بررسی الگوریتم بسیار سریع درخت تصمیم (VFDT) جهت تشخیص مزاحمت شبکه

چکیده مقاله

تکنولوژی شبکه با توجه به پیشرفت های اخیرباید ا ز کامپیوترها و شبکه های حمایت و محافظت کند که تبدیل به یک مسئله بزرگ در جهان امروزی شده اند. کامپیوتر براساس اطلاعاتی که از گروه های مختلف پاسخ دریافت میکند در هر ثانیه یک بار مورد حمله قرار می گیرد و در آن وقفه ایجاد می شود. در این مقاله دو سطح دانه ای سیستم تشخیص مزاحمت (IDS) یعنی ریز دانه و درشت دانه توضیه می شود. مزاحمت ها در نمونه طبیعی تشخیص داده نمی شوند، و در نتیجه سطح درشت دانه مناسبترین سطح IDS جهت افزایش عملکرد است . پس از تشخیص مزاحمت به وسیله IDS درشت دانه، سطح ریز دانه جهت تشخیص اطلاعات و جزئیات حمله احتمالی استفاده می شود. الگوریتم بسیار سریع درخت تصمیم (VFDT) در هر دو سطح تشخیص استفاده می شوند. این سطح به منظور تضمین اثر بخشی مدل پیشنهادی، روی مجموعه داده های KDD CUP 99 و مجموعه داده های حقیقی ترافیک آزمایش شده است. نتایج آزمایشی موفقیت بسیار بالا مدل پیشنهادی را در تشخیص حمله های مشخص و نامشخص ثابت میکند.

کلمات کلیدی : درخت بسیار سریع مزاحمت، تشخیص مزاحمت، مجموعه داده های کشف دانش، IDS درشت دانه و IDS ریز دانه

1-مقدمه

سیستم تشخیص مزاحمت (IDS) تمامی فعالیت شبکه را بررسی و کنترل میکند و الگوهای مشکوک رشد یافته در شبکه ها و افزایش مزاحمت ها را در کامپیوترها شناسایی میکند. سیستم تشخیص مزاحمت مولفه اصی معماری کامل دفاع –عمیق امنیت شبکه است. این سیستم بسته ها را بازبینی و جمع آورد می کند، و اسناد و ادله های رفتارهای مزاحم را جستجو می کند. زنگ خطر پس از تشخیص پیشامد مزاحم به صدا در می آید و فرصت واکنش به موقع را برای تحلیل گر امنیتی فراهم می سازد. اکثر IDS های طراحی شده نمی توانند با شبکه های سریع کنار بیایند. اگر چه سیستم های IDS بسیاری در دسترس و موجود هستند، اما کاهش

تعدادی از هشدارهای کاذب و دروغین و تشخیص حملات جدید به منظور افزایش نسبت تشخیص هدف مشترک این سیستم ها هستند. در این مقاله، توجه به تشخیص حملات در شبکه های سریع به منظور کاهش اثر حمله بر کاهش فاصله زمانی بین حمله حقیقی و تشخیص آن است. این مقاله به ساخت دو سطح دانه های IDS به منظور تشخیص رفتار نابهنجار ترافیک شبکه و سازگاری با شبکه های سریع مثل شبکه ریز و درشت دانه کمک میکند. مشهود است که وقوع مزاحمت در شبکه های به خاطر وجود ترافیک عمومی و کلی در صورت عدم تشخیص مزاحمت موضوع مهمی است. این موارد ما را به ساخت سطوح ریز و درشت دانه IDS بر می انگیزاند. سطح دشت دانه به منظور افزایش عملکرد در نمونه های طبیعی که مزاحمت ها تشخیص داده نمی شوند، مناسبترین سطح IDS است. زمانی که مزاحمت توسط IDS درشت دانه تشخیص داده می شود، از IDS ریز دانه برای تشخیص هر چه بیشتر جزئیات و اطلاعات حمله استفاده می شود. سیستم درشت دانه تشخیص مزاحمت توجه به 5 ویژگی بسته دارد، در حالی که سیستم ریز دانه تشخیص مزاحمت به بررسی 20 ویژگی می پردازد. الگوریتم بسیار سریع درخت تصمیم (VFDT) به عنوان یک طبقه بندی کننده سریع انتخاب شد. مزیت این سیستم پردازش و تحلیل ترافیک سرعت بالا شبکه، کشف و تشخیص دقیق حمله های جدید جهت کاهش هشدار کاذب به منظور بیشینه سازی اندازه و تشخیص مزاحمت ها در زمان حقیقی است.

