

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ
بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ
بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



معاونت پژوهش و فن آوری

به نام خدا

شکوه اخلاق پژوهش

بیایید از خداوند سبحان و اعتماد به این که عالم محضر خداست و همواره ناظر بر اعمال انسان و به منظور پاس داشت تمام بلند انش و پژوهش و نظریه ایست جایگاه دانشگاه و اعتمادی فرهنگ و تمدن بشری، ما دانشمندان و اصناف بیست علمی و اخلاقی دانشگاه آزاد اسلامی مستعدی گردیم اصول زیر را در انجام تعالیات پای پژوهش به نظر قرار داده و از آن تعالی کتیم:-

- ۱- اصل حقیقت جویی: تلاش در راستای پی جویی حقیقت و وفاداری به آن و دوری از حرکت چنان سازی حقیقت.
- ۲- اصل رعایت حقوق: التزام به رعایت کامل حقوق پژوهشگران و پژوهشگران (انسان، حیوان و نبات) و سایر مساجد حق.
- ۳- اصل یکیت لای و مسنوی: تمهید به رعایت کامل حقوق لای و مسنوی دانشگاه و کج دیگران پژوهش.
- ۴- اصل منافع ملی: تمهید به رعایت مصالح ملی و در نظر داشتن میسر و توسعه کشور و کج دیگر ممال پژوهش.
- ۵- اصل رعایت انصاف و انانیت: تمهید به اجتناب از حرکت جانب داری غیر علمی و مسائمت از اموال، تجسرات و منافع در اختیار.
- ۶- اصل رازداری: تمهید به بیانت از اسرار و اطلاعات محرمانه افراد، سازمان ها و کشور و کج افراد و نهادی مرتبط با تحقیق.
- ۷- اصل احترام: تمهید به رعایت حریم ااد و حرمت ااد انجام تحقیقات و رعایت جانب تعد و خودداری از حرکت حرمت شکنی.
- ۸- اصل ترویج: تمهید به رواج دانش و ابداع نتایج تحقیقات و انتقال آن به بچه ها و علمی و دانشمندان به غیر از مواردی که منع قانونی دارد.
- ۹- اصل برت: التزام به برت جویی از حرکت زرقه غیر زردی و اعلام موضع نسبت به کسانی که حوزه علم و پژوهش را به شایه های غیر علمی آلوده.



دانشگاه آزاد اسلامی
واحد علوم تحقیقات

تعهدنامه اصالت رساله یا پایان نامه

اینجانب **مریم بهنام** دانش آموخته مقطع کارشناسی ارشد ناپیوسته در رشته نرم افزار کامپیوتر که در تاریخ ۹۲/۸/۲۸ از پایان نامه / رساله خود تحت عنوان " استفاده از تکنیکهای داده کاوی در سیستمهای تشخیص نفوذ " با کسب نمره ۱۹ و درجه عالی دفاع نموده ام بدینوسیله متعهد می شوم:

- ۱) این پایان نامه / رساله حاصل تحقیق و پژوهش انجام شده توسط اینجانب بوده و در مواردی که از دستاوردهای علمی و پژوهشی دیگران اعم از (پایان نامه، کتاب، مقاله و ...) استفاده نموده ام، مطابق ضوابط و رویه موجود، نام منبع مورد استفاده و سایر مشخصات آن را در فهرست مربوطه ذکر و درج کرده ام.
- ۲) این پایان نامه / رساله قبلاً "برای دریافت هیچ مدرک تحصیلی (هم سطح، پائین تر یا بالاتر) در سایر دانشگاه ها و مؤسسات آموزش عالی ارائه نشده است.
- ۳) چنانچه بعد از فراغت از تحصیل، قصد استفاده و هرگونه بهره برداری اعم از چاپ کتاب، ثبت اختراع و ... از این پایان نامه داشته باشم، از حوزه معاونت پژوهشی واحد مجوزهای مربوطه را اخذ نمایم.
- ۴) چنانچه در هر مقطع زمانی خلاف موارد فوق ثابت شود، عواقب ناشی از آن را می پذیرم و واحد دانشگاهی مجاز است با اینجانب مطابق ضوابط و مقررات رفتار نموده و در صورت ابطال مدرک تحصیلی ام هیچ گونه ادعایی نخواهم داشت.

نام و نام خانوادگی: **مریم بهنام**
تاریخ و امضاء: ۹۲/۸/۲۸



دانشگاه آزاد اسلامی

واحد علوم و تحقیقات شاهرود

دانشکده فنی مهندسی، گروه کامپیوتر

پایان نامه برای دریافت درجه کارشناسی ارشد در رشته مهندسی کامپیوتر (M.Sc)

گرایش: نرم افزار

عنوان:

استفاده از تکنیک های داده کاوی در سیستم های تشخیص نفوذ

استاد راهنما:

دکتر سید هاشم طبسی

استاد مشاور:

دکتر محمود معلم

نگارش:

مریم بهنام

پاییز ۱۳۹۲

سپاسگزاری

خداوند متعال را شاکرم که این توفیق را نصیب من نمود تا بتوانم این پایان نامه را به پایان برسانم .

برخود لازم می‌دانم از اساتید ارجمند جناب آقای دکتر طبسی به عنوان استاد راهنما و جناب آقای دکتر معلم به عنوان استاد مشاور که با راهنمایی های خود باعث فراهم آوردن فرصتی جهت تحقیق در این زمینه شدند و با صبر و حوصله بسیار مرا در این مسیر هدایت فرمودند ، کمال تشکر را داشته باشم .

