخش مدل مخفی مارکوف، تنظیم مقادیر پارامترهای آن است که هر چه این مقادیر بهینه‌تر باشند، مدل مخفی مارکوف با دقت بیشتری قادر به پیش‌بینی احتمال مقادیر بعدی خواهد بود؛ لذا در این پژوهش سعی شده است بر مبنای تحلیل مجموعه‌داده NSL-KDD با استفاده از الگوریتم برنامه‌نویسی تکاملی، پارامترهای بهینه را برای مدل مخفی مارکوف انتخاب کرده و به نوعی آن را تعلیم دهیم؛ سپس با بهره‌گیری از آن، انواع حملات موجود در مجموعه‌داده را شناسایی کنیم. برای ارزیابی میزان موفقیت مدل پیشنهادی EHHM در ارتقای درصد صحت تشخیص نفوذ، سامانه پیشنهادی و همچنین روش قبلی در محیط شبیه‌سازی MATLAB پیاده‌سازی شده‌اند. نتایج پژوهش نشان می‌دهد، مدل EHMM، درصد تشخیص نفوذ را از متوسط ۸۷٪ (در استفاده از مدل مخفی مارکوف معمولی) به بیش از ۹۲٪ (در استفاده از مدل مخفی مارکوف تکاملی) افزایش می‌دهد. همچنین پس از آموزش کامل داده آموزشی به هر دو روش مبتنی بر مدل مارکوف معمولی و تکاملی، زمان آموزش سامانهٔ مورد نظر برای یک مجموعه‌دادهٔ حدود شامل دویست‌هزار رکوردی، از متوسط ۴۸۹ دقیقه در روش معمولی به کم‌تر از چهارصد دقیقه در روش پیشنهادی کاهش یافته است. حصول این نتیجه و عملیاتی‌کردن آن در سامانه‌های تشخیص نفوذ، می‌تواند موجب ارتقای توان دفاعی کشور در مقابل هجمه‌های سایبری دشمن شود.

واژگان کلیدی: امنیت اطلاعات،تشخیص نفوذ، مدل مخفی مارکوف، الگوریتم برنامه‌ریزی تکاملی، مجموعه‌دادهٔ NSL.

۱- مقدمه

با افزایش سرعت، اندازه و تعداد رایانه‌ها بعد از سال ۱۹۷۰، اهمیت امنیت سامانه‌های رایانه‌ای افزایش یافت و با توجه به افزایش حجم داده‌های ذخیره‌شده در دنباله‌های ممیزی، مرور و تحلیل دستی آن‌ها مشکل شد. جیمز اندرسون نخستین کسی بود که مسألهٔ مرور خودکار دنباله‌های ممیزی را مطرح ساخت. تجربیات حاصله از ممیزی سامانه‌ها و ردیابی برخی از وقایع از روی دنباله‌های ممیزی به‌همراه ایده‌های مطرح‌شده توسط اندرسون، منجر به ظهور سامانه‌های تشخیص تهاجم در سال ۱۹۸۰ شد [۱]. متأسفانه سامانه‌های تشخیص نفوذ، حجم غیرقابل‌کنترلی هشدار تولید می‌کنند. در کنار مشکل حجم بالای هشدارها، مشکل مهم دیگر تولید حجم بالای هشدارهای غلط و تکراری است که درنهایت این مشکلات منجر به ناکارایی سامانه خواهد شد [۳]. با توجه به اهمیت کارایی سامانه‌های تشخیص نفوذ، در این مقاله، مسأله اصلی ارائه یک راه‌کار جدید مبتنی بر روش‌های یادگیری ماشینی به‌منظور کنترل حجم هشدارها و کاهش نرخ هشدارهای غلط و تکراری تولیدشده به‌وسیلهٔ سامانهٔ تشخیص نفوذ است. طولانی‌بودن

*نویسندهٔ مسؤول

زمان آموزش موجب کارایی پایین سامانه می‌شود؛ لذا در مدل پیشنهادی سعی می‌شود با استفاده از الگوریتم تکاملی EP [۶, ۷] زمان آموزش را تا حد قابل‌توجهی کاهش دهیم. در قسمت تحلیل‌گر ایده استفاده از مدل مخفی مارکوف تکاملی مطرح شده است، مزیت اصلی استفاده از الگوریتم تکاملی که روش ارائه‌شده نسبت به مدل معمولی مارکوف دارد، نخست این‌که سبب بهبود تشخیص نفوذ و دوم این‌که نرخ هشدار غلط را در سامانه به‌نحو چشم‌گیری کاهش می‌دهد؛ همچنین در مدل پیشنهادی امکان تشخیص حملات با دقت تشخیص بالا وجود خواهد داشت؛ ازاین‌رو، در راه‌کار ارائه‌شده سعی می‌شود، تحلیل دقیق‌تری از هشدارها به‌منظور کاهش حجم هشدارها و تمایز هشدارهای مهم از هشدارهای غلط و تکراری انجام شود که می‌تواند تأثیر به‌سزایی در مدیریت هشدارها داشته باشد. ازجمله روش‌های هوشمند متداول که امروزه مورداستفاده قرار می‌گیرند، شبکه‌های عصبی، منطق فازی، روش‌های داده‌کاوی، الگوریتم ژنتیک و مدل مخفی مارکوف هستند. با توجه به عدم وجود داده آموزشی کافی برای آموزش سامانه‌های تشخیص نفوذ در پاسخ به حملات، نیاز به الگوریتمی داریم که بتواند با تعداد رکورد داده آموزشی، بیشترین کارایی را داشته باشد. مدل مخفی مارکوف یا HMM هم در اواخر دهه ۱۹۶۰ میلادی معرفی شد و در حال حاضر به‌سرعت در حال گسترش دامنه کاربردها است. دو دلیل مهم برای این مسأله وجود دارد: نخست این که این مدل ازلحاظ ساختار ریاضی بسیار قدرتمند و به همین دلیل مبانی نظری بسیاری از کاربردها را شکل داده است. دوم این که مدل مخفی مارکوف برخلاف روش‌های دیگر با تعداد نمونه آموزشی کم، می‌تواند کارایی قابل قبولی داشته باشد [۸]. ادامه این مقاله به‌شرح زیر خواهد بود: در بخش دوم مفاهیم پایه بیان می‌شوند؛ در بخش سوم کارهای انجام‌شده در این زمینه مورد بررسی و ارزیابی قرار می‌گیرند؛ در بخش چهارم روش پیشنهادی و مؤلفه‌های مورد نیاز آن بیان می‌شود؛ در بخش پنجم به ارزیابی و تحلیل روش پیشنهادی خواهیم پرداخت و در بخش آخر نتیجه‌گیری و کارهای آینده آمده است.

۲-پیشینهٔ پژوهش

پژوهش‌های انجام‌شده نشان می‌دهد که در سال‌های اولیهٔ مطرح‌شدن سامانه‌های تشخیص نفوذ، بیش‌تر کارها مبتنی بر فرخوان‌های سیستمی و دنباله‌های ممیزی بودند؛ بعد از آن سامانه‌های تشخیص نفوذ مبتنی بر میزبان مطرح شدند و همچنین در چند سال اخیر با گسترش حملات تحت وب، چندین روش تشخیص نفوذ در این زمینه نیز ارائه شده است. استفادهٔ گسترده از سامانه‌های تشخیص نفوذ مبتنی بر میزبان و شبکه و همچنین از یک طرف حجم تولید هشدار بالا و از طرف دیگر اهمیت تشخیص درست بالای سامانه‌ها، موجب شده بیشتر پژوهش‌گران به سمت روش‌های ترکیبی مبتنی بر یادگیری ماشین برای رسیدن اهداف خود پیش روند؛ لذا بیش‌تر کارهای انجام‌شدهٔ اخیر بر همین اساس هستند؛ بنابراین کارهای انجام‌شده در چهار دسته قرار می‌گیرند:

- سامانه‌های تشخیص نفوذ مبتنی بر فرخوان‌های سیستمی
- سامانه‌های تشخیص نفوذ مبتنی بر میزبان
- سامانه‌های تشخیص نفوذ مبتنی بر وب
- سامانه‌های تشخیص نفوذ مبتنی با رویکرد ترکیبی

آقای چوی و همکارش [۱۳] در دانشگاه ینسی، سئول-کره جنوبی، یک سامانهٔ تشخیص ناهنجاری مبتنی بر فراخوان‌های سیستمی و بر پایه روش‌های هوش مصنوعی ارائه کرده‌اند. در این روش داده‌های ورودی به‌وسیله شبکه عصبی پیش‌پردازش، سپس داده‌های پیش‌پردازش‌شده به‌وسیلهٔ چندین مدل مخفی مارکوف پردازش و تحلیل می‌شوند، بعد از تحلیل فعالیت‌های ورودی، رأی‌گیری از تمامی خروجی‌های به‌دست‌آمده برای تعیین این‌که فعالیت جاری نرمال یا حمله است، انجام می‌شود. نگارندگان این مقاله بر این باورند که به‌طورمعمول روش‌های مبتنی بر رأی‌گیری می‌توانند نرخ تشخیص بیشتری را داشته باشند. از مزایای این روش سادگی و مقرون به صرفه‌بودن و همچنین در جایی که داده‌های سنگین نباشد، مناسب است؛ از معایبی که می‌توان برای این روش بیان کرد، این است که اندازه بردارهای شبکه عصبی SOM باید از قبل مشخص باشد؛ بنابراین در هر بار تحلیل فعالیت‌های جاری نیاز مجدد به تعیین تعداد بردارهای ورودی برای پیش‌پردازش فعالیت‌ها دارد؛ همچنین با توجه به این که فعالیت‌های ورودی الگوهای یکسانی ندارند، فرآیند یافتن تعداد بردار بهینه برای SOM به‌ازای هر ورودی تکرار می‌شود؛ بنابراین زمان زیادی صرف مرحله پیش‌پردازش می‌شود. از معایب دیگر این روش استفاده از الگوریتم بام-ولش برای آموزش HMM است. با توجه به اینکه بام-ولش یک الگوریتم مبتنی بر شیب نزولی است و این که امکان گیرافتادن در بهینگی محلی را دارد، در بسیاری از موارد نمی‌تواند پارامترهای بهینه HMM را پیدا

[DOR: 20.1001.1.24763047.1398.8.2.1.9]

کند؛ بنابراین کارایی این روش با داشتن معایب بالا نمی‌تواند درحد معقول باشد.