مجموعه داده های DARPA KDD CUP 99 به عنوان آزمون کارایی IDS استفاده می شود که شامل 41 ویژگی است. ما این ویژگی ها را تحلیل و 20 ویژگی را انتخاب کردیم که دارای نسبت بهره اطلاعات بر میانگین مجموعه داده ها است. سپس، سیستم مورد نظر را آموزش دادیم و آزمایش کردیم.

2- تحقیق مرتبط

1- تشخیص مزاحمت و حمله طبقه بندی شده در سه تکنیک

2- روشهای مختلف محاسبات – نرم در سالهای اخیر برای توسعه سیستم های تشخیص مزاحمت طراحی شده اند. هدف اصلی این تحقیق توسعه، اجرا و ارزیابی سیستم های غیر متعارف، ناپیوسته تشخیص مزاحمت (IDS) با استفاده از سه تکنیک، قانون های ارتباط و پیوند داده کاوی، درختان تصمیم (الگوریتم ID3) ، و شبکه عصبی مصنوعی است و سپس مقایسه بین آنها جهت تصمیم گیردی در مورد برتری هر یک از آنها در اجرا سیستم تشخیص مزاحمت است. روشهای بسیاری جهت اصلاح این تکینک ها به منظور توسعه پروسه طبقه بندی

طراحی شده است. طبقه بندی کننده مهم در قوانین پیوند و ارتباطات به منظور ساخت طبقه بندی کننده جدید قادر به تشخیص نابهنجاری های و نادرستی ها اصلاح شد.الگوریتم ID3 توسط درختان تصمیم نه تنها برای بررسی داده های محتاطانه بلکه نیز برای بررسی داده های عددی اصلاح شد. الگوریتم پس انتشار در شبکه های عصبی به عنوان الگوریتم یادگیری با تعداد متفاوتی ورودی جهت آغا ز و شروع دانش مهم مزاحم در در شبکه های عصبی استفاده شده است. روشهای مختلف هنجار سازی در الگوهای ورودی جهت تسریع پروسه یادگیری اجرا شدند.مجموعه کامل داده های آموزشی 10% KDD 99 و مجموعه داده های کامل آزمون اصلاح و تصحیح در این تحقیق استفاده می شود. نتایج تکنیک های پیشنهادی افزایشس عملکرد این تکنیک ها را نسبت به تکنیک های استاندارد، برتری روشهای درصد پیش بینی موفق (PSP) و آزمون برمبنای هزینه شبکه های عصبی، و درختان تصمیم را بر قوانین پیوند و تداعی نشان می دهد. از طرف دیگر، آموزش شبکه های عصبی زمان بر تر از درختان تصمیم است.

2-سیستم های چند سطی تشخیص مزاحم (ML-IDS)

با افزایش استقرار سیستم های با مرکزیت شبکه، شدت و پیچیدگی حمله های شبکه افزایش می یابد. تکنیک های تشخیص حمله به صورت تکینک های امضاء گرا، طبقه بندی محور، یا غیر متعارف طبقه بندی می شوند. ما د راین تحقیق سیستم چند سطحی تشخیص مزاحمت (ML-IDS) را ارایه می دهیم که از محاسبات مستقل و خودکار جهت خودکار سازی کنترل و مدیریت IDS چند سطحی استفاده میکند. این اتوماسیون منجر می شود تا ML-IDS مزاحمت را در حمله های شبکه شناسایی کند و از آنها محافظت کند. ML-IDS ترافیک شبکه را با استفاده از سه سطح گرانولیته (جریان ترافیک شبکه، سر پیام بسته اطلاعات، و ظرفیت باری) بررسی و تحلیل میکند، و از الگوریتم سازشی درخت تصمیم جهت افزایش نسبت کلی تشخیص مزاحمت و کاهش وقوع هشدارهای دروغین استفاده میکند. ما هر یک از چهار روش را در دامنه وسیعی از حملات شبکه ای بررسی کردیم و سپس نتایج این روشها را با نتایج الگوریتم مرکیب سازشی تصمیم مقایسه کردیم.