از استاد ارجمند جناب آقای دکتر پور قلی که زحمت مطالعه و داوری این پایان نامه را بر عهده داشته اند، تشکر و قدردانی می‌نمایم .

همچنین از خانواده عزیزم که همیشه حامی و مشوق من بوده اند و در راه پیشرفتم از هیچ تلاشی فرور گذار نکرده اند سپاسگزاری می‌نمایم. در نهایت از دوستان خوبم که همیشه یار و یاورم بوده اند سپاسگزارم .

مریم بهنام – آبان ۱۳۹۲

تقدیم به پدر و مادر عزیزم
که زحماتشان جبران نشدنی است

فهرست مطالب

شماره صفحه

عنوان

فصل اول : کلیات

۱	چکیده فارسی
۱-۱	مقدمه
۲-۱	مروری بر پیشینه تحقیق
۳-۱	هدف و موضوع تحقیق
۴-۱	نتایج حاصل شده
۵-۱	مروری بر فصول پایان نامه

فصل دوم : سیستم های تشخیص نفوذ

۱-۲	مقدمه
۲-۲	دلایل استفاده از سیستم تشخیص نفوذ
۳-۲	نگاهی بر سیستمهای تشخیص نفوذ (IDS)
۴-۲	انواع روشهای تشخیص نفوذ
۱-۴-۲	روش تشخیص رفتار غیر عادی
۲-۴-۲	روش تشخیص سوءاستفاده یا تشخیص مبتنی بر امضاء
۵-۲	انواع معماری سیستم های تشخیص نفوذ
۱-۵-۲	سیستم تشخیص نفوذ مبتنی بر میزبان (HIDS)
۲-۵-۲	سیستم تشخیص نفوذ مبتنی بر شبکه (NIDS)
۳-۵-۲	سیستم تشخیص نفوذ توزیع شده (DIDS)
۶-۲	روش های برخورد و پاسخ به نفوذ
۱-۶-۲	پاسخ غیرفعال در سیستم تشخیص نفوذ
۲-۶-۲	پاسخ فعال در سیستم تشخیص نفوذ

فصل سوم : مفاهیم داده کاوی

۱-۳	مقدمه
۲-۳	تعریف تئوری از داده کاوی
۳-۳	تکنیک های داده کاوی
۴-۳	چرخه پروژه داده کاوی
۱-۴-۳	مرحله اول: جمع آوری داده
۲-۴-۳	مرحله دوم: پاک سازی و تبدیل داده
۱-۲-۴-۳	تبدیل نوع داده
۲-۲-۴-۳	تبدیل ستون پیوسته
۳-۲-۴-۳	گروه بندی
۴-۲-۴-۳	از بین بردن داده های پرت

۱۵ ۳-۴-۳- مرحله سوم: ساخت مدل
۱۵ ۳-۴-۴- مرحله چهارم : ارزیابی مدل
۱۵ ۳-۴-۵- مرحله پنجم : گزارش گیری
۱۵ ۳-۴-۶- مرحله ششم : مدیریت مدل
۱۵ ۳-۵- پایه های یک فرآیند داده کاوی
۱۶ ۳-۶- داده کاوی در سیستم های تشخیص نفوذ
۱۶ ۳-۷- وظایف داده کاوی
۱۶ ۳-۸- طبقه بندی و پیش بینی داده ها
۱۷ ۳-۸-۱- مراحل یک الگوریتم طبقه بندی
۱۷ ۳-۹- خوشه بندی
۱۸ ۳-۹-۱- فرآیند خوشه بندی :
۱۸ ۳-۹-۲- کیفیت خوشه بندی
۱۸ ۳-۱۰- انواع کاربردهای داده کاوی
۱۹ ۳-۱۱- نتیجه گیری

فصل چهارم : مفاهیم و الگوریتمهای یادگیری ماشین

۲۰ ۴-۱- مقدمه
۲۰ ۴-۲- نگاهی به یادگیری ماشین
۲۱ ۴-۳- یادگیری ماشین
۲۱ ۴-۳-۱- یادگیری نظارت شونده
۲۲ ۴-۳-۲- یادگیری غیر نظارت شونده
۲۲ ۴-۴- مسائل اساسی یادگیری
۲۲ ۴-۵- روند عملیات
۲۳ ۴-۶- برخی تکنیک های یادگیری ماشین
۲۳ ۴-۷- یادگیری با استفاده از الگوریتم های ژنتیک
۲۴ ۴-۸- اجزای الگوریتم های تکاملی
۲۴ ۴-۹- الگوریتم ژنتیک چیست؟
۲۶ ۴-۱۰- چارچوب کلی الگوریتم ژنتیک
۲۷ ۴-۱۱- ارائه ساختار مناسب برای هر فرد
۲۷ ۴-۱۲- عملگرهای ژنتیک
۲۷ ۴-۱۲-۱- تابع شایستگی (Fitness Function)
۲۸ ۴-۱۲-۲- انتخاب (Select)
۲۸ ۴-۱۲-۳- تلفیق (Crossover)
۳۰ ۴-۱۲-۴- جهش (Mutation)
۳۱ ۴-۱۲-۵- جایگزینی (Reinsertion)
۳۱ ۴-۱۲-۵-۱- جایگزینی سراسری

