بررسی کاربرد الگوریتم ژنتیک در تشخیص نفوذ شبکه های کامپیوتری

زهرا سیوندیان[۱]، حمید رستگاری[۲]

[۱]دانشکده کامپیوتر دانشگاه آزاد اسلامی واحد نجف‌آباد؛ zsv۹۷@yahoo.com
[۲]دانشکده کامپیوتر دانشگاه آزاد اسلامی واحد نجف‌آباد؛ Rastegari@iaun.ac.ir

چکیده

شبکه‌ها و سیستم‌های اطلاعاتی در معرض حملات الکترونیکی هستند. وقتی حملات شبکه اتفاق می‌افتد سازمان‌ها به حالت بحرانی در می‌آیند. متدهای سنتی که برای غلبه بر این تهدیدها استفاده می‌شوند امنیت کاملی را برای سیستم فراهم نمی‌کنند. این امر محققان را برای توسعه‌ی یک سیستم تشخیص نفوذ که قادر به تشخیص و پاسخ به این چنین وقایعی باشد تشویق می‌کند. سیستم تشخیص نفوذ یکی از روش‌های عمده برای مسئله‌ی مهم و در حال گسترش امنیت کامپیوتر می‌باشد که عبارت است از تشخیص دسترسی غیرمجاز به یک سیستم کامپیوتری. معماری‌های مختلف و روش‌های مبتنی بر محاسبات نرم متعددی بویژه الگوریتم ژنتیک برای تشخیص نفوذهای شبکه استفاده شده‌اند. این مقاله‌ی مروری مطالعه‌ای جامع روی سیستم تشخیص نفوذ مبتنی بر الگوریتم ژنتیک ارائه می‌دهد، همچنین تکنیک‌های تشخیص نفوذ شامل الگوریتم ژنتیک در دهه‌ی اخیر را به طور مختصر بررسی می‌کند.

کلمات کلیدی

سیستم تشخیص نفوذ، نفوذ، محاسبات نرم، الگوریتم ژنتیک

۱- مقدمه‌ای بر IDS

در عصر اطلاعات شبکه‌های کامپیوتری و کاربردهای مربوطه بسیار محبوب شده‌اند اما تهدیدهای مختلفی به آنها وارد است. وقتی یک سیستم کامپیوتری به یک شبکه متصل می‌شود در معرض یک ریسک بالایی قرار می‌گیرد. از جمله تهدیدهایی که به یک سیستم کامپیوتری وجود دارند ویروس‌ها و نفوذها هستند. ویروس‌ها می‌توانند به طور گسترده با نصب نرم‌افزار آنتی‌ویروس و بروزرسانی بطور منظم کنترل شوند.

هر دسترسی غیرمجاز به منابع یک کامپیوتر، نفوذ[۱] به یک کامپیوتر گفته می‌شود. برای دفاع در مقابل این تهدیدها ، تکنیک‌های امنیتی بسیاری در دهه‌های گذشته مطالعه شده‌اند از قبیل رمزنگاری[۲]، دیواره آتش[۳]، تشخیص نفوذ و ناهنجاری[۴]. از میان آن‌ها تشخیص نفوذ در شبکه به عنوان محتمل‌ترین و امیدبخش‌ترین متد برای دفاع در مقابل رفتارهای نفوذ دینامیکی و پیچیده ملاحظه می‌شود[۳] ـ یک سیستم تشخیص نفوذ[۵]، فعالیت‌های یک محیط داده شده را مانیتور می‌کند و تصمیم می‌گیرد که این فعالیت‌ها بر اساس یکپارچگی[۶] ، قابلیت اعتماد[۷]، دسترس‌پذیری[۸] منابع اطلاعات، مخرب یا طبیعی هستند.

سیستم تشخیص نفوذ اطلاعاتی درباره‌ی سیستم مشاهده شده گردآوری می‌کند. این داده‌ی audit[۹] جمع‌آوری شده توسط detector[۱۰] پردازش می‌شود. detector، اطلاعات غیر لازم را از داده‌ی audit از بین می‌برد و سپس یک تصمیم برای ارزیابی احتمال اینکه این فعالیت‌ها می‌توانند به عنوان یک نشانه‌ی نفوذ ملاحظه شوند را در بر می‌گیرد[۲].