آقای هوانگ [۱۴] در دانشگاه ملبورن-استرالیا، یک سامانهٔ تشخیص نفوذ مبتنی بر HMM و فراخوان‌های سیستمی پیشنهاد کرده است. این مقاله مدل مخفی مارکوف را یک الگوریتم اثبات‌شده می‌داند که می‌تواند به‌خوبی رفتارهای نرمال را از رفتارهای ناهنجار تشخیص بدهد. ایشان معتقد است که یکی از معایبی که در روش‌های قبلی مبتنی بر فراخوان‌های سیستمی وجود داشته است، نیاز به محاسبات سنگین منابع در فرآیند آموزش HMM است که باعث ناکارآمدی سامانهٔ تشخیص نفوذ خواهد شد. بنابراین در این مقاله روشی برای بهبود فرآیند آموزش HMM پیشنهاد شده است. روش ارائه‌شده از ایده تقسیم و غلبه مشاهده‌های طولانی استفاده کرده است. یکی از مزایایی که این مقاله تضمین داده امکان استفاده در محیط‌های بلادرنگ است؛ اما در کنار این مزیت یکی از معایب عمده این روش عملیات سنگین تبدیل فراخوان‌های سیستمی طولانی به چندین بخش مجزا است که باعث تأخیر در عملیات سیستم و درنهایت خروجی سامانهٔ تشخیص نفوذ و کارایی آن می‌شود.

آقای ارستون و همکارانش [۱۵] در دانشگاه تگزاس-آمریکا، یک سامانهٔ تشخیص نفوذ مبتنی بر وب بر پایه HMM برای تشخیص حملات چندمرحله‌ای پیشنهاد کرده است. این مقاله با استفاده از فراخوان‌های سیستمی برای تشخیص حملات پیچیده اینترنتی چندمرحله‌ای از مدل مخفی مارکوف استفاده کرده است. حملات چندمرحله‌ای، برخلاف سایر حملات شبکه مثل Dos و Scan، در چندین بازه‌های زمانی رخ دهد. با توجه به اینکه حملات پیچیده اینترنتی به نسبت سایر حملات معمول شبکه کمتر رخ می‌دهد و الگوهای حملات کمتری دارند؛ بنابراین یکی از دلایل اصلی استفاده از HMM در این مقاله این است که برخلاف سایر الگوریتم‌های یادگیری ماشین، این الگوریتم با داده آموزشی خیلی کم هم می‌تواند به‌خوبی تعلیم داده شود و حملات یادشده را با درصد احتمال موفقیت بیشتری شناسایی می‌کند. از معایب این روش وجود هشدارهای تکراری با حجم بالاست و هیچ سازوکاری برای کنترل و مدیریت آن‌ها وجود ندارد که این امر باعث می‌شود زمان پردازش سامانهٔ تشخیص نفوذ افزایش یابد.

آقای سونگ و همکارانش [۱۶] در دانشگاه برکلی، کالیفرنیا-آمریکا، یک سامانهٔ تشخیص نفوذ مبتنی بر میزبان طراحی کردند. هدف آن‌ها تشخیص حملات تقلید است؛ ایده اصلی این روش استفاده از مدل مخفی مارکوف یا HMM در محاسبه فاصله رفتاری داده‌های سطح پایین مثل فراخوان‌های سیستمی است. با محاسبه فاصله رفتاری داده‌هایی را که ازلحاظ رفتار باهم شباهت دارند، یعنی فاصله رفتاری کمی دارند، شناسایی می‌کند و به‌احتمال زیاد امکان به‌روزرسانی حمله تقلب وجود دارد.

آقای هسلوم و همکارانش [۱۷] در دانشگاه تروندهایم-نروژ، یک سامانهٔ تشخیص نفوذ بلادرنگ ترکیبی مبتنی بر مدل مخفی مارکوف و فیلترینگ تجمعی را پیشنهاد دادند. اساس کار این مقاله به این صورت است که داده‌های شبکه توسط حسگرهای فعال جمع‌آوری‌شده و سپس به HMM داده می‌شود تا بتواند با تحلیل آن حملات مخرب را شناسایی کند. در این روش نقش HMM به‌عنوان یک پیش‌بینی‌کننده است که با دریافت اطلاعات جاری مهاجم، نقشه او را برای گام بعدی پیش‌بینی می‌کند. برای رسیدن به این هدف ابتدا HMM با استفاده از یک سری داده آموزشی مربوط به اطلاعات رفتاری نفوذگران تعلیم می‌بینند، سپس در مرحلهٔ بعدی HMM به‌صورت برخط و با استفاده از داده‌های تعلیم‌یافته می‌تواند رفتار مهاجم را در گام بعدی پیش‌بینی کند؛ یعنی رویکرد اصلی این مقاله بر مبنای یافتن نوعی تعامل بین سامانه و مهاجمان شبکه در مراحل مختلف است. از معایب این روش می‌توان به مواردی مثل: امکان تبدیل تمامی اطلاعات حس‌گرها به مشاهدات قابل‌فهم HMM وجود ندارد و همچنین با توجه به برخط‌بودن سامانهٔ تشخیص نفوذ طراحی‌شده عملیات تبدیل اطلاعات ورودی حس‌گرها به دنباله مشاهده‌ها زمان‌بر خواهد بود.

آقای توسان [۱۸] در دانشگاه آنکارا-ترکیه، یک سامانهٔ تشخیص نفوذ ترکیبی برخط مبتنی بر خوشه‌بندی و مدل مخفی مارکوف پیشنهاد کرده‌اند. این روش بر مبنای تجزیه‌وتحلیل بسته‌های اطلاعاتی جریانی سرورها و تشخیص نوع پروتکل‌ها کار می‌کند. با توجه به نیازمندی یک سازمان به پایبندی به سیاست‌های امنیتی و همچنین جلوگیری از فاش‌شدن اطلاعات درون‌سازمانی که درنهایت موجب به خطرافتادن منابع سیستم می‌شود، وجود یک سامانهٔ تشخیص نفوذ کارا و مطمئن بیش‌ازپیش موردتوجه خواهد بود. برای رسیدن به این دستاورد، آقای توسان ایده استفاده مدل مخفی مارکوف را برای تشخیص رفتار کاربران در شبکه پیشنهاد کرده است. از مزایای این روش امکان به‌کارگیری

[DOR: 20.1001.1.24763047.1398.8.2.1.9]

در محیط‌های رمزشده و همچنین تشخیص حملات چندمرحله‌ای است. از معایب این روش، با توجه به امکان تغییر شماره بسته توسط مهاجم، این روش نیازمند پشتیبانی دائمی مدیر سامانه دارد.

آقای دبابراتا و همکارنش [۱۹] در دانشگاه سامبلپور- هندوستان، یک سامانهٔ تشخیص نفوذ فقط برای شناسایی حملات تزریق[1] SQL با استفاده از مدل مخفی مارکوف پیشنهاد کرده‌اند. تزریق SQL یک روش حمله است که هدف آن داده‌های ساکن در پایگاه داده‌ای است که از طریق فایروال محافظت می‌شود. آقای دبابراتا یک سامانهٔ تشخیص نفوذ مبتنی بر مدل مخفی مارکوف ارائه کرده که قادر است تا حدود قابل‌توجهی حملات تزریق را که به‌طورمعمول به‌صورت کوئری‌های مشکوک هستند، حتی در محیط‌های چندگانه شناسایی کند. تمرکز اصلی این روش برای شناسایی کوئری‌های مشکوک بر روی سرویس‌های وب و سرورهای پایگاه داده است. با این کار می‌توان کوئری‌های دریافتی را ازلحاظ مشکوک یا نرمال‌بودن قبل از ارسال به سرورهای پایگاه داده بررسی کرد. از مزایای این روش با توجه به این‌که به‌صورت بلادرنگ عمل می‌کند، می‌توان به نرخ پیش‌بینی بالا و همچنین نرخ بالای تشخیص درست حملات اشاره کرد. یکی از عیب‌های این روش عدم مقایسه با روش‌های مشابه دیگر است. همچنین عیب دیگر آن سربار زیادی است که به سامانه وارد می‌شود، این سربار ناشی از استخراج کامل کوئری‌ها، نرمال‌سازی و تبدیل به دنباله قابل‌فهم است، که این عملیات زمان و هزینه زیادی را می‌طلبد.