۳۱ جایگزینی محلی - ۲-۵-۱۲-۴
۳۱ شرط پایان الگوریتم - ۶-۱۲-۴
۳۱ کاربرد های الگوریتم ژنتیک - ۱۳-۴
۳۲ مزایا و معایب - ۱۴-۴
۳۳ نتیجه - ۱۵-۴
فصل پنجم : تجزیه و تحلیل داده ها و نتیجه گیری	
۳۴ ۱-۵- مقدمه
۳۴ ۲-۵- برنامه نویسی ژنتیک
۳۵ ۳-۵- برنامه نویسی بیان ژن
۳۶ ۴-۵- فرآیند برنامه نویسی بیان ژن
۳۷ ۵-۵- اجزای اصلی برنامه نویسی بیان ژن
۴۲ ۱-۵-۵- بررسی ساختار ژن ها
۴۶ ۶-۵- مجموعه داده های تشخیص نفوذ KDDCUP99
۴۸ ۷-۵- تجزیه تحلیل داده ها
۵۴ ۸-۵- تنظیمات پارامتر ها در روش GEP اصلاح شده
۵۵ ۹-۵- نتیجه گیری
۵۵ ۱۰-۵- پیشنهادات
۶۲ فهرست منابع فارسی
۶۳ فهرست منابع انگلیسی
۶۵ واژنامه انگلیسی به فارسی
۶۷ چکیده انگلیسی

فهرست شکل ها

عنوان	شماره صفحه
(شکل ۱-۲) مثالی از سیستم های تشخیص نفوذ مبتنی بر میزبان	۹
(شکل ۲-۲) مثالی از سیستم های تشخیص نفوذ مبتنی بر شبکه	۱۰
(شکل ۳-۲) مثالی از يك سیستم تشخیص نفوذ توزیع شده
(شکل ۱-۳) مراحل فرآیند داده کاوی	۱۶
(شکل ۲-۳) نمونه ای از خوشه بندی	۱۸
(شکل ۱-۴) نمونه ای از یک فرایند یادگیری نظارت شونده	۲۱
(شکل ۲-۴) روند پردازش از داده های یادگیری تا آزمون	۲۲
(شکل ۳-۴) عملکرد کلی الگوریتم ژنتیک	۲۶
(شکل ۴-۴) تلفیق تک نقطه	۲۹
(شکل ۵-۴) تک نقطه ای (بالا) و چند نقطه ای (پایین)	۲۹
(شکل ۶-۴) نمونه ای از تغییر ۱ به ۰ در کروموزوم فرزند نسبت به والد پس از عمل جهش	۳۰
(شکل ۱-۵) طرح کلی گامهای مقدماتی برنامه نویسی بیان ژن	۳۵
(شکل ۲-۵) فلوچارت از گام های اجرایی برنامه نویسی بیان ژن	۴۱
(شکل ۳-۵) پیمایش درختی با رشته ورودی $Q*c-abde/+$	۴۲
(شکل ۴-۵) پیمایش درختی با رشته ورودی $/aQ/b*ab/Qa*b*-ababaababbabbba$	۴۳
(شکل ۵-۵) پیمایش درختی با رشته ورودی $/a+/b*ab/Qa*b*-ababaababbabbba$	۴۴
(شکل ۶-۵) پیمایش درختی با رشته ورودی $/aQ/bbab/Qa*b*-ababaababbabbba$	۴۵
(شکل ۷-۵) پیمایش درختی با رشته ورودی $/aQ/bbab/Qa*b*-ababaababbabbba$	۴۵
(شکل ۸-۵) پیمایش درختی حاصل از رشته $+Q-/b*aaQbaabaabbaaab$	۴۶
(شکل ۹-۵) طرح سلسه مراتبی از دسته بندی حملات	۵۶

فهرست جداول

<u>شماره صفحه</u>	<u>عنوان</u>
۲۵	(جدول ۱-۴) شبه کد الگوریتم ژنتیک.....
۴۰	(جدول ۱-۵) شبه کد الگوریتم GEP.....
۴۷	(جدول ۲-۵) تعداد حملات در مجموعه داده 10-percent و Corrected.....
۴۹	(جدول ۳-۵) ۴۱ ویژگی مربوط به تشخیص نفوذ.....
۵۱	(جدول ۴-۵) متغیرهای اصلاح شده و ۱۹ متغیر باقی مانده.....
۵۲	(جدول ۵-۵) تعداد و نحوه نمایش عملگرهای الگوریتم.....
۵۳	(جدول ۶-۵) حذف ۱۶ ویژگی های با بررسی معیار اول.....
۵۳	(جدول ۷-۵) حذف ۶ ویژگی های با بررسی معیار دوم.....
۵۷	(جدول ۸-۵) مقادیر پارامترهای روش GEP اصلاح شده.....
۵۷	(جدول ۹-۵) مقایسه درصد صحت نرخ آلام در GEP و GEP اصلاح شده.....
۵۷	(جدول ۱۰-۵) مقادیر TP , FP برای Dataset_Test.....
۵۷	(جدول ۱۱-۵) مقایسه دقت طبقه بندی بین دو الگوریتم.....