---

[۱] intrusion
[۲] cryptography
[۳] firewall
[۴] anomaly
[۵] Intrusion Detection System
[۶] integrity
[۷] confidentiality
[۸] availability

[۹] ممیزی
[۱۰] تشخیص دهنده

سیستم‌های تشخیص نفوذ سنتی محدود هستند و راه حلی کامل برای مسئله فراهم نمی‌کنند. آنها برای فعالیت‌های بد[۱۱] بالقوه روی ترافیک‌های شبکه جستجو می‌کنند و گاهی اوقات موفق می‌شوند تا حملات امنیتی و ناهنجاری‌ها را درست پیدا کنند. هرچند در بسیاری از موارد آنها شکست می‌خورند تا رفتارهای بد(false negative) را تشخیص دهند یا هشداری را می‌دهند زمانیکه هیچچیز اشتباهی در شبکه وجود ندارد(false positive) ـ علاوه بر این آنها پردازش دستی جامع و استنباط تخصصی انسان را نیاز دارند. بکار بردن تکنیک‌های داده‌کاوی روی داده‌ی ترافیک شبکه یک راه حل محتمل(امیدبخش) است که به توسعه‌ی سیستم‌های تشخیص نفوذ بهتر کمک می‌کند.

تکنیک‌های محاسبات نرم مختلفی شبیه الگوریتم ژنتیک، شبکه‌های عصبی مصنوعی، ماشین‌های بردار پشتیبان و منطق فازی برای ایجاد سیستم تشخیص نفوذ استفاده می‌شوند ـ الگوریتم ژنتیک به تنهایی یا در ترکیب با بعضی تکنیک‌های هوش مصنوعی دیگر کارآترین روش برای تشخیص نفوذ کشف شده است.

هدف یک IDS[۱۲] تشخیص ترافیک بد است. به منظور تحقق این هدف، IDS تمام ترافیک ورودی و خروجی را نظارت می‌کند. IDS بر اساس روش‌های تشخیص می‌تواند به دسته‌های زیر طبقه‌بندی شود[۴] :

- Anomaly detection[۱۳]: این تکنیک بر اساس تشخیص ناهنجاریهای ترافیک است. انحراف ترافیک نظارت شده از شکل نرمال اندازه‌گیری می‌شود.
- Misuse/Signature detection[۱۴]: این تکنیک به دنبال الگوها و امضاهای حملات معروف در ترافیک شبکه جستجو می‌کند. یک پایگاه داده‌ی همیشه به‌روز معمولاً برای ذخیره‌ی امضاهای حملات معروف استفاده می‌شود. روشی که این تکنیک با تشخیص نفوذ برخورد می‌کند همانند روشی است که نرم‌افزار آنتی‌ویروس عمل می‌کند.

IDS بر طبق منابع داده نیز می‌تواند به دو دسته‌ی تشخیص مبتنی بر میزبان و تشخیص مبتنی بر شبکه تقسیم شود[۲].

- ✓ در تشخیص مبتنی بر میزبان ـ پَروَنجاه‌ای داده و فرآیندهای سیستم‌عامل میزبان به‌طور مستقیم نظارت می‌شوند تا تعیین شود دقیقا کدام منابع میزبان، اهداف یک حمله خاص هستند.
- ✓ در سیستم‌های تشخیص نفوذ مبتنی بر شبکه، داده‌ی ترافیک شبکه را با استفاده از یک مجموعه از سنسورهایی که به شبکه ضمیمه شده نظارت می‌کنند تا هرگونه فعالیت بد را ثبت کنند.

گرایش جاری در تشخیص نفوذ، ترکیب هر دوی اطلاعات مبتنی بر میزبان و شبکه برای توسعه‌ی سیستم‌های ترکیبی است. در ادامه ابتدا الگوریتم ژنتیک در حالت کلی شرح داده شده است و سپس کاربردهای مختلف آن در تشخیص نفوذ، فواید و معماری آن برای IDS ارائه می‌شود هم چنین برخی کارهای مرتبط در این زمینه در دهه‌ی اخیر مطرح شده‌اند، بخش ۵ مجموعه داده گان مورد استفاده در این زمینه را معرفی می‌کند و در پایان برخی نتایج ارائه شده‌اند.

۲- الگوریتم ژنتیک

محدوده‌ی کاری الگوریتم ژنتیک بسیار وسیع می‌باشد و هر روز با پیشرفت روز افزون علوم و تکنولوژی استفاده از این روش در بهینه‌سازی[۱۵] و حل مسائل بسیار گسترش یافته است. الگوریتم ژنتیک را می‌توان یک روش جستجوی کلی نامید که از قوانین تکامل بیولوژیک طبیعی تقلید می‌کند. الگوریتم ژنتیک بر روی یک سری از جواب‌های مساله به امید بدست آوردن جواب‌های بهتر قانون بقای بهترین را اعمال می‌کند. درهر نسل به کمک فرآیند انتخابی متناسب با ارزش جواب‌ها و تولید مثل جواب‌های انتخاب شده به کمک عملگرهایی که از ژنتیک طبیعی تقلید شده‌اند، تقریب‌های بهتری از جواب نهایی بدست می‌آید. این فرآیند باعث می‌شود که نسل‌های جدید با شرایط مساله سازگارتر باشد[۶].