آقای ریچارد و همکارش [۲۰] در دانشگاه مرگان - امریکا، یک سامانهٔ تشخیص نفوذ مبتنی بر مدل مخفی مارکوف پیشنهاد کرده‌اند؛ ایشان بر این باور است که روش‌های تشخیص نفوذ مبتنی بر مدل مخفی مارکوف کارایی بالای آن در عملیات تشخیص نفوذ و همچنین موفقیت‌های آن در کاربردهایی چون تشخیص الگو، استفاده گسترده در تشخیص گفتار و همچنین تشخیص دستخط است. به‌طورمعمول در مواردی که از HMM در فرآیند تشخیص نفوذ استفاده می‌شود، شامل دو مرحله کلی است: مرحله نخست فرآیند آموزش HMM است و مرحله دیگر فرآیند تصمیم‌گیری HMM براساس آموزش مرحله قبلی و پایگاه داده آموزشی آن است. گفتنی است که این مقاله به‌عنوان مقاله پایه و مقایسه نتایج آن با روش پیشنهادی انتخاب شده است.

[1] SQL Injection

۳- مفاهیم پایه

۳-۱- تشخیص نفوذ

امروزه بیش از گذشته، حجم داده‌های موجود در سامانه‌های سایبری افزایش‌یافته است؛ هم‌زمان تهدیدهای پیچیده و وسیع امنیتی جهت نفوذ به این سامانه‌ها نیز افزایش یافته است. در حال حاضر با توجه به پیشرفته و پیچیده‌شدن حملات، استفاده از سامانه تشخیص نفوذ جهت شناسایی حملات و جلوگیری از نفوذگران، اجتناب‌ناپذیر است. با توجه به گسترش شبکه‌های رایانه‌ای منابع اطلاعاتی نیز گسترش یافته و زمان تحلیل آن‌ها طولانی‌تر و پیچیده‌تر شده است. سامانه‌های تشخیص نفوذ برای این که قادر به تشخیص یک حمله باشند به یک مجموعه‌داده برای مطابقت حملات با آن نیاز دارند یا به‌طورکلی هر سامانه مبتنی بر پردازش داده برای انجام عملیات خود به منابع اطلاعاتی برای ردگیری اهداف خود نیاز دارد؛ بنابراین یکی از نیازهای اولیه سامانه‌های تشخیص نفوذ به‌دست‌آوردن منابع اطلاعاتی است که متناسب با نوع سامانهٔ تشخیص نفوذ (تشخیص حمله یا جلوگیری از حمله) ممکن است متفاوت باشد. سامانه‌های تشخیص نفوذ به‌طور کلی به سه دسته تقسیم می‌شوند [۴]:

۳-۱-۱- سامانه‌های تشخیص نفوذ مبتنی بر میزبان

از زمان مطرح‌شدن سامانه‌های تشخیص نفوذ تا سال ۱۹۹۰ سامانه‌های تشخیص تهاجم به‌طور عمده، مبتنی بر میزبان بودند که تحلیل آن‌ها فقط بر اساس داده‌های حاصل از دنباله‌های ممیزی سیستم‌عامل یا منابع اطلاعاتی میزبان‌های دیگر بود. سامانه‌های تشخیصی مبتنی بر میزبان وظیفه نظارت بر اتفاقات رخ‌داده‌شده بر روی میزبان را بر عهده‌دارند. که این نظارت با جستجوکردن در فایل‌های سیستمی انجام می‌شود و با این کار هرگونه تغییرات ایجادشده در آن‌ها توسط نفوذگران را پیدا می‌کند. سامانه‌های تشخیص نفوذی که از منابع اطلاعاتی مبتنی بر میزبان بهره می‌گیرند، روی اطلاعات جمع‌آوری‌شده از داخل یک سیستم (میزبان) عمل می‌کند. این روش نظارتی، به سامانه‌های مبتنی بر میزبان این امکان را می‌دهد که فعالیت‌ها را با قابلیت اطمینان و دقت بالایی تحلیل کند و به‌طوردقیق مشخص کنند که کدام پردازه‌ها و کاربران در انجام یک حمله خاص دست داشته‌اند؛ درضمن، این سامانه‌ها به‌راحتی می‌توانند نتیجه یک حمله را (چه با موفقیت و چه با شکست باشد)، بر روی سیستم موردنظر ببینند. سامانه‌های تشخیصی از این نوع دارای

[DOR: 20.1001.1.24763047.1398.8.2.1.9]

معماری ساده و کم‌هزینه‌ای هستند؛ البته از معایب روش این است که در کشف حملاتی که کل شبکه را مورد هدف قرار می‌دهند، نمی‌تواند عملکرد خوبی داشته باشد.

۲-۱-۳- سامانه‌های تشخیص نفوذ مبتنی بر شبکه

روال کار سامانه‌های تشخیصی مبتنی بر شبکه با کارکردن مستقیم بروی ترافیک شبکه انجام می‌شود. با توجه به این که یکی از منابع مهم اطلاعاتی این سامانه‌ها ترافیک شبکه است، بنابراین قادر به استفاده از بسته‌های ترافیک موجود بر روی یک بخش به‌عنوان داده‌های خود است. که این کار با قراردادن کارت واسط شبکه در مُد تصادفی صورت می‌پذیرد؛ مُد تصادفی حالتی است که در آن کارت شبکه به‌گونه‌ای تنظیم شود که برای کل ترافیک شبکه وقفه تولید کند. یکی از مهم‌ترین مزیت‌های این روش نسبت به روش قبلی این است که هزینه نظارت بر منابع اطلاعاتی خیلی کم است؛ چون نظارت می‌تواند به‌سادگی با خواندن بسته‌هایی که در بخشی از شبکه جابه‌جا می‌شوند، بدون آنکه این نظارت تأثیری بر کارایی دیگر سیستم‌های موجود در شبکه داشته باشد صورت پذیرد، به همین دلیل اضافه‌کردن یک سامانهٔ تشخیص تهاجم مبتنی بر شبکه، به یک شبکه موجود با کمترین تغییر در آن شبکه امکان‌پذیر است. همچنین یکی از مهم‌ترین معایب این روش این است که سامانه‌های تشخیص تهاجم مبتنی بر شبکه با وجود سادگی و قدرتشان، به‌طورمعمول نمی‌توانند در شبکه‌های مدرن سوئیچی کار کنند؛ چون سوئیچ‌ها، اتصالات بین میزبان‌ها را از هم مجزا می‌سازند؛ به‌طوری‌که یک میزبان فقط بتواند ترافیکی که به مقصد آن نشانی‌دهی شده ببیند.

۳-۱-۳- سامانه‌های تشخیص نفوذ ترکیبی

سامانهٔ تشخیص نفوذ ترکیبی که گاهی توزیع‌شده گفته می‌شود از ترکیب دو روش قبلی بهره می‌برد؛ به‌گونه‌ای که سطح امنیت و انعطاف‌پذیری بیشتری را به‌دنبال خواهد داشت؛ که این توزیع‌شدگی نه‌تنها در جمع‌آوری اطلاعات بلکه در تحلیل داده‌ها نیز استفاده می‌شود. از خصوصیات مهم این روش امکان ردیابی کاربر در شبکه تحت نظارت است؛ که با توجه به دشواربودن ردیابی کاربران و فایل‌های ردوبدل‌شده بین آن‌ها در یک شبکه بزرگ، قابلیت مهمی به‌شمار می‌رود. بدین ترتیب تشخیص و جلوگیری از حمله‌های چندنفره و توزیع‌شده، می‌تواند به‌صورت کارا به‌وسیلهٔ این سامانه صورت پذیرد. برای بهبود کارایی این روش دو روش وجود دارد: الف) تقسیم ترافیک: این روش بیشتر براساس جریان‌های داده‌ای و سیاسیت‌های امنیتی و ساختار IDS کار می‌کند. ب) متعادل‌کردن بار: در هر زمان مقدار بار مناسبی برای هر یک کدام از حس‌گرها در نظر می‌گیرد، به‌نحوی‌که از ظرفیت سامانه به‌طور بهینه استفاده شود. این روش به‌دلیل ترکیب دو روش قبلی، پیچیدگی زیادی را دارد. در جدول (۱) نقاط ضعف و قوت هر کدام از روش‌ها بیان شده است.

(جدول-۱): مقایسه روش‌های تشخیص نفوذ با منابع اطلاعاتی مختلف

معیارهای کارایی	تشخیص مبتنی بر میزبان	تشخیص مبتنی بر شبکه	تشخیص توزیع شده
بازدارندگی در برابر نفوذ	بازدارندگی بالا در برابر نفوذهای داخلی.	بازدارندگی بالا در برابر نفوذهای خارجی.	بازدارندگی بالا در برابر نفوذهای داخلی و خارجی.
زمان پاسخ	تأخیر بالا در محیطهای بلادرنگ اما مناسب برای محیطهایی با داشتن زمان کافی است.	مناسب برای محیطهای بلادرنگ.	برای محیطهایی با حجم تراکنش بالا مثل بانک ها و بیمارستان ها.
ارزیابی آسیب	مناسب در تعیین نقاط آسیب دیده .	خیلی ضعیف در تعیین نقاط آسیب دیده.	ضعیف در تعیین نقاط آسیب دیده.
جلوگیری از نفوذ	مناسب در جلوگیری ازنفوذهای داخلی	مناسب در جلوگیری ازنفوذهای خارجی	مناسب در جلوگیری ازنفوذهای داخلی و خارجی
پیش بینی تهدید	مناسب در پیش بینی و تشخیص الگوهای رفتاری مشکوک داخلی.	مناسب در پیش بینی و تشخیص الگوهای رفتاری مشکوک خارجی.	مناسب در پیش بینی و تشخیص الگوهای رفتاری مشکوک داخلی و خارجی

۲-۳- مدل مخفی مارکوف تکاملی

یکی از چالش های اصلی مدل مخفی مارکوف[1] سنتی، عدم انتخاب پارامترهای بهینه است که درنهایت موجب ناقص‌بودن فرآیند تعلیم برای HMM می شود؛ لذا در این مقاله ما ایده استفاده از الگوریتم برنامه‌ریزی تکاملی[2] را که یکی از الگوریتم‌های تکاملی محسوب می‌شود، در فرآیند تعلیم HMM پیشنهاد کرده‌ایم؛ که در بخش بعدی به تشریح اجزای آن می پردازیم.