فهرست نمودارها

<u>شماره صفحه</u>	<u>عنوان</u>
۵۸	Error vs Iteration, Normal_Train (نمودار ۱-۵)
۵۸	Error vs Iteration, Dos_Train (نمودار ۲-۵)
۵۹	Error vs Iteration, Probe_Train (نمودار ۳-۵)
۵۹	Error vs Iteration, U2R,R2l_Train (نمودار ۴-۵)
۶۰	Error vs Iteration, Normal_Test (نمودار ۵-۵)
۶۰	Error vs Iteration, Dos_Test (نمودار ۶-۵)
۶۱	Error vs Iteration, Probe_Test (نمودار ۷-۵)
۶۱	Error vs Iteration, U2R,R2l_Test (نمودار ۸-۵)

چکیده فارسی:

فرآیند تشخیص نفوذ نظارت بر رویدادهایی است که در سیستم های کامپیوتری و شبکه اتفاق می افتد و آنها را برای شناسایی نفوذ تجزیه و تحلیل می کند. چگونگی تشخیص حملات شبکه بخش مهمی در سیستم های تشخیص نفوذ است.

در این پایان نامه سیر تکامل قوانین بر اساس برنامه نویسی ژنتیک برای تشخیص نفوذ در شبکه معرفی شده است که الگوهای حملات شناخته شده را تشخیص می دهد و با ارتقا روش جدید به نام برنامه نویسی بیان ژن (Gene Expression Programming) از ژنوم خطی به همراه عملگرهای ژنتیکی مانند جهش، ترکیب، و ارونسازي و جابجایی برای این کار استفاده می کند. در نهایت با شاخص های دقت طبقه بندی، نرخ مثبت صحیح (TPR) و نرخ مثبت کاذب (FPR) بازدهی سیستم تشخیص نفوذ پیشنهادی نسبت به سیستم های مرسوم قبلی نشان داده شده است. نتایج نشان می دهد تکنیک های برنامه نویسی ژنتیکی استفاده شده برای تشخیص نفوذ از روشهای مشابه خود که بر اساس الگوهای یادگیری ماشین هستند از دقت بالاتری برخوردار است. از مجموعه داده به نام KDDCUP99 که انواع حملات شبکه را در خود جای داده است استفاده می نمایم.

کلمات کلیدی

سیستم تشخیص نفوذ، یادگیری ماشین، برنامه نویسی ژنتیک، برنامه نویسی بیان ژن

فصل اول

۱-۱- مقدمه

همگام با رشد شبکه های کامپیوتری حملات و نفوذ ها به این شبکه گسترش یافته و به روشهای متعددی انجام می شود. نفوذ مجموعه اقدامات غیر قانونی است که صحت، محرمانگی و یا دسترسی به یک منبع را به خطر می اندازد. نفوذ گران را می توان به دو دسته نفوذگران داخلی و نفوذگران خارجی دسته بندی کرد.

نفوذگران خارجی کسانی هستند که اجازه استفاده از سیستم را ندارند اما سعی می کنند سیستم را مورد دسترسی قرار دهند و نفوذ گران داخلی کسانی هستند که به سیستم اختیارات محدودی دارند اما سعی می کنند به منابعی که اجازه دسترسی به آن را ندارند، دسترسی پیدا کنند.

به منظور مقابله با نفوذ گران به شبکه ها و سیستم های کامپیوتری، روشهای متعددی تعیین شده است که روشهای تشخیص نفوذ نامیده میشوند. هدف از تشخیص نفوذ این است که استفاده غیر مجاز، سوء استفاده و آسیب رساندن به سیستم ها و شبکه های کامپیوتری توسط هر دو دسته کاربران داخلی و حمله کنندگان خارجی شناسایی شود.

بطور کلی روشهای تشخیص نفوذ به دو دسته اصلی تشخیص سوء استفاده و تشخیص رفتار غیر عادی تقسیم میشوند. در روش تشخیص سوء استفاده از الگوهای نفوذ شناخته شده برای شناسایی نفوذ ها استفاده می شود. در حالی که روشهای تشخیص رفتار غیر عادی، رفتار عادی کاربران ملاک عمل قرار داده میشود و در نتیجه هرگونه مغایر با آن به عنوان تلاش جهت نفوذ به سیستم شناسایی می گردد.

به منظور پیاده سازی روشهای تشخیص نفوذ، سیستم های متعددی طراحی و ساخته شده اند. در حوزه امنیت کامپیوتر، سیستمهای تشخیص نفوذ نقش هشدار دهنده را ایفا می کنند و هر زمان که امنیت شبکه در معرض خطر قرار می گیرد، آن را اعلام می کنند. نهاد دیگری که مسئول امنیت سایت نامیده می شود، می تواند به این هشدار ها پاسخ داده و راهکارهای مناسب را اجرا کند.

سیستم های تشخیص نفوذ اولیه با تحلیل فایل رخدادهای که توسط سیستم عامل و برنامه های کاربردی ایجاد میشوند، کار می کردند. اما به مرور که شبکه ها پیچیده شدند به داده های کافی برای شناسایی قاطعانه یک حمله دسترسی نداشتند. بنابراین با توجه به روشهای تشخیص نفوذ پیچیده تر مبتنی بر تحلیل داده های شبکه یا میزبان معطوف شد.