بدن هر موجود زنده‌ای از سلول تشکیل شده و هر سلول نیز شامل کروموزوم‌ها می‌باشد.کروموزوم‌ها هم از ژن‌ها تشکیل شده‌اند . بطور کلی

[۱۱] malicious
[۱۲] Intrusion Detection System
[۱۳] تشخیص ناهنجاری
[۱۴] تشخیص امضا
[۱۵] optimization

اجزای الگوریتم‌های ژنتیک به شرح ذیل می‌باشند:

- کروموزوم[۱۶] : در الگوریتم‌های ژنتیک، هر کروموزوم نشان دهنده یک نقطه در فضای جستجو و یک راه‌حل ممکن برای مسئله مورد نظر است. خود کروموزوم‌ها (راه حل‌ها) از تعداد ثابتی ژن[۱۷] (متغیر) تشکیل می‌شوند.
- جمعیت[۱۸] : مجموعه‌ای از کروموزوم‌ها یک جمعیت را تشکیل می‌دهند. با تاثیر عملگرهای ژنتیکی بر روی هر جمعیت, جمعیت جدیدی با همان تعداد کروموزوم تشکیل می‌شود.
- تابع برازندگی[۱۹] : به منظور حل هر مسئله با استفاده از الگوریتم‌های ژنتیک، ابتدا باید یک تابع برازندگی برای آن مسئله ابداع شود. برای هر کروموزوم، این تابع عددی غیر منفی را برمی‌گرداند که نشان دهنده شایستگی یا توانایی فردی آن کروموزوم است.

عملگرهای الگوریتم ژنتیک

در الگوریتم‌های ژنتیک، در طی مرحله تولید مثل[۲۰] از عملگرهای ژنتیک استفاده می‌شود. با تاثیر این عملگرها بر روی یک جمعیت، نسل[۲۱] بعدی آن جمعیت تولید می‌شود. عملگرهای انتخاب[۲۲]، ادغام[۲۳] و جهش[۲۴] معمولاً بیشترین کاربرد را در الگوریتم‌های ژنتیک دارند.

- عملگر انتخاب: این عملگر از بین کروموزوم‌های موجود در یک جمعیت، تعدادی کروموزوم را برای تولیدمثل انتخاب می‌کند. کروموزوم‌های برازنده‌تر شانس بیشتری دارند تا برای تولیدمثل انتخاب شوند.
- عملگر ادغام: در جریان عمل تلفیق به صورت اتفاقی بخش‌هایی از کروموزوم‌ها با یکدیگر تعویض می‌شوند.
- عملگر جهش : پس از اتمام عمل ادغام، عملگر جهش بر روی کروموزوم‌ها اثر داده می‌شود. این عملگر یک ژن از یک کروموزوم را به طور تصادفی انتخاب نموده و سپس محتوای آن ژن را تغییر می‌دهد. اگر ژن از جنس اعداد دودویی باشد آن را به وارونش تبدیل می‌کند و چنانچه متعلق به یک مجموعه باشد مقدار یا عنصر دیگری از آن مجموعه را به جای آن ژن قرار می‌دهد.

روند کلی الگوریتم ژنتیک : کار الگوریتم ژنتیک با جمعیت تولید شده بطور تصادفی از کروموزوم‌ها شروع می‌شود. از طریق تولیدات مختلف این جمعیت نمو می‌یابد و کیفیت کروموزوم‌ها بهبود می‌یابد.

قبل از این که یک الگوریتم ژنتیک بتواند اجرا شود، ابتدا باید کدگذاری(یا نمایش) مناسبی برای مسئله مورد نظر پیدا شود. معمولی‌ترین شیوه نمایش کروموزوم‌ها در الگوریتم ژنتیک به شکل رشته‌های دودویی است. هر متغیر تصمیم‌گیری به صورت دودویی در آمده و سپس با کنار هم قرار گرفتن این متغیرها کروموزوم ایجاد می‌شود. همچنین یک تابع برازندگی نیز باید ابداع شود تا به هر راه‌حل کدگذاری شده ارزشی را نسبت دهد. در طی اجرا، والدین برای تولیدمثل انتخاب می‌شوند و با استفاده از عملگرهای ادغام و جهش با هم ترکیب می‌شوند تا فرزندان جدیدی تولید کنند. این فرآیند چندین بار تکرار می‌شود تا نسل بعدی جمعیت تولید شود. سپس این جمعیت بررسی می‌شود و در صورتی که ضوابط همگرایی برآورده شوند، فرآیند فوق خاتمه می‌یابد. همگرایی را می‌توان به این صورت بیان کرد که وقتی یک درصد ثابتی از سطر و ستون‌های ماتریس جمعیت شبیه هم می‌شوند می‌توان فرض کرد که همگرایی صورت گرفته است. این درصد ممکن است ۸۰٪ یا ۸۵٪ باشد[۲].