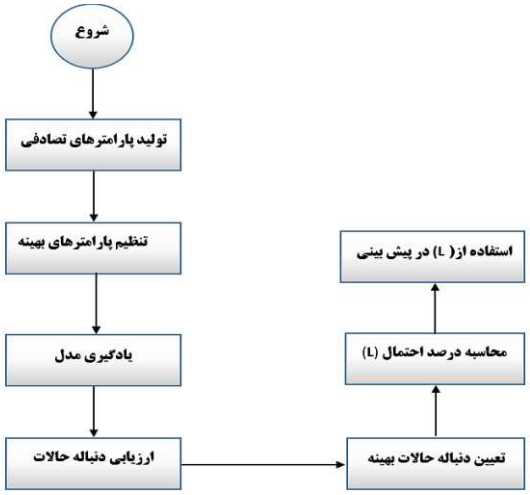
[1] Hidden Markov Model (HMM)

[2] Evolutionary Programming (EP)

[DOR: 20.1001.1.24763047.1398.8.2.1.9]

۱-۲-۳- الگوریتم مدل مخفی مارکوف

مدل مخفی مارکوف در اواخر دهه ۱۹۶۰ میلادی معرفی شد و کاربردهای درخشانی در مسائل پردازش سیگنال گفتار، تشخیص دست‌خط و تشخیص نقص دستگاه‌های الکترونیکی کسب کرد و در حال حاضر به‌سرعت در حال گسترش دامنهٔ کاربرد است. HMM به فرایندی تصادفی گفته می‌شود که احتمالات آیندهٔ آن از طریق مقادیر اخیر آن محاسبه می‌شود. به‌عبارتی دیگر یک مدل مارکوف آماری است که در آن سامانه مدل‌شده به‌صورت یک فرایند مارکوف با حالت‌های مشاهده‌نشده (پنهان) فرض می‌شود. یک مدل پنهان مارکوف می‌تواند به‌عنوان ساده‌ترین شبکه بیزی پویا در نظر گرفته شود [۱۱]. برای مثال فرض کنید، که به شما سکه‌هایی برای پرتاب‌کردن داده می‌شود، در اینجا برای هر حالت متناظر(شیر/خط) یک احتمال قابل‌مشاهده و مشخص وجود دارد، که می‌توان با مدل مارکوف معمولی آن را مدل کرد؛ اما اگر شما در یک اتاق باشید و در اتاق مجاورشان فرد دیگری سکه‌هایی را به هوا پرتاب می‌کند و بدون اینکه به شما بگوید این کار را چگونه انجام می‌دهد و تنها نتایج را به اطلاع شما می‌رساند. در این حالت شما با فرآیند مخفی انداختن سکه‌ها و با دنباله‌ای از مشاهدات شیر یا خط مواجه هستید، در این شرایط ما با یک فرآیند دوگانه (انداختن سکه و دنباله مشاهدات) سروکار داریم که این همان HMM است.



(شکل-۱): روند اجرای الگوریتم HMM

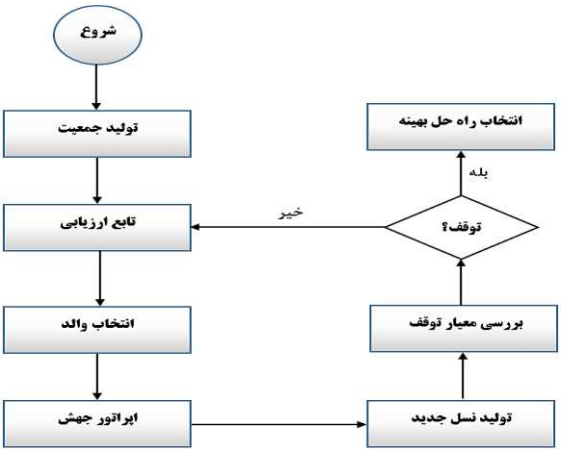
مطابق شکل (۱) در به‌کارگیری HMM سه مسألهٔ اساسی وجود دارد که هریک به‌نوعی در کاربردهای مختلف بر روی کارایی مدل اثر می‌گذارند؛ اما جهت فرآیند تشخیص نفوذ، مسأله یادگیری مدل از اهمیت ویژه‌ای برخوردار است. الگوریتم بام-ولش[1]، به‌طورمعمول به‌عنوان الگوریتم یادگیر یا همان تنظیم‌کننده پارامتر، در HMM استفاده می‌شود؛ اما این الگوریتم مبتنی بر شیب نزولی بوده و امکان گیرکردن در بهینگی محلی و درنهایت عدم انتخاب پارامتر بهینه برای HMM است [۱۲,۱۰]؛ لذا ایده استفاده از EP برای این کار در ادامه مقاله پیشنهاد شده است.

۳-۳- الگوریتم برنامه‌ریزی تکاملی

امروزه، گونه‌های زیادی از الگوریتم‌های تکاملی وجود دارند، که پایه و اساس همه آن‌ها برای رسیدن به هدف یکی است. با داشتن جمعیتی از گونه‌ها[2]، فشار محیطی باعث انتخاب می‌شود (القای بهترین[3]) و این، افزایش شایستگی[4] جمعیت را نتیجه می‌دهد. با داشتن یک تابع کیفیتی که می‌خواهیم بیشینه شود، می‌توان مجموعه‌ای از جواب‌های نامزد را به‌طور تصادفی تولید کرد و تابع کیفیت را به‌عنوان معیاری برای محاسبه شایستگی به کاربرد. (هر چه بیشتر، بهتر) بر اساس این شایستگی، بعضی از نامزدهای بهتر انتخاب می‌شوند، تا به‌عنوان هسته‌ای برای تولید نسل بعد به‌کار روند. دلیل انتخاب الگوریتم تکاملی این است که الگوریتم‌های تکاملی در مقایسه با سایر الگوریتم‌های بهینه‌سازی برتری‌هایی دارند که موجب شده است، به‌طور گسترده مورداستفاده قرار بگیرند. برای مثال، این الگوریتم‌ها، نیاز به معرفی کامل مسأله ندارند و تنها با داشتن اطلاعات چندی در مورد تعریف مسأله می‌توانند کار کنند. همچنین محدودیتی در مورد تابع شایستگی ندارند و لزومی ندارد که این تابع برای مثال مشتق‌پذیر باشد. علاوه‌بر این موارد، چون الگوریتم‌های تکاملی دارای جمعیتی از موجودات هستند و روی بخش‌های مختلفی از جمعیت به‌طور موازی کار می‌کنند، احتمال کمتری برای قرارگرفتن در بهینه‌های محلی[5] دارند. این قابلیت الگوریتم‌های تکاملی اجازه می‌دهد که کار بهینه‌سازی را به‌طور موازی روی چندین بخش جمعیت انجام داد. الگوریتم برنامه‌ریزی تکاملی یا EP یکی از الگوریتم‌های تکاملی است، که در سال ۱۹۶۰ توسط آقایان فوگل و لورانس در آمریکا ارائه شده است [۹]. هدف اصلی EP استفاده از تکامل به‌منظور فرآیند یادگیری در چارچوب تولید هوش مصنوعی ایجاد شده است، که این کار را از

[1] Baum-welch
[2] individual
[3] Survival of the fittest
[4] Fitness
[5] Local optima

طریق تعدادی حالت قابل شمارش یا FSM به‌عنوان موجودیت‌های پیش‌بینی‌کننده و تکامل آن‌ها در آینده، انجام می‌دهد. EP از بسیاری از جهات شبیه الگوریتم ژنتیک است، اما تفاوت‌هایی نیز دارد که موجب انتخاب آن بجای ژنتیک شده است [۱۲]: مهم‌ترین دلیل آن این است که، EP در جستجوی فضای حالت برای یافتن نسل جدید، فقط از عمل‌گر جهش[1] بهره می‌برد؛ درحالی‌که در GA هم از جهش و هم از ترکیب[2] استفاده می‌کند. دلیل دیگر این است که EP برای شکل گیری فرآیند تکامل خود نسبت به GA ساختار پایدارتری دارد. شکل (۲) فرآیند اجرای الگوریتم EP را نشان داده است.