نفوذگرها معمولاً از عیوب سخت افزاری، شکستن کلمات رمز، استراق سمع ترافیک شبکه و نقاط ضعف طراحی برای نفوذ به سیستم و شبکه های کامپیوتری استفاده می کنند.

اغلب لازم است با کشف هر حمله جدید تشخیص نفوذ بروز رسانی شود. سازندگان این سیستم ابتدا سناریوهای حمله و نقاط ضعف سیستم را تجزیه تحلیل و طبقه بندی می کنند. به دلیل ماهیت دستی و غیر الگوریتمی فرآیند توسعه در سیستمهای تشخیص نفوذ، اعمال تغییرات در آن کند و پرهزینه است.

۱-۲- مروری بر پیشینه تحقیق

(عظیمیان و دیگران ۱۳۸۹) از تکنیکهای داده کاوی در زمینه یادگیری ماشین و از الگوریتمهای ماشین بردار پشتیبان (SVM)^۱، الگوریتم درخت تصمیم (C4.5) استفاده کردند و از روش تشخیص سوء استفاده، به منظور تشخیص نفوذ در شبکه های کامپیوتری مقایسه و ارزیابی شده است. نتایج حاصل از آزمایشات از بین معیار های ارزیابی آنها نشان می دهد که در حالت میانگین، C4.5 در معیار های Detection rate و Precision بهتر عمل نموده است، اما در معیاری False alarm rate، SVM عملکرد بهتری داشته است. همچنین با توجه به معیار Accuracy عملکرد C4.5 کمی بهتر از SVM می باشد.

(معدنی پور، ابوالحسني و شیرازي ۱۳۸۹) روشی از مفاهیم داده کاوی و تحلیل شبکه های اجتماعی پیشنهاد کرده اند. روشهای موجود برای انتخاب ویژگی ذکر شدند که هر کدام مزایا و معایبی دارند. تنها برخی از آنها قادر به رتبه بندی ویژگیها از نظر اهمیت هستند و هیچ یک قادر به بیان نحوه ارتباط ویژگی های تشخیص نفوذ نمی باشند. با توجه به نتایج نشان داده شده که روش (Particle Swarm optimization) می تواند مجموعه بهینه مرتبط با هر نوع حمله را بیان کند که هزینه اجرای آن نمایی است.

(علی زاده، مقدادی ۲۰۱۲) به مطالعه ی انواع حملات و سیستم های تشخیص ناهنجاری و سیستم های تشخیص سوء استفاده تاکید دارند و به معرفی انواع تکنیک های مورد استفاده در سیستم های تشخیص نفوذ می پردازند. برخی از این تکنیک ها مبتنی بر محاسبات هستند نظیر منطق فازی و شبکه های بیزین، برخی مبتنی بر هوش مصنوعی هستند نظیر سیستم های خبره و شبکه های عصبی و برخی دیگر مفاهیم بیولوژیکی نظیر سیستم های ژنتیکی و ایمنی می باشند.

(سلطانی زاده، برادران شکوهی و زارع ۱۳۹۰) الگوریتم جدیدی جهت دسته بندی داده ها بر مبنای روش ایمنی مصنوعی و تلفیق آن با روش تپه نوردی (Hill Climbing) به منظور پیاده سازی سیستم تشخیص نفوذ ارائه نمودند. داده های مورد تست و آزمون این الگوریتم NSL-KDD می باشد که بر روی ۷۴۵۳ داده TCP در مجموعه داده تست گردیده است.

(ابطحی، میبیدی ۱۳۸۷) یک مدل سلسله مراتبی از عامل های مبتنی بر اتوماتای یادگیر برای تشخیص نفوذهای از نوع انکار سرویس مطرح نمودند. در مدل پیشنهادی هر عامل دارای یک اتوماتای یادگیر است که پارامترهای مدل را نگهداری کرده و آنها را با توجه به بازخوردهایی که از محیط دریافت می کنند به روز می رسانند. در مدل ارائه شده مسئله تشخیص نفوذ به صورت یک مسئله دسته بندی دوکلاسه مطرح می گردد. برای آموزش و آزمایش این مدل، از مجموعه داده های KDD99 استفاده شده که یک مجموعه داده استاندارد برای کاربردهای امنیتی می باشد.

(زندگی، فتحی و شرایی ۱۳۹۰) با تعریف نفوذ، و داده کاوی به بررسی کاربردهای داده کاوی در تشخیص نفوذ پرداخته اند. سیستم های تشخیص نفوذی که بر اساس داده کاوی طراحی شده اند در شبکه های محلی دارای عملکرد بسیار خوبی هستند. در سیستم های تجاری از الگویی و امضای حملات شناخته شده استفاده می گردد. به اینگونه امضاها، امضاهای ایستا گویند، این بدان معناست که امضا و الگویی هر حمله پس از مشاهده رفتار آن حمله و ثبت داده های مربوط به آن با تحلیل هایی که بر روی داده ها صورت می گیرد تهیه می گردد، بدینوسیله سیستم در آینده به راحتی آن حمله را خواهد شناخت و جلوی

^۱Support vector machine

نفوذ و حمله به شبکه را خواهد گرفت. تحلیل بر روی داده ها جهت استخراج امضا توسط افراد خبره صورت می گرفت ولی جدیداً در این حوزه نیز از داده کاوی استفاده می گردد. (بلدرن، تالبوت و دبار ۲۰۰۶) هدف از ضمانت اطلاعات، محرمانه نگه داشتن اطلاعات و کنترل سطح دسترسی معرفی نمودند که اطلاعات خاص فقط توسط کاربران مجاز قابل دسترسی می باشد و کاربرد سیستم امنیتی تشخیص و جلوگیری از وقایع نامطلوب می باشد. بطور کلی سیستم نفوذ شامل ۳ عملکرد (نظارت، کشف، واکنش) است. تشخیص وقایع در امنیت شبکه توسط روشی از داده کاوی با عنوان یادگیری ماشین صورت گرفته است.