۳- الگوریتم ژنتیک و IDS

الگوریتم‌های ژنتیک در ابتدا در رشته زیست‌شناسی محاسباتی معرفی شده بودند. از آن به بعد آنها در رشته‌های مختلفی با نتایج محتمل به کاربرده شده‌اند. اخیراً هم محققان تلاش دارند تا این الگوریتم‌ها را با IDS ترکیب کنند. محققان به منظور تشخیص نفوذ از GA[۲۵] یا برای تولید قوانین classification یا برای انتخاب ویژگی‌های مناسب کروموزوم‌ها استفاده کرده‌اند. GA به دلیل قدرت و سادگی عملیات و برتری نتیجه به

[۱۶] Chromosome
[۱۷] Gene
[۱۸] Population
[۱۹] Fitness Function
[۲۰] Reproduction
[۲۱] Generation
[۲۲] Selection
[۲۳] Crossover
[۲۴] Mutation
[۲۵] Genetic Algorithm

عنوان تکنیکی کارا برای تشخیص نفوذ استفاده شده است[۱۵].

۳-۱- کاربرد GA در انتخاب ویژگی :

تشخیص نفوذ در اصل یک مسئله‌ی classification است[۵] و اولین مسئله برای حل در classification، انتخاب[۲۶] و استخراج ویژگی[۲۷] است. است. یک ارتباط خطی بین ویژگی‌ها و عملکردهای classifier وجود ندارد. اما وقتی که تعداد feature ها از یک حد مشخصی تجاوز کند تغییر در عملکرد classifier ایجاد خواهد کرد. اصطلاحاً feature selection ، انتخاب خروجی مربوط یا زیرمجموعه‌ی feature مهم از مجموعه‌ی ویژگی اصلی بر طبق یک تابع ارزیابی مشخص است وتا زمانیکه امکان دارد کاهش ابعاد فضای ویژگی به فرض عدم کاهش دقت classification انجام می‌گیرد.

برای داده‌ی با ابعاد زیاد، feature selection نه تنها می‌تواند زمان تشخیص و هزینه را کاهش دهد بلکه کارایی classification را توسعه می‌دهد و زیرمجموعه‌ی حاوی اطلاعات مفیدی را کشف می‌کند.

۳-۲- کاربرد GA برای تولید قوانین classification :

الگوریتم ژنتیک برای استنتاج قوانین جدید برای IDS استفاده می‌شود[۶]. با استفاده از این قوانین، ترافیک نرمال شبکه یا داده‌ی audit از ترافیک/داده غیر نرمال تفکیک می‌شود. قوانین در مجموعه قانون الگوریتم ژنتیک بصورت if-then هستند.

شرح ذیل ترکیب کلی برای قانون در الگوریتم ژنتیک هست :

if { condition } then { act }

condition به داده‌ای که باید بررسی شود برمی‌گردد و اگر شرط قانون true هست، act عکس‌العملی است که باید انجام شود. یک condition می‌تواند برای مواردی از قبیل شماره‌های پورت پروتکل‌های شبکه، پروتکل‌های استفاده شده، مدت زمان اتصال، آدرس IP منبع و مقصد چک کند. حال آنکه act به عکس‌العملی که باید انجام شود وقتی که شرط true هست مانند فرستادن پیغام alert و ایجاد پیغام‌های log بر می‌گردد.

ویژگی‌های مختلف شبکه می‌توانند برای تشخیص نفوذ در شبکه بررسی شوند همانند مدت زمان اتصال، پروتکل‌های مورد استفاده ، پورت‌های مبدأ و مقصد، IP مبدأ و مقصد و نام حمله. به عنوان نمونه اگر اولین ۶ ویژگی با استفاده از عملیات AND منطقی برای ساختن بخش condition قانون به هم متصل شده‌اند، آنگاه ویژگی نام حمله به عنوان بخش act قانون استفاده می‌شود. شرح ذیل مثال ساده‌ای است که یک اتصال شبکه را به عنوان حمله انکار سرویس Neptune طبقه‌بندی می‌کند.

if (duration = “۰:۰:۱” and protocol = “finger” and source_port = ۱۸۹۸۹ and destination_port = ۷۹ and source_ip = “۹۹.۱۹.۹۹.۱۹” and destination_ip=“۱۹۲.۱۶۸.۲۵۴.۱۰”) then (attack_name = “Neptune”)

بعد از تولید قوانین classification در مرحله‌ی قبل ، قوانین fittest برای هدف تشخیص بکار برده می‌شوند.