3- نشخیص مزاحمت درخت تصمیم (الگوریتم ID3)

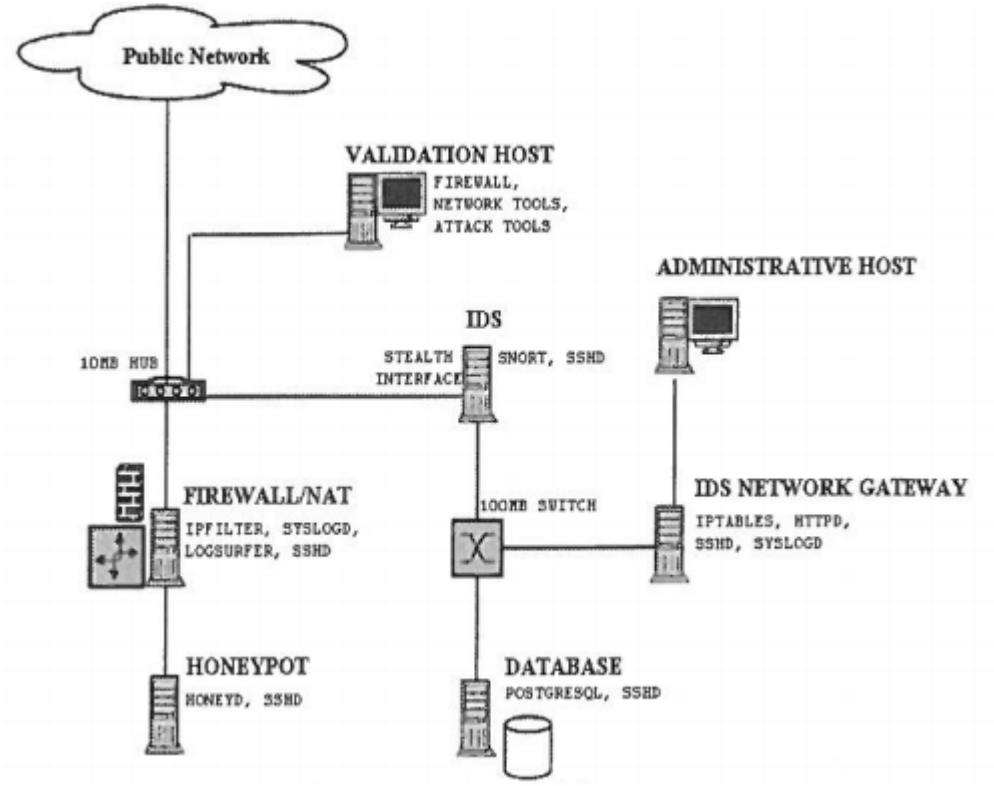
الگوریتم های طبقه بندی درخت تصمیم را مانند درختی می سازنند که به وسیله الگوهای مختلف تشیخصی در مجموعه داده موجود ارایه شده و از اطلاعات جهت ساخت درخت استفاده میکند. الگوریتم های درخت تصمیم از