(هوگ، موکیت و بیکاس ۲۰۱۲) روشهای انتخاب معیار های ارزیابی کارایی مورد بررسی قرار دادند. برای ارزیابی و مقایسه کارایی الگوریتم ها از دومعیار استفاده شده است. برای برچسب گذاری از روش مبتنی بر فاصله استفاده شده که در این روش خوشه هایی که از سایر خوشه ها جدا افتاده اند و فاصله آنها از سایر خوشه ها زیاد است به عنوان خوشه های غیر نرمال و یا حمله تلقی می شوند و خوشه های نزدیک به هم شامل داده های نرمال هستند. این معیار ها نرخ کشف (DR) و نرخ مثبت کاذب (FPR) هستند که پس از تشکیل ماتریس پراکنندگی برای خوشه ها قابل محاسبه هستند.

(مالوف ۲۰۰۶) در شبکه MITRE از روشهای داده کاوی برای بهبود فرآیند شبکه استفاده می کند و پیش زمینه ای در خصوص شبکه MITRE و ویژگی های اصلی شبکه شامل: انتخاب، تراکم، طبقه بندی و فرآیند رتبه بندی ارائه کرده است. هدف این مقاله تشخیص آنومالی و کاهش تعداد نرخ آلام کاذب که توسط سنسورهایی در شبکه ایجاد شده برای تشخیص هشدارهای غیر ضروری می باشد. و بدین منظور از ترکیب دو روش درخت تصمیم و دسته بندی K-MEANS استفاده شده است.

(هوگ، موکیت و بیکاس ۲۰۱۲) برای سیستم تشخیص نفوذ با استفاده از الگوریتم ژنتیک با کارایی شناسایی انواع مختلف نفوذ شبکه ارائه کردند، برای پیاده سازی و اندازه گیری عملکرد سیستم از استاندارد KDD99 استفاده شده است و برای اندازه گیری تابع برآزش کروموزوم از معادله انحراف با فاصله استفاده شده است. نرخ تشخیص و فرآیند تا حد زیادی بهبود داده شده و نرخ مثبت کاذب بسیار کاهش یافته است.

۱-۳- هدف و موضوع تحقیق

هدف از این پایان نامه آشنایی با روشهای تشخیص نفوذ، بررسی سیستمهای تشخیص نفوذ، بررسی روشهای داده کاوی در تشخیص نفوذ، و در نهایت ارائه روش جدید با رویکرد مبتنی بر الگوریتم ژنتیک بنام Gene Expression Programming در تشخیص نفوذ مورد استفاده قرار می گیرد.

۴-۱ - نتایج حاصل شده

در این پایان نامه با ارائه مدل جدیدی از الگوریتم GEP برای دسته بندی رکوردهای اطلاعاتی شبکه به منظور انجام فرآیند تشخیص نفوذ استفاده می شود و با تکیه بر مجموعه داده استاندارد KDDCUP99 به مطالعه انواع حملات می پردازیم.

۵-۱ - مروری بر فصول پایان نامه

این پایان نامه مشتمل بر پنج فصل است. در فصل دوم روش های استفاده شده در تشخیص نفوذ معرفی می شود. در فصل سوم مفاهیم داده کاوی بررسی شده است و سعی شده است تمامی موارد مربوط به داده کاوی بصورت خلاصه پوشش داده شود. در فصل چهارم مفاهیم و الگوریتم های یادگیری ماشین ذکر شده است و به طور کامل مفاهیم الگوریتم ژنتیک توضیح داده شده است. در فصل پنجم روشی بر مبنای الگوریتم برنامه نویسی بیان ژن (GEP) پیشنهاد شده است. همچنین در این فصل به طور کامل الگوریتم GEP توضیح داده شده است و گام های پیاده سازی تشریح شده است. در نهایت نتایج حاصل از اجرای الگوریتم بروی مجموعه حملات داده استاندارد KDDCUP99 نشان داده شده است.

سیستمی دارای امنیت است که بتوان نسبت به عملکرد آن اعتماد نمود. به منظور رسیدن به درجه ی بالای اعتماد معمولاً سیاستهای امنیتی در نظر گرفته می شود. این سیاست ها عملکرد بخش های مختلف سیستم را کنترل کرده و نیازمندی های لازم جهت اعمال نظارت را مشخص می کنند. تعریفی دقیق از سیستم کامپیوتری امن به این صورت است: سیستمی امن است که بتواند محرمانگی، جامعیت و در دسترس بودن را برای کاربران خود فراهم آورد. محرمانگی به این معنی است که اطلاعاتی فقط در اختیار افرادی قرار گیرد که مجوز دسترسی به آن داده ها را قبلاً از مدیر سیستم داشته باشند.