(شکل-۲): روند اجرای الگوریتم EP

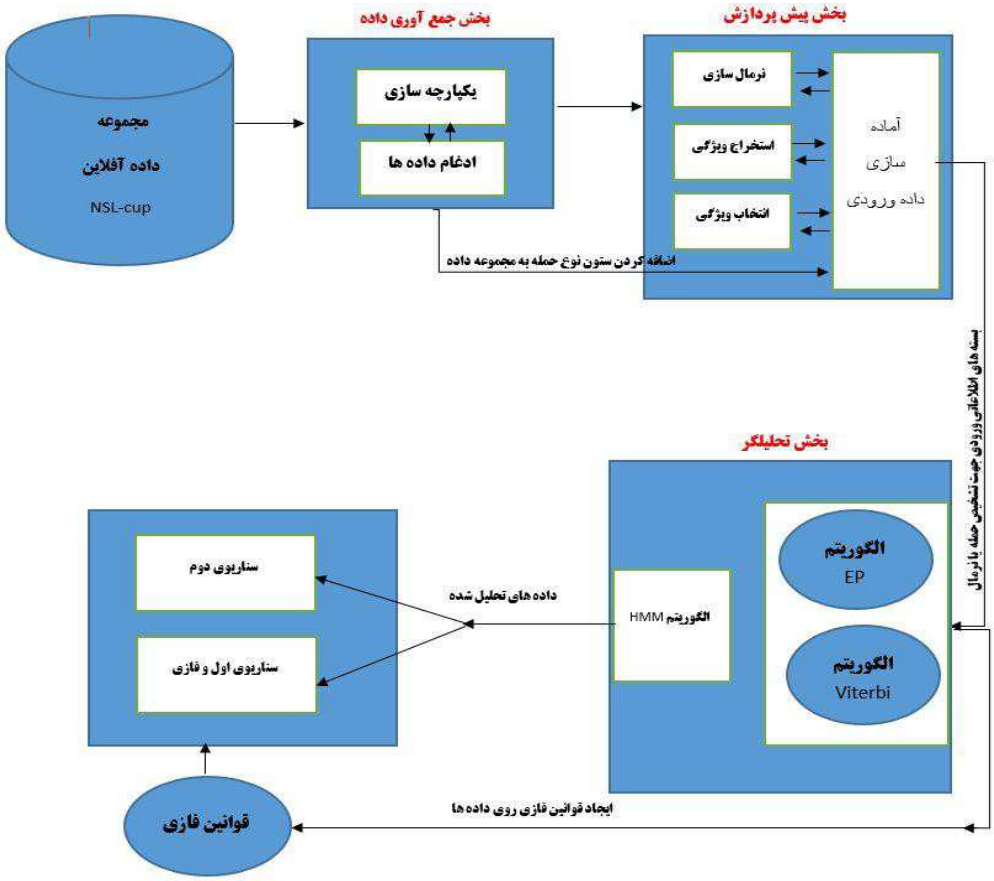
۴- روش پیشنهادی

در طرح پیشنهادی، قصد داریم با استفاده از روش‌های داده‌کاوی، الگویی از یک سامانه تشخیص نفوذ مبتنی بر تشخیص ناهنجاری ارائه کنیم که در آن با بهره‌گیری از مدل مخفی مارکوف و الگوریتم تکاملی EP، درصد تشخیص هشدارهای صحیح را ارتقا دهیم. گفتنی است، روش پیشنهادی بر مبنای دو سناریوی مختلف شبیه‌سازی شده است (جهت امکان مقایسه با روش‌های دیگر)، به‌نوعی که در سناریوی نخست فقط تشخیص نرمال یا غیرنرمال داریم، اما در سناریوی دوم در صورت غیرنرمال‌بودن رکورد جاری نوع حمله را نیز مشخص خواهیم کرد. از آنجا که در طرح پیشنهادی، آموزش مدل مخفی مارکوف، به نوعی توسط الگوریتم EP انجام می‌شود، لذا نام این طرح را EHMM[3] انتخاب کرده‌ایم.

با توجه به این که در سامانه‌های تشخیص نفوذ مبتنی بر HMM کارایی در بخش آموزش مدل، نکته‌ای بسیار کلیدی محسوب می‌شود، لذا مسأله حفظ کارایی به‌همراه افزایش سرعت در فرآیند آموزش باید مورد توجه باشد؛ بنابراین هدف اصلی در این مقاله استفاده از الگوریتم تکاملی EP، درجهت بهبود کارایی سامانهٔ تشخیص نفوذ مبتنی بر HMM در بخش آموزش مدل آن دارد. معماری کلی روش پیشنهادی در شکل (۳) قابل مشاهده است، که شامل چندین بخش است و هر کدام وظایف خاصی را بر عهده دارند. در بخش جمع‌آوری داده[4] داده‌های مورد نیاز سامانهٔ تشخیص نفوذ پس از انجام یک سری عملیات غربال‌گری به‌دست می‌آیند، که در این بخش ما از مجموعه داده آماده NSL-CUP، که توسط دانشگاه MIT جهت ارزیابی سامانه‌های تشخیص نفوذ طراحی‌شده، استفاده می‌شود. در بخش پیش‌پردازش[5] نیاز است که یک سری رکوردهای غیرضروری که موجب پیچیدگی مدل می‌شود، حذف شود و درنهایت عملیات انتخاب ویژگی نیز انجام می‌شود؛ اما بخش اصلی این معماری بخش تحلی‌گر است، جایی که الگوریتم‌های EP و HMM قرار دارند، که با دریافت داده‌ها از بخش پیش‌پردازش عملیات تحلیل روی آن‌ها را انجام می‌دهند و ماهیت هر کدام از رکوردها از لحاظ نرمال یا حمله‌بودن را تعیین می‌کند. همچنین در سناریوی نخست، با استفاده از فرآیند داده‌کاوی در مجموعه‌دادهٔ مورد استفاده، تعداد هشت قانون فازی با احتمال موفقیت بالای نود درصد مطابق جدول (۲) استخراج شده‌اند؛ با این کار تا حدودی نرخ تشخیص غلط سامانه کاهش می‌یابد. در بخش خروجی براساس خروجی الگوریتم EHMM، ماهیت بسته‌های اطلاعاتی را مشخص می‌کند، با توجه به این که به‌طورکلی در سامانه‌های تشخیص نفوذ، به‌طورمعمول دو روش پاسخ‌دهی به حملات، فعال و منفعل وجود دارد، به‌نوعی می‌توان سازوکار پاسخ‌دهی در سامانهٔ پیشنهادی را منفعل دانست؛ زیرا هیچ عملیات متقابلی برای جلوگیری از حمله وجود ندارد و فقط وظیفه تعیین ماهیت بسته جاری را ازلحاظ نرمال یا حمله‌بودن با توجه به نتیجه خروجی مؤلفه تحلیل‌گر بر عهده دارد؛ نتیجه این مؤلفه یک پایگاه داده از رکوردهای تحلیل‌شده خواهد بود، که ماهیت هر سطر آن بر اساس تحلیل‌ها و آموزش‌هایی که EHMM دیده، تعیین شده که در این بخش خروجی تولیدشده براساس دو

[1] Mutation
[2] Crossover
[3] Evolutionary Hidden Markov Model
[4] Data collection
[5] Preprocessing

سناریوی مختلف سناریوی انخست (فقط تشخیص نرمال و غیرنرمال)، سناریوی دوم(تشخیص نرمال و نوع حملات موجود) است؛ سپس در انتها می‌توان با مقایسه مقادیر پیش‌بینی‌شده به‌وسیلهٔ EHMM با مقادیر واقعی هر سطر از رکوردهای مجموعه‌داده میزان کارایی الگوریتم ارائه‌شده را با استفاده از معیارهایی مثل Recall، Accuracy، Precision و .... مورد سنجش قرار داد.



(شکل-۳): معماری پیشنهادی سامانهٔ تشخیص نفوذ مبتنی بر EHMM

(جدول-۲): قوانین فازی استخراج‌شده از مجموعه داده NSL

قوانین استخراج شده از دیتاست	درصد=قبول/ تعداد کل
If (SF='NO' , http='YES' , REJ='YES') then Normal	۳۲۶/۳۰٤ =۹۳٪
If(SF='YES',ICMP='NO',Pop_3='NO',telnet='NO',ftp='NO',ftp_data='NO',imap4='NO',tcp='YES',Login='NO') then Normal	٤۹۱۳/٤۷۹۹ =۹۷٪
If(SF='YES',ICMP='NO',Pop_3='NO',Private='NO',telnet='NO',ftp='NO',ftp_data='YES') then Normal	٦۳۳/٥٦۰=۸۸٪
If(SF='NO',http='NO',Login_yes='YES',IRC ='NO',S1='NO',SMTP='NO',X11='NO') then Abnormal	٥٥٤۲/٥٥۲۱=۹۹٪
If(SF='YES',ICMP='NO',Pop_3='YES',Private='NO') then Abnormal	۳٤۲/۳۲٤=۹٤٪
If(SF='YES',ICMP='NO',Pop_3='NO',Private='NO',telnet='YES',login_no='NO') then Abnormal	٥۰۷/٥۰٦=۹۹٪
If(SF='YES',ICMP='NO',Pop_3='NO',Private='NO',telnet='NO',ftp='NO',ftp_data='NO',gopher='NO',login='NO') then Normal	٦۰۸٥/٥۷٤٤=۹٤٪
If(SF='YES',ICMP='YES',urp_i='NO') then Abnormal	۹۲۹/۸٤۰=۹۰٪

۴-۱- تعیین پارامترهای بهنیه با EP

به‌طورکلی فرآیند آموزش HMM، به‌صورت تعیین پارامترهایی در طول یک فرآیند تکراری مشخص می‌شود. درعین‌حال مسأله پیداکردن پارامترهای بهینه برای مدل مخفی مارکوف به‌عنوان یک چالش مطرح می‌شود، و انتخاب پارامتر بهینه تأثیر قابل‌توجهی در عملکرد HMM دارد؛ بنابراین به‌کارگیری یک الگوریتم کارا و مطمئن که بتواند با بررسی تمامی معیارهای موجود در انتخاب پارامتر بهینه، بهینه‌ترین آن‌ها را برگزیند، تا در ادامه HMM به بهترین شکل تعلیم داده شود، موردتوجه است؛ لذا برای تحقق این امر الگوریتم تکاملی EP پیشنهاد داده شده است، که این الگوریتم علاوه‌بر برداشتن ویژگی جستجوی سراسری مسأله

(برای یافتن بهینه‌ترین پارامتر)، به‌علت نداشتن فرآیند تقاطع، به نسبت الگوریتم GA، سرعت بیشتری در به‌دست‌آوردن پارامترهای بهینه دارد. ازآنجاکه هدف اصلی، تکامل یافتن پارامترهای HMM در راستای انتخاب بهینه آن‌هاست، بنابراین EP با استفاده از عملیات تولید دنباله‌ای از جمعیت‌ها و همچنین عملیات انتخاب و جهش، قادر است بهینه‌ترین پارامترها را برگزیند. در ادامه مراحل به‌دست‌آوردن پارامترهای بهینه را به‌وسیله EP تشریح می‌کنیم.