تابع fitness :

هر کروموزوم بعد از بکاربردن تابع fitness برای آن انتخاب می‌شود. اگر یک قانون بصورت if A then B نمایش داده شود آنگاه fitness قانون مطابق ذیل است :

support = | A and B | / N
confidence = | A and B | / | A|
fitness = w۱ * support + w۲ * confidence

در اینجا N تعداد کل اتصالات شبکه در داده‌ی audit، | A | تعداد اتصالات شبکه منطبق با شرط A و | A and B | تعداد اتصالات شبکه که با قانون if A then B مطابقت می‌کند. وزنهای W۱,W۲ نیز برای کنترل توازن بین دو عبارت استفاده می‌شوند.

۳-۳- فواید بکارگیری GA برای تشخیص نفوذ :

۱)الگوریتم‌های ژنتیک فی‌نفسه موازی هستند. به دلیل چندین زادوولد آنها می‌توانند فضای راه‌حل را در چندین جهت باهم جستجو کنند.
۲)موازی‌سازی به الگوریتم ژنتیک اجازه می‌دهد تا به طور تلویحی طرح‌های بسیاری را یک‌دفعه و باهم ارزیابی کند. این امر آنها را مناسب

[۲۶] selection
[۲۷] feature extraction

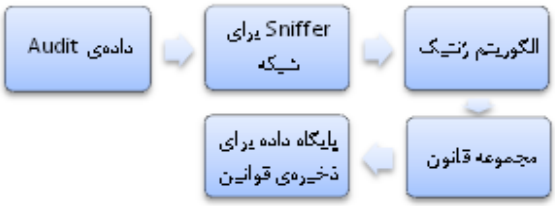
برای حل مسائلی که فضای راه‌حل بالقوه حقیقتاً بزرگ هست می‌سازد.

۳)سیستم‌های مبتنی بر الگوریتم ژنتیک می‌توانند به آسانی دوباره آموزش (retraining) ببینند. این امر امکان افزودن قوانین جدید را توسعه می‌دهد و سیستم تشخیص نفوذ را پیشرفت می‌دهد.

۴) پرهیز از گیرکردن در ماکسیمم‌های محلی[۲].

۳-۴- معماری GA برای IDS :

این معماری نیازمند جمع‌آوری داده‌ی شبکه برای audit که داده‌ی نرمال و غیر نرمال را در برمی‌گیرد است. بعد از جمع‌آوری داده، Sniffer[۲۸] شبکه ، داده را آنالیز خواهد کرد و آن را به الگوریتم ژنتیک خواهد فرستاد . بعد از بکارگیری تابع fitness ، قوانین به مجموعه قانون اضافه می‌شوند که در پایگاه قانون ذخیره می‌شوند. شکل زیر مراحل ذکر شده به منظور تشخیص نفوذ را به تصویر می‌کشد[۶].



شکل (۱) : الگوریتم ژنتیک برای IDS[۶]

۴- کارهای مرتبط

Ajith Abraham در سال ۲۰۰۴ یک روش محاسبه نرم برای تشخیص نفوذها در شبکه ارائه داد[۷]. نویسنده مناسب بودن تکنیک برنامه‌نویسی ژنتیک خطی را برای مدلسازی سریع وکارای این سیستم‌ها بررسی کرده است. در مقایسه با بیانات عملیاتی یا درخت‌های syntax مورد استفاده در برنامه‌نویسی ژنتیک سنتی، برنامه‌نویسی ژنتیک خطی یک ساختار برنامه‌ی خطی به عنوان ماده ژنتیک که ویژگی‌های اصلی آن برای کسب هردوی زمان اجرا و پیشرفت کامل به کار گرفته شده، استفاده می‌کند. کارایی این روش قابل مقایسه با نتایج حاصله از شبکه‌های عصبی مصنوعی و متدهای درخت رگرسیون می‌باشد.