جامعیت بدین معنی است که اطلاعات در اثر حوادث و یا فعالیتهای خرابکارانه بدون کوچکترین تغییر باقی بمانند و در دسترس بودن نیز بدین معنی است که سیستم باید در مواقع نیاز فعال بوده و بدون هیچ مشکلی منابع و اطلاعات را در اختیار استفاده کنندگان قرار دهد. در جوامع امروز سرعت دسترسی به اطلاعات و پردازش سریع آنها یک نیاز بديهی است. این نیازمندی باعث شده است که اطلاعات بیشتری در سیستم های کامپیوتری ذخیره و بازیابی شوند و برای استفاده از این اطلاعات نیازمند به اشتراک گذاردن آنها هستیم. به همین دلیل شبکه های کامپیوتری به سرعت در حال رشد و تکامل هستند.

با افزایش دانش مربوط به چگونگی عملکرد شبکه ها و برنامه های کامپیوتری، نفوذکنندگان ضعف های سیستم را به خوبی شناسایی کرده و از آنها برای نفوذ به شبکه استفاده می کنند. در برخی از موارد این ضعف ها به گونه ای است که نفوذ کننده را قادر می سازد هر عمل دلخواه را در سیستم انجام دهد. نفوذگرها ممکن است از الگوهای خاص نفوذ استفاده کنند که ردیابی و شناسایی آنها بسیار مشکل است. همچنین یک نفوذگر ماهر ابتدا برخی حرکات فریبنده انجام می دهد و آنگاه در فرصت مناسب حمله اصلی را به اجرا در می آورد و در برخی از موارد نفوذگرها با پنهان کردن ردپای خود کار تشخیص نفوذ را بسیار مشکل می کنند. (مالوف ۲۰۰۶).

نفوذتوسط سه گروه صورت می گیرد: الف) حمله کنندگان که از طریق اینترنت دسترسی پیدا می کنند.
ب) کاربران مجازی که در تلاش برای کسب دسترسی بیشتری هستند، در حالیکه اجازه چنین کاری ندارند.

ج) کاربران مجازی که از دسترسی خود سوء استفاده می کنند.

سیستم تشخیص نفوذ ابزار سخت افزاری یا نرم افزاری است که این تحلیل را خودکار انجام می دهد (طاهری منفرد ۱۳۸۷).

برای ایجاد امنیت کامل در يك سیستم کامپیوتری، علاوه بر دیواره های آتش (Firewall) و دیگر تجهیزات جلوگیری از نفوذ، سیستمهای دیگری به نام سیستم های تشخیص نفوذ (IDS)^۱ مورد نیاز می باشند تا بتوانند در صورتی که نفوذگر از دیواره ی آتش، آنتی ویروس و دیگر تجهیزات امنیتی عبور کرد و وارد سیستم شد، آن را تشخیص داده و چاره ای برای مقابله با آن بی اندیشند. سیستم های تشخیص نفوذ را می توان از

^۱Intrusion Detection System (IDS)

سه جنبه ي روش تشخیص، معماری و نحوه ي پاسخ به نفوذ طبقه بندی کرد. انواع مختلفی از معماری سیستمهای تشخیص نفوذ وجود دارد که به طور کلی می توان آن ها را در سه دسته ي مبتنی بر میزبان (HIDS)^۳، مبتنی بر شبکه (NIDS)^۴ و توزیع شده (DIDS)^۵ تقسیم بندی نمود.

۲-۲- دلایل استفاده از سیستم تشخیص نفوذ

دلایل متعددی را برای لزوم استفاده از IDS میتوان نام برد که اهم آنها در ادامه آمده است :

الف) تشخیص حمله و سایر فعالیتهای خصمانه برای نقض امنیت میزبان که دیگر مکانیزمهای امنیتی با آن مقابله نکرده اند.

ب) شناسایی و مقابله با فعالیتهای پیش نیاز برای حمله

ج) گزارش خطرات موجود به سازمان

د) کنترل کیفیت طرح امنیتی و مدیریتی، بخصوص برای سازمانهای بزرگ و پیچیده

ه) تامین اطلاعات مفید درباره ي نفوذهایی که رخ داده است. این کار باعث بهبود شناسایی، بازیابی و تصحیح دلایل نفوذ می شود.

۲-۳- نگاهی بر سیستمهای تشخیص نفوذ (IDS)

سیستمهای تشخیص نفوذ (IDS) وظیفه ي شناسایی و تشخیص هر گونه استفاده ي غیرمجاز به سیستم، سوء استفاده و یا آسیب رسانی توسط هر دو دسته ي کاربران داخلی و خارجی را بر عهده دارند. سیستم های تشخیص نفوذ به صورت سیستم های نرم افزاری و سخت افزاری ایجاد شده و هر کدام مزایا و معایب خاص خود را دارند. سرعت و دقت از مزایای سیستمهای سخت افزاری است و عدم شکست امنیتی آن ها توسط نفوذگران، قابلیت دیگر این گونه سیستم ها می باشد. اما استفاده ي آسان از نرم افزار، قابلیت انطباق پذیری در شرایط نرم افزاری و تفاوت سیستم های عامل مختلف، عمومیت بیشتری را به سیستمهای نرم افزاری می دهد و عموماً این گونه سیستم ها انتخاب مناسب تری هستند.