در ابتدا، EP یک جمعیت تصادفی از کروموزوم‌ها را به‌عنوان پارامترهای اولیه HMM تولید می‌کند، در گام بعدی اندازه ابعاد(تعداد حالات) HMM با توجه به نیاز تعیین می‌شود که ما در اینجا ابعاد را (۳×۳) مطابق مقاله پایه [۲۰] انتخاب کرده‌ایم. در مرحله بعدی ماتریس انتقال[1] در ابتدا با عناصر تصادفی تشکیل داده می‌شود؛ به‌گونه‌ای که عناصر این ماتریس ۳×۳ بیان‌گر احتمال انتقال از هر حالت به هر حالت دیگر است.

۵- ارزیابی روش پیشنهادی

مشخصات سیستمی که EHMM روی آن اجرا شده در سیستم‌عامل ویندوز نسخه 8,1 و شامل پردازنده Core i5، حافظه 6G، هارد 1TB است. شبیه‌سازی سیستم EHMM، با استفاده از زبان برنامه‌نویسی Mathlab R2016(9,1,0,441655) انجام پذیرفته و جهت آماده‌سازی و پیش‌پردازش مجموعه‌دادهٔ مورداستفاده، از SQL-Server 2012 استفاده شده است؛ همچنین جهت بررسی الگوریتم J48 جهت انتخاب ویژگی از Weka(3,9,0) استفاده شده است؛ درنهایت برای صحت مقایسه و ارزیابی روش پیشنهادی، مقالهٔ پایه نیز در متلب شبیه‌سازی‌شده و در آن سعی شده است تمامی مقایسه‌ها در شرایطی به‌طورکامل یکسان با مجموعه داده یکسان انجام گیرد؛ همچنین برای صحت مقایسه با [۲۰]، تعداد حالات HMM برابر سه حالت و عملیات انتخاب ویژگی نیز براساس آن، برابر با هشت ویژگی مطابق جدول (۳) برگزیده شده است.

برای ارزیابی طرح پیشنهادی EHMM، ما از مجموعه داده NSL استفاده کردیم، که نحوه تحلیل آن در بخش‌های قبلی تشریح شد. به‌طورکلی در سامانه‌های تشخیص نفوذ پس از آموزش و تحلیل و بررسی مدل به‌دست‌آمده، نیازمند معیارهایی جهت تعیین اعتبار یا به‌عبارتی اعتبارسنجی مدل توسعه داده شده است؛ لذا معیارهای ارزیابی همه سامانه‌های تشخیص نفوذ، بر مبنای TP,TN,FN,FP، است؛ بنابراین با توجه به نیاز به مقایسه طرح پیشنهادی با مقاله پایه، باید تمامی معیارهای بیان‌شده آن را نیز به‌دست آوریم.

(جدول-۳): ویژگی‌های انتخاب‌شده از میان ۴۱ ویژگی موجود

duration
dst_host_count
hot
dst_host_same_src_port_rate
counts
dst_bytes
num_file_creations
attack_category

۵-۱- آزمایش

مطابق جدول (۴) ،پس از تعیین عملیات انتخاب ویژگی و نرمال‌سازی یا به‌اصطلاح همان پیش‌پردازش، باید مقادیر اولیه پارامترها برای راه‌اندازی HMM تعیین شوند؛ که این کار در ابتدا به‌صورت تصادفی انجام و در ادامه مقادیر پارامترها مطابق جدول (۵) توسط الگوریتم تکاملی EP بهینه اما قبل از بهینه‌سازی مقادیر پارامترها باید ابتدا به‌نوعی تمامی احتمالات رفتن از هر حالت به حالت دیگر، مطابق جدول (۶) تعیین شوند؛ سپس پارامترهای مدل مربوطه توسط الگوریتم EP که در طرح EHMM استفاده شده است، بهینه‌سازی می‌شوند؛ که در فرآیند تشخیص تهاجم، هر چه مقادیر این پارامترها، بهینه‌تر باشد، HMM با دقت بیشتری حملات را تحلیل و ارزیابی خواهد کرد و درنتیجه سامانهٔ تشخیص نفوذ کاراتری خواهیم داشت.

(جدول-۴): مقداردهی اولیه پارامترهای مدل

حالات	مقادیر اولیه
S1 (State)	0.7057
S2	0.6879
S3	0.1842

(جدول-۵): مقداردهی اولیه پارامترها، برای احتمال رفتن از هر حالت به حالات دیگر

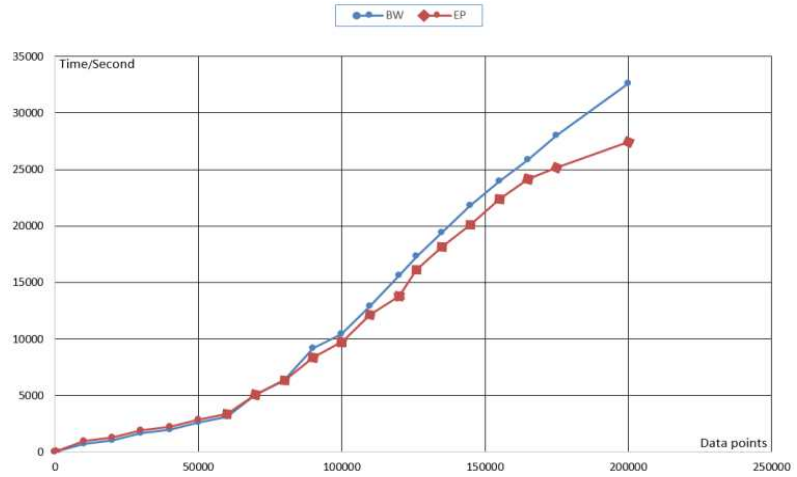
State	S1	S2	S3
S1	0.1944	0.4582	0.5794
S2	0.8791	0.1196	0.4363
S3	0.2734	0.4579	0.1133

(جدول-۶): بهینه‌سازی مقادیر توسط الگوریتم EP

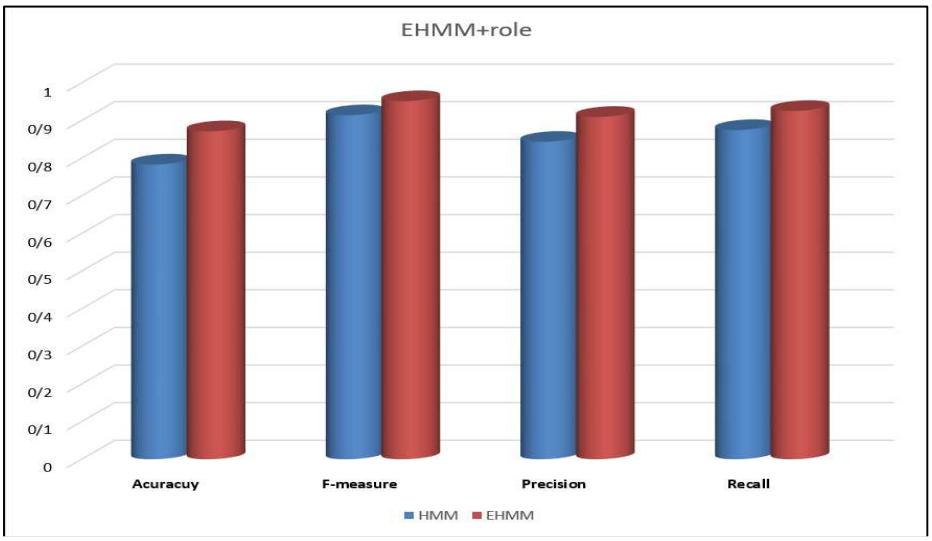
State	S1	S2	S3
S1	0.3401	0.3409	0.3999
S2	.03411	0.3433	0.3411
S3	0.3402	0.3393	0.3996