در سال ۲۰۰۵، Ren Hui Gong و همکارانش [۸] در ابتدا ۶ ویژگی وزن‌دار را بر اساس تجربه برای شرکت در قوانین classification انتخاب کردند سپس از یک الگوریتم ژنتیک ساده برای استنتاج یک مجموعه از قوانین classification از داده‌ی ممیزی شبکه استفاده‌کرده و چارچوب support-confidence را به عنوان تابع fitness برای قضاوت در مورد کیفیت هر قانون بکار بردند. قوانین تولید شده سپس برای تشخیص یا طبقه‌بندی نفوذهای شبکه در یک محیط زمان واقعی استفاده شدند. این روش نیز مشکلاتی به همراه دارد مثلاً درحالیکه چارچوب-support confidence آسان برای پیاده‌سازی است و دقت پیشرفته‌ای برای قوانین نهایی فراهم می‌کند، نیاز به این دارد که کل داده آموزشی قبل از هر محاسبه‌ای در حافظه بارگذاری شود که برای مجموعه داده‌گان آموزشی بزرگ این امر کارا نیست یا غیرممکن است. استفاده از بعضی مرتب‌سازی‌های تکنولوژیهای cash می‌تواند این مسئله را حل کند.

Jungtaek Seo و همکاران در سال ۲۰۰۵ الگوریتم ژنتیک را برای انتخاب ویژگی به کار بردند[۹] و سپس time delay preprocessing را بر اساس جریان packet پیشنهاد داده و در نهایت از SVM[۲۹] برای classification استفاده کردند. در این روش نرخ false positive هنوز هم قابل قبول نیست و svm classifier نیاز به اصلاح دارد.

Yong Wang و همکاران در سال ۲۰۰۹ یک تابع fitness پیشنهاد کردند [۱۰]، یک تولید کننده‌ی قانون کارا برای حمله‌ی انکار سرویس. در عمل این روش، روشی کارا برای مقابله با تمام انواع حملات نیست و جنبه‌ای کاربردی برای چند نوع حمله‌ی خاص دارد. در سال ۲۰۰۹ همچنین یک مدل الگوریتم آموزشی بر اساس تشخیص نابهنجاری توسط Chen Zhongmin و همکارانش طراحی شد[۱۱]. مدل آزمایش پیشنهادی بر اساس این فرضیه است که اگر متغیر x ، بارها بیشتر از یک مقدار مطلوب به نظر بیاید، احتمال اتفاق نابهنجاری وجود دارد. این روش نیز پایه‌ی

[۲۸] نرم‌افزاری برای نشان دادن آمار سایت در شبکه

[۲۹] Support Vector Machine

محکمی ندارد و مبنایش تجربی است.

Siva S. Sivatha Sindhu و همکارانش[۱۲] با هدف ساخت سیستم IDS آزمایشی برای طبقه‌بندی multiclass ، در ابتدا الگوی ترافیک ورودی را پیش پردازش کرده و نمونه‌های زائد آن را حذف نمودند سپس یک الگوریتم انتخاب ویژگی مبتنی بر الگوریتم ژنتیک طراحی کردند که یک تأثیر بیشتر روی حداقل سازی پیچیدگی محاسباتی classifier داشت. در نهایت یک مدل neurotree به عنوان موتورclassification بکار گرفتند که نرخ تشخیص را بالاتر از NN* وC۴۵ گسترش یافته، رساند.

در سال ۲۰۱۲ Sriparna Saha و همکارانش یک سیستم تشخیص نفوذ مبتنی بر یادگیری ماشین توسعه دادند[۱۶]. آنها الگوریتم ژنتیک را به همراه SVM برای تعیین اتوماتیک مجموعه‌ی مناسب از ویژگی‌ها به کار بردند. در این روش تابع هدف منطبق با یک کروموزوم خاص برابر است با f = error_rate . هدف به حداقل رساندن این تابع هدف می‌باشد. نتایج نشان داد که error rate بعد از بکارگیری GA به اندازه ۹٪ کاهش می‌یابد. این مسئله از کاربرد GA برای انتخاب ویژگی classifier مبتنی بر SVM حمایت می‌کند. همچنین یک دیکشنری از پیش تعریف شده از انواع حمله وجود دارد و با استفاده از روش GA وSVM متمایزترین ویژگی‌ها برای هر نوع حمله در دسترس است. بنابراین یک پاکت ورودی می‌تواند تحت SVM برای classify کردن آن به عنوان نرمال یا یکی از انواع حمله اجرا شود.

Bharat S. Dhak و Shrikant Lade در سال ۲۰۱۲ یک تکنیک کاربرد الگوریتم تکاملی یعنی الگوریتم ژنتیک به سیستم تشخیص نفوذ ارائه دادند[۱۷]. آنها برای ساخت یک پروفایل اتصال فقط از۵ فیلد مهم موجودیت‌های بسته‌ی ورودی استفاده کردند همچنین یک تابع fitness خاص برای روش خود بکار بردند. بنابراین در این مقاله به‌طور موفق یک مجموعه قانون و پروفایل اتصال شبکه که بتواند به‌خوبی نفوذهای جدید را شناسایی کند استنتاج شده است. این سیستم می‌تواند با هر سیستم IDS یکپارچه شود تا کارایی و عملکرد آن را توسعه دهد.