به طور کلی سه عملکرد اصلی IDS عبارتند از:

۱- ارزیابی و نظارت ۲- کشف ۳- واکنش

بر همین اساس هر IDS را می توان بر اساس روشهای تشخیص نفوذ، معماری و انواع پاسخ به نفوذ دسته بندی کرد.

۲-۴- انواع روشهای تشخیص نفوذ

نفوذ به مجموعه ي اقدامات غیرقانونی که صحت و محرمانگی و یا دسترسی به يك منبع را به خطر می اندازد، اطلاق می گردد. نفوذ ها می توانند به دو دسته ي داخلی و خارجی تقسیم شوند. نفوذهای خارجی به آن دسته نفوذهایی گفته می شود که توسط افراد مجاز و یا غیرمجاز از خارج شبکه به درون شبکه ي داخلی صورت می گیرد و نفوذهای داخلی توسط افراد مجاز در سیستم و شبکه ي داخلی، از درون خود شبکه انجام می پذیرد.

نفوذ گر ها عموماً از عیوب نرم افزاری، شکستن کلمات رمز، استراق سمع ترافیک شبکه و نقاط ضعف طراحی در شبکه، سرویس ها و یا کامپیوترهای شبکه برای نفوذ به سیستم ها و شبکه های کامپیوتری بهره

^۳ Host-based Intrusion Detection System (HIDS)

^۴ Network-based Intrusion Detection System (NIDS)

^۵ Distributed Intrusion Detection System (DIDS)

می‌برند به منظور مقابله با نفوذگران به سیستم‌ها و شبکه‌های کامپیوتری، روش‌های متعددی تحت عنوان روش‌های تشخیص نفوذ ایجاد گردیده است که عمل نظارت بر وقایع اتفاق افتاده در یک سیستم یا شبکه کامپیوتری را بر عهده دارد. روش‌های تشخیص مورد استفاده در سیستم‌های تشخیص نفوذ به دو دسته تقسیم می‌شوند:

۱- روش تشخیص رفتار غیر عادی

۲- روش تشخیص سوءاستفاده یا تشخیص مبتنی بر امضاء

۲-۴-۱- روش تشخیص رفتار غیر عادی

در این روش، یک نما از رفتار عادی ایجاد می‌شود. یک ناهنجاری ممکن است نشان دهنده یک نفوذ باشد. برای ایجاد نماهای رفتار عادی از رویکردهایی از قبیل شبکه‌های عصبی، تکنیک‌های یادگیری ماشین و ... استفاده می‌شود.

برای تشخیص رفتار غیر عادی، باید رفتارهای عادی را شناسایی کرده و الگوها و قواعد خاصی برای آن‌ها پیدا کرد. رفتارهایی که از این الگوها پیروی می‌کنند، عادی بوده و رویدادهایی که انحرافی بیش از حد معمول آماری از این الگوها دارند، به عنوان رفتار غیر عادی تشخیص داده می‌شود. نفوذهای غیر عادی برای تشخیص بسیار سخت هستند، چون هیچگونه الگوی ثابتی برای نظارت وجود ندارد. معمولاً رویدادی که بسیار بیشتر یا کمتر از دو استاندارد انحراف از آمار عادی به وقوع می‌پیوندد، غیر عادی فرض می‌شود. به عنوان مثال اگر کاربری به جای یک یا دو بار ورود و خروج عادی به سیستم در طول روز، بیست بار این کار را انجام دهد، و یا کامپیوتری که در ساعت ۲ بعد از نیمه شب مورد استفاده قرار گرفته در حالی که قرار نبوده کامپیوتر فوق پس از ساعت اداری روشن باشد. هر یک از این موارد می‌تواند به عنوان یک رفتار غیر عادی در نظر گرفته شود. یک معیاری که در تشخیص رفتار غیر عادی به کار می‌رود تشخیص سطح آستانه می‌باشد.

➤ تشخیص سطح آستانه

تعداد ورود/ خروج به سیستم و یا زمان استفاده از سیستم، از مشخصه‌های رفتار سیستم و یا استفاده کننده است که می‌توان با شمارش آن به رفتار غیر عادی سیستم پی برد و آن را ناشی از یک نفوذ دانست.

۲-۴-۲- روش تشخیص سوءاستفاده یا تشخیص مبتنی بر امضاء

در این تکنیک که معمولاً با نام تشخیص مبتنی بر امضاء شناخته شده است، الگوهای نفوذ از پیش ساخته شده به صورت قانون نگهداری می‌شوند. به طوری که هر الگو انواع متفاوتی از یک نفوذ خاص را در بر گرفته و در صورت بروز چنین الگویی در سیستم، وقوع نفوذ اعلام می‌شود. در این روش‌ها، معمولاً تشخیص دهنده دارای پایگاه داده‌ای از امضاءها یا الگوهای حمله است و سعی می‌کند با بررسی ترافیک شبکه، الگوهای مشابه با آن چه را که در پایگاه داده‌ی خود نگهداری می‌کند، بیابد. این دسته از روش‌ها تنها قادر به تشخیص نفوذهای شناخته شده می‌باشند و در صورت بروز حملات جدید در سطح شبکه، نمی‌توانند آن‌ها را شناسایی کنند و مدیر شبکه باید همواره الگوی حملات جدید را به سیستم تشخیص نفوذ اضافه کند. از مزایای این روش دقت در تشخیص نفوذهایی است که الگوی آن‌ها عیناً به سیستم داده شده است.