[1] Transport Matrix

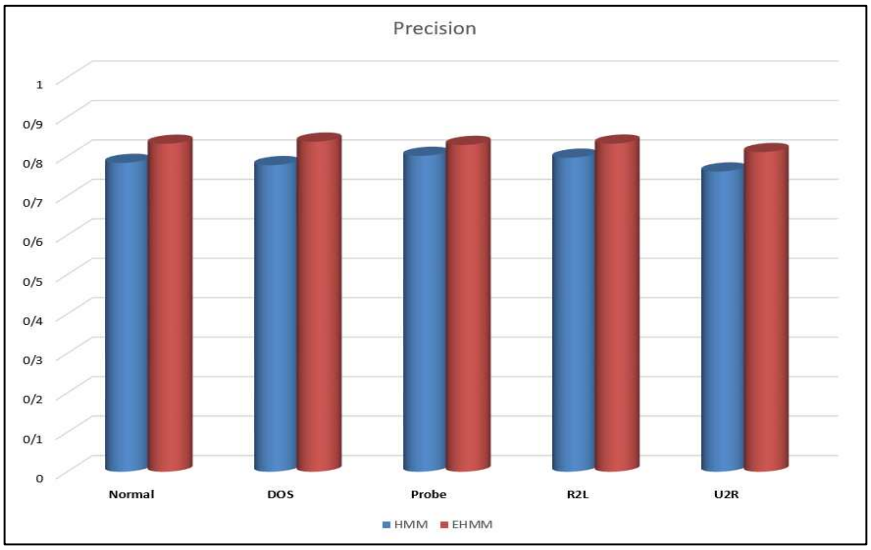
پس از تعیین پارامترها، اکنون مجموعه‌دادهٔ آموزشی را به HMM تعلیم می‌دهیم؛ لذا یکی از فاکتورهای مطرح زمان آموزش مدل است؛ که هرچه زمان کمتر باشد، مدل با سرعت بیشتری تعلیم می‌بیند. مطابق شکل (۴) نمودار رشد زمان آموزش سامانه EHMM، نشان داده شده که برای یک مجموعه‌دادهٔ دویست‌هزار عضوی، در نقاط داده‌ای مختلف، آزمایش گرفته شده است که نشان از برتری روش پیشنهادی(EP)، نمودار قرمز با افزایش تعداد رکوردهای شبکه، به نسبت روش قبلی (BW)، نمودار آبی دارد؛ اما روش قبلی(BW) در جاهایی که تعداد بسته‌ها کم باشد، سریع‌تر عمل خواهد کرد؛ که با توجه به حجم داده تولیدی در عصر امروزی، بهره‌گیری از آن منتفی خواهد بود. همان‌طور که درقبل هم بیان شد، چارچوب آزمایش‌های انجام‌شده در این مقاله بر مبنای دو سناریوی مختلف است، که در سناریوی نخست فقط تشخیص حمله از نرمال را داریم؛ اما در سناریوی دوم علاوه‌بر تشخیص حمله و نرمال، نوع حملات نیز مشخص خواهند شد. با این حال مطابق شکل (۵) نتایج آزمایش حاصل از سناریوی نخست را نشان می‌دهد که بیان‌گر بهبود چشم‌گیر روش پیشنهادی (EHMM)، به نسبت روش قبلی (HMM) [20] است که این بهبودی نه‌تنها در دقت روش بلکه در تمامی معیارهای موجود، روش پیشنهادی عملکرد بهتری داشته است.



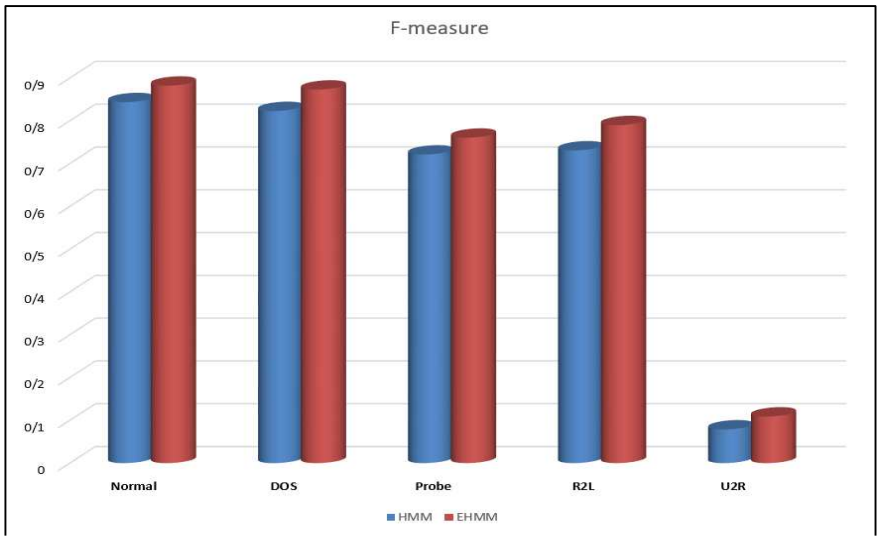
(شکل-۴): نمودار رشد مدت‌زمان آموزش و تعداد نقاط داده‌ای HMM و EHMM



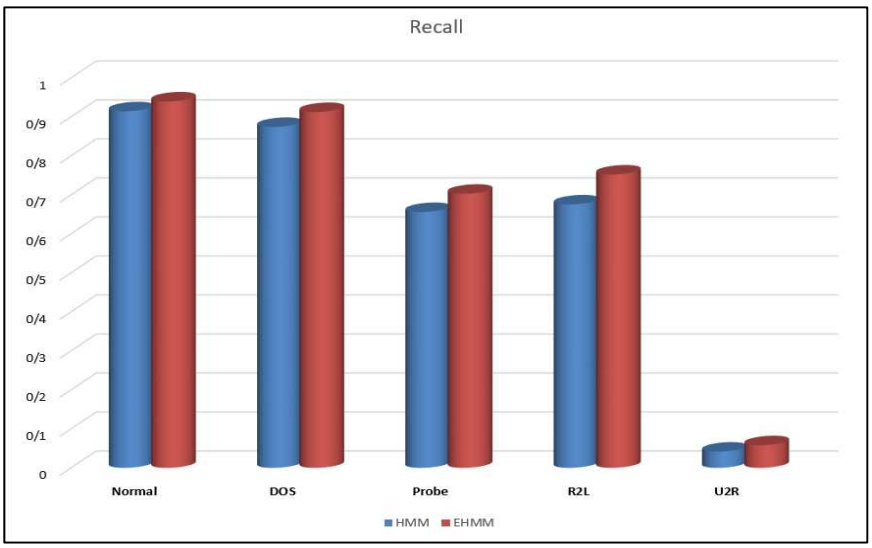
(شکل-۵): نمودار کارایی سناریوی نخست روش ارائه‌شده (EHMM) در مقایسه با روش قبلی (HMM) [20]



(شکل-۶): نمودار Recall روش ارائه‌شده (EHMM)، به تفکیک نوع حمله در مقایسه با روش قبلی (HMM) [20]

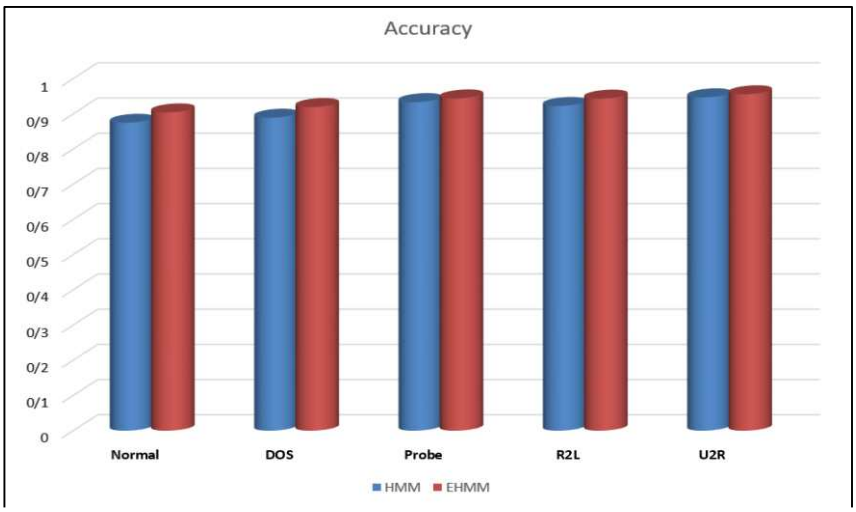


(شکل-۷): نمودار Precision روش ارائه‌شده (EHMM)، به تفکیک نوع حمله در مقایسه با روش قبلی (HMM) [20]



(شکل-۸): نمودار F-measure روش ارائه‌شده (EHMM)، به تفکیک نوع حمله در مقایسه با روش قبلی (HMM) [20]

[DOR: 20.1001.1.24763047.1398.8.2.1.9]



(شکل-۹): نمودار دقت یا درصد تشخیص حملات، روش پیشنهادی (EHMM) در مقایسه با روش قبلی (HMM)[20]

نتایج آزمایش سناریوی دوم، بر اساس تفکیک حمله و محاسبه معیارهای سنجش کارایی به‌ترتیب: recall، Precision، F-measure، به‌ترتیب مطابق شکل‌های (۶، ۷ و ۸) انجام شده است، که در هر سه معیار سنجش کارایی، نشان از برتری روش پیشنهادی EHMM، به نسبت روش HMM دارد. همچنین مطابق شکل (۹)، نمودار دقت روش پیشنهادی و روش قبلی نشان داده شده است. این معیار دقت، از مهم‌ترین معیار برای سنجش عملکرد یک سامانه تشخیص نفوذ است؛ که روش EHMM، در تشخیص حملات از نرمال‌ها و همچنین تفکیک نوع حمله، از HMM بهتر عمل کرده است. اوج کار روش پیشنهادی در تشخیص حملات از نوع U2R است؛ که با درصد تشخیص بالای ۹۵ درصد توانسته، عملکرد قابل قبولی را داشته باشد و یا به‌عبارتی‌دیگر طرح پیشنهادی EHMM، در تفکیک انواع حملات به موفقیتی بالای نود درصد رسیده است؛ که همه این نمودارها بیان‌گر تأثیر بسیار زیاد پارامترهای بهینه بر روی سامانه‌های تشخیص نفوذ مبتنی بر مدل مخفی مارکوف دارد؛ که در این پژوهش این مسأله را با استفاده از الگوریتم تکاملی EP، مرتفع کردیم.