در سال ۲۰۱۳ Shuxin ZHU و Bin Hu با ترکیب ویژگیهای نوع filter وwrapper، یک نوع روش ترکیبی برای انتخاب ویژگی با استفاده از یک الگوریتم ژنتیک توسعه‌یافته شامل مکانیزم تنبیه و پاداش را پیشنهاد دادند[۵]. این مکانیزم می‌تواند همگرایی سریع این الگوریتم روی راه‌حل بهینه‌ی کلی تقریبی را تضمین کند. بر طبق نتایج آزمایشی این الگوریتم بسیار خوب عمل می‌کند و پیچیدگی زمانی آن کم است اما عیبی که دارد این است که روند کار شکل ساده‌ای ندارد .

۵- مجموعه داده‌گان مورد استفاده

۵-۱- DARPA ۱۹۹۸[۳۰] :

تمام محققان الگوریتم ژنتیک شان را روی داده‌ی offline شبیه به داده‌ی DARPA ۱۹۹۸ یا داده‌ی KDD CUP ۹۹ پیاده‌سازی کرده‌اند. MIT Lincoln Laboratory تحت آژانس پروژه‌های تحقیق پیشرفته‌ی دفاع (DARPA) و تکفل AFRL[۳۱] ، اولین داده‌ی استاندارد را برای سیستم‌های تشخیص نفوذ شبکه کامپیوتر جمع‌آوری و توزیع کرده‌اند. این داده، داده‌ی DARPA ۱۹۹۸ است و شامل فایلهای لیست BSM و tcp dump می‌شود. هر خط در یک فایل لیست به یک نشست[۳۲] جدا متعلق است و هر نشست به یک اتصال TCP/IP منحصربه فرد بین دو کامپیوتر مرتبط می‌شود.[۱۳]

۵-۲- KDD CUP ۹۹ :

این داده قسمتی از داده‌ی جمع‌آوری شده از برنامه‌ی ارزیابی تشخیص نفوذ DARPA ۱۹۹۸ MIT Lincoln Lab است. به عنوان یک ورژن از مجموعه داده‌ی DARPA ۱۹۹۸، KDD ۹۹ اولین بار در سومین رقابت ابزارهای داده‌کاوی و کشف دانش بین‌المللی[۳۳] استفاده شده بود و حالا به‌عنوان یک معیار استاندارد برای ارزیابی IDS مبتنی بر داده‌کاوی ملاحظه می‌شود. درKDD ۹۹، رکوردهای داده‌ی حمله‌ها به چهار دسته‌ی اصلی تقسیم می‌شوند : DOS ، R۲L ، U۲R ، Probing . به منظور تشخیص اتصالات نرمال از حملات، هر آیتم داده از اتصال شبکه در KDD ۹۹ با ۴۱ ویژگی سطح بالا شرح داده می‌شود. در ارتباط با ۴ نوع حمله، ویژگی‌های آنها به صورت چهار دسته‌ی زیر مشتق می‌شوند : ۱)ویژگی‌های اصلی اتصال TCP منحصر به فرد ۲)ویژگی‌های ترافیک محاسبه شده با استفاده از یک پنجره زمان دو ثانیه‌ای ۳) ویژگی‌های ترافیک مبتنی بر

[۳۰] Defence Advanced Research Projects Agency
[۳۱] Air Force Research Laboratory
[۳۲] session
[۳۳] Third International Knowledge Discovery and Data Mining Tools Competition

میزبان ۴)ویژگی‌های محتوا در بین یک اتصال پیشنهاد شده با دانش دامنه

علاوه بر این هر آیتم داده‌ی این مجموعه داده به نوان نرمال یا یک حمله با یک نوع حمله‌ی خاص برچسب‌گذاری می‌شود. بطورکلی ۲۳ نوع حمله وجود دارد و تمام آنها متعلق به ۴ دسته حمله اصلی ذکر شده هستند[۱۴].

داده بطور کامل (۷۴۳M از اتصالات شبکه) یا در تعداد مجموعه‌های داده‌ی کوچکتر قابل دسترس است.