داده های از پیش طبقه بندی شده به عنوان ورودی استفاده میکنند. آنها الگوهای موجود در داده ها را یاد می گیرند، و قوانین ساده را جهت تمیز بین انواع مختلف داده ها در مجموعه داده های از پیش طبقه بندی شده اجرا میکنند. الگوریتم IDS با مجموعه اصلی به عنوان گره ریشه شروع می شود. این الگوریتم روی هر یک از تکرارهای الگوریتم، ویژگی های استفاده نشده مجموعه داده ها را تکرار میکند، و آنتروپی آن ویژگی را محاسبه میکند. سپس این الگوریتم ویژگی دارای کمترین مقدار آنتروپی (یا بیتشرین بهره اطلاعاتی) را انتخاب میکند. سپس، مجموعه براساس ویژگی های انتخاب شده (مثلاً، سن کمتر از 50 سال، سن بین 50 و 100 سال، وسن بیش از 100 سال) جهت تهیه زیر مجموعه داده ها به دو نیم تقسیم می شود. الگوریتم به توسل و مراجعه به هر یک از زیر مجموعه ها ادامه می دهد، و فقط ویژگی های را بررسی میکند که هر گز قبلاً انتخاب نشده اند. خود بازگشتی به زیر مجموعه در یکی از نمونه های زیر متوقف می شود:» هر جزء زیر مجموعه متعلق به یک طبقه است، سپس، گرده به یک برگ تبدیل می شود، و با گروه نمونه های برچسب زده می شود. هر چند، ویژگی های بیشتری برای انتخاب وجود ندارد، اما نمونه ها هنوز به یک گروه تعلق ندارند، و گروه به برگ تبدیل میوشد و با مشترک ترین گروه نمونه ها در زیر مجموعه نشانه گذاری می شود.

4- سیستم تشخیص مزاحمت مبتنی بر KDD-99 و با استفاده از تکنیک های داده کاوی و انتخاب ویژگی

اینترنت روزانه و کاربران اینترنت در حال افزایش هستند. امنیت به علت توسعه سریع تکنولوژی اینترنت تبدیل به یک مسئله بزرگ می شود. مزاحم ها به طور پیوسته سیستم های شبکه کامپیوتری را برا ی حمله به آنها کنترل می کنند. تحقیق جامع و فراگیری از دست نوشته های علمی ثابت می کند که تکنیک های داده کاوی قوی ترین تکنیک برای توسعه IDS به عنوان طبقه بندی کننده هستند. اثر بخشی عملکرد طبقه بندی کننده یک مسئله مهم و حیاتی است، و نیز تعداد ویژگی های اسکن شده توسط IDS نیز باید بهینه شوند. در این تحقیق، دو تکنیک C5.0 و شکبه عصبی مصنوعی (ANN) با تکنیک انتخاب ویژگی استفاده می شوند. تکنیک های انتخاب ویژگی از برخی از ویژگی ها صرفنظر و آنها را کنار خواهند گذاشت، در حالی که تکینک C5.0 و ANN جهت طبقه بندی داده ها در گروه نرمال یا یکی از 5 نوع حمله به عنوان یک طبقه بندی کننده و دسته بندی کننده عمل می کنند. از مجموعه داده های KDD99 جهت آموزش و آزمایش مدلها استفاده می شود،

مدل C5.0 با تعدادی ویژگی و با حداکثر 100٪ دقت بهترین نتایج را بدست می آورد. عملکرد نیز از نظر اندازه تقسیم داده ها تایید می شود.



شکل 1: معماری سیستم

3- معماری سیستم

سیستم طرح شده از چهار مرحله پردازش یعنی جمع آوری داده ها، پیش پردازش، طبقه بندی، و پاسخ تشکیل شده است. هر دو سطح IDS جهت آموزش مدل نیاز مند اطلاعات ارتباطی هستند. بنابرین، سیستم اطلاعات را در هر زمانی جهت دستیابی به اتصال های کافی که منجر به ساخت درخت تصمیم حمله می شود، به روز درمی آورند. الگوریتم VFDT پس از دسترسی به اطلاعات اتصال / بسته در یکی از سطوح IDS به منظور طبقه بندی و تصمیم گیری (نرمال یا حمله) اجرا می شود. در صورت تشخیص حمله، گزارشی با ارایه اطلاعات حمله مثل آدرس های IP و زمان.... وغیره تهیه می شود.