۶- نتیجه‌گیری و کارهای آینده

روند روبه‌رشد فناوری اطلاعات و پیشرفت فناوری و علوم رایانه و همچنین گرایش بشر به بهره‌گیری از سامانه‌ها و شبکه‌های رایانه‌ای در جمع‌آوری، نگه‌داری و انتقال اطلاعات، موجب احساس نیاز به تأمین امنیت در سامانه‌ها و شبکه‌های رایانه‌ای شده است؛ لذا در این مقاله یک سامانه تشخیص تهاجم مبتنی بر مدل مخفی مارکوف تکاملی ارائه شد؛ لذا پس از بیان چالش‌های موجود در زمینه سامانه‌های تشخیص تهاجم، مفاهیم و اصطلاحات مطرح در امینت رایانه، انواع حملات مطرح در شبکه، نحوه ظهور سامانه‌های تشخیص نفوذ و روند تکاملی این سامانه‌ها در قالب سه نسل، شامل سامانه‌های مبتنی بر میزبان، سامانه‌های مبتنی بر شبکه و سامانه‌های ترکیبی (ترکیب میزبان و شبکه) موردبررسی قرار گرفت؛ همچنین همه پیشرفت‌های دو دهه اخیر، در زمینه سامانه‌های تشخیص نفوذ مبتنی بر مدل مخفی مارکوف، موردبررسی و ارزیابی قرار گرفته و مزایا و معایب هرکدام از آن‌ها بیان شد و به این نتیجه رسیدیم که نکته بسیار کلیدی در سامانه‌های تشخیص تهاجم مبتنی بر HMM، کارایی در بخش آموزش مدل است، که باید به آن توجه ویژه داشت. با توجه به کمبود دادهٔ آموزشی در زمینه تشخیص نفوذ و همچنین لزوم تشخیص هرچه بیشتر حملاتی که هر روز بر تعدادشان افزوده می‌شود، طراحی یک سامانهٔ تشخیص نفوذ مطمئن که بتواند با تعداد رکورد آموزشی کم، تعلیم داده شود، بیش‌ازپیش موردنیاز است؛ لذا در همین راستا در این پژوهش، یک سامانهٔ تشخیص تهاجم مبتنی بر مدل مخفی مارکوف که با استفاده از الگوریتم‌های تکاملی آموزش داده‌شده، طراحی شده است. با توجه به گسترش حملات در دنیای فناوری امروز و با توجه به عدم پوشش تمام حملات در مجموعه داده‌های موجود، لزوم ساخت یک مجموعه‌دادهٔ استاندارد که در آن الگوهای مختلف حملات قرار داشته، یکی از پروژه‌های مطرح در این زمینه است.

[DOR: 20.1001.1.24763047.1398.8.2.1.9]

یکی از معایب همه سامانه‌های تشخیص نفوذ مبتنی بر یادگیری ماشین، تولید نرخ هشدار غلط زیاد آن است؛ بنابراین یکی از پیشنهادهای مطرح در این زمینه کاهش این نرخ با استفاده از الگوریتم‌های داده‌کاوی است.

یکی از اهداف به‌وجودآمدن بخش تشخیص نوع حمله در سامانه پیشنهادی EHMM، استفاده هوشمندانه از بازخوردهای حاصل از تشخیص حملات با توجه به زمان و بهبود کشف و دسته‌بندی حملات بود؛ لذا می‌توان با بهره‌گیری از این ویژگی، با استفاده از روش‌های داده‌کاوی تحلیل دقیق‌تری در برخورد با حملات نشان داد.

علاوه‌براین، سامانه پیشنهادی ارائه‌شده، به‌نوعی استفاده از الگوریتم‌های تکاملی به‌جای الگوریتم‌های جستجوی محلی را مدنظر قرار داده، اما راه‌حل بهتر آن است که بتوان به‌گونه‌ای، از هردوی آن‌ها در کنار هم استفاده کرد؛ که هم بتواند درجاهایی که با مسأله تعداد رکوردهای آموزشی کم یا زیاد مواجهه هستیم، کارایی لازم را داشته باشد. به‌یقین با این نگرش می‌توان قابلیت‌های سامانهٔ موردنظر را افزایش داد.

۷- مراجع

[1] D. E. Denning, "An intrusion-detection model," *IEEE Transactions on software engineering*, pp. 222-232, 1987.

[2] R. G. Bace, "Survey Intrusion Detection," in Macmillan Technical Publishing, ed. USA, 2000.

[3] A. A. Ghorbani, W. Lu, and M. Tavallaee, Network intrusion detection and prevention: concepts and techniques vol. 47: Springer Science & Business Media, 2009.

[4] A. Lazarevic, V. Kumar, and J. Srivastava, "Intrusion detection: A survey," in Managing Cyber Threats, ed: Springer, 2005, pp. 19-78.

[5] E. Hodo, X. Bellekens, A. Hamilton, C. Tachtatzis, and R. Atkinson, "Shallow and Deep Networks Intrusion Detection System: A Taxonomy and Survey," arXiv preprint arXiv: 1701.02145, 2017.

[6] E. Figielska and W. Kasprzak, "An evolutionary programming based algorithm for HMM training," ed: in series: Challenging problems of Science–Computer Science, L. Bolc (series editor), Academic Publishing House EXIT, 2008, pp. 166-175.

[7] A. A. Thawerani and S. Ghani, "Evolving HMM for ranking Twitter influence," *in Information and Communication Technologies (ICICT)*, 2015 International Conference on ,2015,pp. 1-7.

[8]R. J. Elliott, L. Aggoun, and J. B. Moore, Hidden Markov models: estimation and control vol. 29: Springer Science & Business Media, 2008.

[9] G. B. Fogel, "Evolutionary programming," Handbook of Natural Computing, pp. 699-708, 2012.

[10] H. Farhadi, M. AmirHaeri, and M. Khansari, "Alert correlation and prediction using data mining and HMM," *The ISC International Journal of Information Security*, vol. 3, 2015.

[11] L. Nguyen, "Tutorial on Hidden Markov Model," *Applied and Computational Mathematics*, vol. 6, pp. 16-38, 2017.

[12] U. S. K. P. M. Thanthrige, "Hidden Markov Model Based Intrusion Alert Prediction," The University of Western Ontario, 2016.

[13] J. Choy and S.-B. Cho, "Anomaly detection of computer usage using artificial intelligence techniques," *in Pacific Rim International Conference on Artificial Intelligence*, 2000, pp. 31-43.

[14] X. Hoang and J. Hu, "An efficient hidden Markov model training scheme for anomaly intrusion detection of server applications based on system calls," in Networks, 2004,(ICon 2004), *Proceedings 12th IEEE International Conference on*, 2004, pp. 470-474.

[15] D. Ourston, S. Matzner, W. Stump, and B. Hopkins, "Applications of hidden markov models to detecting multi-stage network attacks," in System Sciences, 2003, *Proceedings of the 36th Annual Hawaii International Conference on,* 2003, pp. 10

[16] D. Gao, M. K. Reiter, and D. Song, "Beyond output voting: Detecting compromised replicas using HMM-based behavioral distance," *IEEE Transactions on Dependable and Secure Computing*, vol. 6, pp. 96-110, 2009.

usiness Media, 2008.

[17] K. Haslum, M. E. Moe, and S. J. Knapskog, "Real-time intrusion prevention and security analysis of networks using HMMs," in Local Computer Networks, 2008. LCN 2008. *33rd IEEE Conference on*, 2008, pp. 927-934.

[18] U. Tosun, "Policy misuse detection in communication networks with hidden Markov models," *Proceeding Computer Science*, vol. 32, pp. 947-952, 2014

[DOR: 20.1001.1.24763047.1398.8.2.1.9]

[19] D. Kar, K. Agarwal, A. K. Sahoo, and S. Panigrahi, "Detection of SQL injection attacks using Hidden Markov Model," in Engineering and Technology (ICETECH), *2016 IEEE International Conference on*, 2016, pp. 1-6.

[20] Zegeye, Wondimu K., Farzad Moazzami, and Richard Dean. "Hidden Markov Model (HMM) based Intrusion Detection System (IDS)." (2018).

محمد درویشی درجه کارشناسی ارشد خود را در رشته فناوری اطلاعات گرایش شبکه رایانه‌ای از دانشگاه جامع امام حسین (ع) اخذ کرد. علایق پژوهشی ایشان شبکه و امنیت شبکه است.

مجید غیوری ثالث درجه دکترای خود را از دانشگاه علم و صنعت اخذ کرد. ایشان در حال حاضر دانشیار دانشکده فناوری اطلاعات و ارتباطات دانشگاه جامع امام حسین (ع) هستند. سامانه‌های تشخیص نفوذ در شبکه‌های رایانه‌ای، سامانه‌های تشخیص نفوذ در پایگاه داده، مدل‌سازی امنیت با استفاده از داده‌کاوی، بررسی مناقشات سایبری با استفاده از نظریه بازی‌ها از جمله زمینه‌های پژوهشی ایشان است.

[DOR: 20.1001.1.24763047.1398.8.2.1.9]