۶- نتیجه‌گیری

در این مقاله، مروری بر تشخیص نفوذ ارائه شد و روشهای مختلف برای کاربرد الگوریتم ژنتیک برای تشخیص نفوذ در شبکه بحث شد. الگوریتم ژنتیک متد جستجوی تصادفی است که اغلب برای مسئله‌ی بهینه‌سازی استفاده می‌شود. آن بطور موفق در IDS پیاده‌سازی شده تا قوانین classification تولید کند یا برای انتخاب ویژگی‌های مناسب استفاده شود. این دو کاربرد الگوریتم ژنتیک در سیستم تشخیص نفوذ در متن شرح داده شد. سه فاکتور که روی کارایی الگوریتم ژنتیک اثر دارند انتخاب تابع fitness، نمایش کروموزوم‌ها و مقادیر پارامترهای GA می‌باشند. تعیین این فاکتورها اغلب به موارد کاربرد بستگی دارد. طراحی تابع fitness دقیق، یک چالش عمده برای حل یک مسئله‌ی خاص است. یک سیستم ترکیبی با استفاده از الگوریتم ژنتیک می‌تواند راه‌حل بهتری برای کاهش نرخ false positiveها باشد.

۷- مراجع

[۱] Huy Anh Nguyen and Deokjai Choi, Application of Data Mining to Network Intrusion Detection: Classifier Selection Model, ۳۰۰ Yongbong-dong, Buk-ku Gwangju ۵۰۰-۷۵۷, Korea,۲۰۰۸

[۲] Jyotiprakash Sahoo, Subasish Mohapatra, Radha Lath ,A Survey on Evolutionary Approaches to Intrusion Detection Systems,۲۰۱۰

[۳] Ahmed Youssef and Ahmed Emam,NETWORK INTRUSION DETECTION USING DATA MINING AND NETWORK BEHAVIOUR ANALYSIS ,۲۰۱۱

[٤] Theodoros Lappas and Konstantinos Pelechrinis , Data Mining Techniques for Network Intrusion Detection Systems,۲۰۰۵

[٥] Shuxin ZHU,Bin HU ,Hybrid Feature Selection Based on Improved Genetic Algorithm,۲۰۱۳

[٦] Vivek K. Kshirsagar, Sonali M. Tidke & Swati Vishnu, Intrusion Detection System using Genetic Algorithm and Data Mining: An Overview, International Journal of Computer Science and Informatics ISSN: ۲۲۳۱ –٥۲۹۲, Vol-۱, Iss-٤, ۲۰۱۲

[۷] A. Abraham, ―Evolutionary Computation in Intelligent Network Management , in Evolutionary Computing in Data Mining, Springer, pp. ۱۸۹—۲۱۰, ۲۰۰٤

[۸] Ren Hui Gong, Mohammad Zulkernine, Purang Abolmaesumi ,A Software Implementation of a Genetic Algorithm Based Approach to Network Intrusion Detection,۲۰۰۵

[۹] Taeshik Shon, Jungtaek Seo۲, and Jongsub Moon ,SVM Approach with a Genetic Algorithm for Network Intrusion Detection,۲۰۰۵

[۱۰] Yong Wang, Dawu Gu,Xiuxia Tian ,Jing Li ,Genetic Algorithm Rule Definition for Denial of Services Network Intrusion Detection,۲۰۰۹

[۱۱] Chen Zhongmin, Feng Jianyuan, Xu Sheng and Xu Renzuo ,"The research of Intrusion Detection Technology Based on Genetic Algorithms",۲۰۰۹

[۱۲] Siva S. Sivatha Sindhu , S. Geetha , A. Kannan ,Decision tree based light weight intrusion detection using a wrapper approach ,۲۰۱۲

[۱۳] S. N. Pawar, INTRUSION DETECTION IN COMPUTER NETWORK USING GENETIC ALGORITHM APPROACH: A SURVEY, International Journal of Advances in Engineering & Technology, May ۲۰۱۳

[۱٤] Wenying Feng, Qinglei Zhang, Gongzhu Hu, Jimmy Xiangji Huang, Mining Network Data for Intrusion Detection through Combining SVM with Ant Colony,۲۰۱۳

[۱٥] S. N. Pawar and R. S. Bichkar, "Using Enumeration in a GA based Intrusion Detection", International Journal of Computer Applications (IJCA), October, ۲۰۱۲.

[۱٦] Sriparna Saha, Ashok Singh Sairam, Asif Ekbal," Genetic Algorithm Combined with Support Vector Machine for Building an Intrusion Detection System",International Conference on Advances in Computing, Communications and Informatics (ICACCI-۲۰۱۲)

[۱۷] Bharat S. Dhak , Shrikant Lade," An Evolutionary Approach to Intrusion Detection System using Genetic Algorithm", International Journal of Emerging Technology and Advanced Engineering, ISSN ۲۲۵۰-۲۴۵۹, ISO ۹۰۰۱:۲۰۰۸ Certified Journal, Volume ۲, Issue ۱۲, December ۲۰۱۲