وظیفه اصلی دو سطح دانه ای IDS شناسایی و تشخیص الگوهای مزاحمت با بررسی مشخصات نمونه است. درخت به شیوه بازگشتی و با جایگزینی برگها به جای گره های تصمیم ساخته می شود. آمارهای کافی مقادیر ویژگی در هر برگ ذخیره می شود. تابع ارزیابی رهگشار و مکاشفه ای جهت تعیین تبدیل ویژگی تقسیم شده از

برگها به گره ها استفاده می شود. گره ها حاوی ویژگی های تقسیم شده و برگها فقط حاوی نشـان هـای طبقـه هستند. برگ طبقه ای را نشان می دهد که نمونه نشان دار می کند. نمونه پس از ورود به درخـت، از ریشـه بـه برگ حرکت میکند، ویژگی مربوطه را در هر کی از گره ها بررسی می کند. آمارهای موجود پس از رسید ن نمونه به برگ به روز در آورده می شود. در این زمان، سیستم هر یک از شرایط و حالت های احتمالی را براساس مقادیر خصیصه بررسی می کند. اگر آمارها برای حمایت از یک آزمون بر دیگر ی کافی باشند، آنگاه برگ به گره تصمیم تبدیل می شود. گره تصمیم حاوی تعداد مقادیر احتمالی خصیصه منتخب آزمون تقسیم نصب شده است. درخت تصمیم برای طبقه بندی حمله ها می تواند بزرگ شود. البته نسبت خطا به علت پیچیده تر شدن درخت افزایش می یابد. بنابراین، درخت تصمیم مستعد و آمادگی کمینه سازی نسبت خطا را دارد و به اینگونه ساده تر و آسانتر قابل درک می شود.

فراوانی مزاحمت های کامپیوتری در طی دو دهه اخیر بـه سـرعت افـزایش یافتـه انـد. سیسـتم هـای تشـخیص مزاحمت (IDS ها) مولفه اصلی معماری کامل دفاع – عمیق امنیت شبکه هستند. آنها بسته ها را جمع آوری و کنترل می کنند، و شواهد وادله های رفتار مزاحم را جستجو میکنند. به محض تشخیص رخداد و اتفاق مـزاحم، هشدار اعلام شود و فرصت واکنش آنی را برای تحلیل گر امنیتی فراهم می سـازد. متاسـفانه، اکثـر IDD هـای طراحی شده نمی توانند با شبکه های سریع کنار بیایند. اگر چه سیستم هـای IDS بسـیاری وجـود دارنـد، امـا هدف مشترک این سیستم ها کاهش مقدار هشدارهای کاذب و دروغین و تشـخیص حمـلات جدیـد بـه منظـور افزایش نسبت تشخیص است. در این مقاله توجه به تشخیص حملات شناخته و شـناخته نشـده در شـبکه هـای سریع به منظور کاهش اثر حمله با کاهش فاصله زمانی بین حمله حقیقی و تشخیص آنها می شود.

الگوریتم VFDT (درخت تصمیم بسیار سریع)

VFDT سیستم عملکرد داده کاوی مبتنی بر درخت های تصمیم است. بنابراین، روشهای یادگیری طبقه بندی زیادی طراحی شده است، که روش یادگیری درخت تصمیم یکی از معمول ترین و پـر کـاربردترین ایـن روشـها است. سرعت طبقه بندی این الگوریتم و توصیفی از طبقه بندی ها در آن منجر شده که ایـن روش آسـانتر مـی شود. VFDT یکی از الگوریتم های جریان داده ای اکه از روش یادگیری درخت تصمیم حمایت میکند. بـا ورود داده ها، این جریان داده ای در ضمن طبقه بندی داده ها، به تدریج رشد می کند و افزایش مـی یابـد. VFDT

امکان استفاده از ملاک ارزیابی خصیصه را به عنوان استفاده از بهره اطلاعات یا شاخص Gini فراهم مـی سـازد.

این روش منجر به اصلاحات الگوریتم می شود.

الگوریتم VFD (S,X,G,δ)

ورودی : S توالی بردار ویژگی ها، x است.

تابع ارزیابی و بررسی، δ

گره

خروجی : HT درخت تصمیم است

آغاز پروسه

1: فرض کیند که HT درختی با برگهای مجزاl_1 (ریشه) است

2: فرض کنید که $X_1 = XU\{X_0\}$

3:فرض کنید که $\overline{G}_1(X_0)$، همان $\overline{G}$ بدست آمده از پیش بینی معمولی ترین گروه در s است

4: در هر گروه y_k

5: در هر مقدار X_{ij} هر خصیصه $X_i \in X$ است

6: فرض کنید که $n_{ijk}(l_I) = 0$ است

7: در هر نمونه(x,y_k) در s

8:(x,y) را به استفاده از HT در برگ l دسته بندی کنید

9: هر X_{ij} در X به گونه ای است که $X_i \in X_1$ است

10:عبارت $n_{ijk}(l)$ را اضافه کنید

11:l را با اکثر یت گروه های بین نمونه نشانه دار کنید

12: در صورت مشاهده نمونه ها در l

13:$\overline{G}_1(X_i)$ را در هر خصیصه $X_i \in X_1$ محاسبه کنید

14: از شمارش $n_{ijk}(l)$ استفاده کنید

15:فرض کنید که X_a ویژگی با بالاترین G_I است

16:فرض کنید که X_b ویژگی با دومین مقدار بالا G_I است

17: عبارت زیر را محاسبه کنید

$$\in = \sqrt{\frac{\left(R^2 In\left(1/\delta\right)\right)}{2n}}$$

18: اگر $G_l(X_a) - G_I(X_b) > \in$ و $X_a \neq X_0$ باشد، آنگاه

19:گره داخلی را جایگزین l کنید که روی X_3 به دو نیم تقسیم می شود

20: در هر شاخه از تقسیم

21: برگ جدید l_m را اضافه کنید، و $X_m = X - \{X_a\}$

22:فرض کنید که $\overline{G}_m(X_o)$ مقدار $\overline{G}$ بدست آمده از پیش بینی معمول ترین طبقه

23: در هر گروه y_k و هر مقدار x_{lj} است

24: فرض کنید

$$n_{ijk}(l_m) = 0$$

25: به HT برگردید

پایان

الگوریتم دو سطحی IDS

سیستم های ریز دانه IDS روی مجموعه SF ویژگی ها عمل می کنند. سیستم درشت دانه IDS روی مجموعه BF ویژگی ها عمل می کنند. سیستم دارای دو سطح دانه ای IDS هستند که امکان تحلیل ترافیک شبکه را روی گرانولیته های مختلف برای سیستم فراهم می سازد. IDS درشت و ریز دانه سطوح تشخیص سیستم هستند. در نمونه طبیعی که مزاحمت های شبکه تشخیص داده نمی شوند، IDS درشت دانه مناسبترین سطح است که در آن 5 ویژگی جهت افزایش عملکرد IDS کنترل می شوند. زمانی که مزاحت به وسیله IDS_ درشت دانه تشخیص داده می شود، IDS ریز دانه فعال می شود که در آن 20ویژگی به منظور تشخیص هر چه بیشتر جزئیات حمه کنترل می شوند. الگوریتم VFDT قادر به پردازش و تحلیل ترافیک های پرسرعت شبکه و تشخیص مزاحمت در زمان حقیقی هستند.

الگوریتم دو دانه

1-مد= سطح _H

2-مادامی که (P= ضبط بسته ()) {

3- پیش پردازش (PO

4- اگر (مد = سطح _H)

5- // IDS ریز دانه

6- E=BF(P)

7- E را با استفاده از مدل VFDT طبقه بندی کنید

8- اگر (حمله تشخیص داده شود) {

9- مد: سطح _C

10- هشدارها تولید و تهیه می شوند }

11- در غیر این صورت {

12- //IDS ریز دانه

13- E=SF(P) که E از مجموعه P ساخته می شود

14- E را با استفاده از مدل VFDT بسازید

15- اگر (حمله تشخیص داده شود)

16- هشدارهای تولید شده

17- اگر (حمله ای در دوره خاص تشخیص داده نشود)

18- مد= سطح _H

19- }

پایان

4- نتیجه گیری

IDS درشت دانه و ریز دانه به منظور فراهم سازی امکان تحلیل ترافیک شبکه روی گرانولیته های مختلف برای سیستم طراحی شد. این نوع IDS متفاوت از IDS های موجود است که در شرایط حمله یا موارد دیگر خـود را

با موقعیت و موضع شبکه سازگار میکند. IDS درشت دانه و IDS ریز دانه سطوح تشخیص آن هستند. این سطوح دقت درخت تعمیم و تشخیص حمله ها را افزایش می دهند.