سیزدهمین کنفرانس ملی انجمن کامپیوتر ایران
جزیره کیش، خلیج فارس، ایران ۱۹ الی ۲۱ اسفند ۱۳۸۶
انجمن کامپیوتر ایران
Computer Society of Iran
دانشگاه صنعتی شریف
دانشکده مهندسی کامپیوتر
مقاله کامل

سیستم تشخیص نفوذ پایگاه داده‌ها مبتنی بر داده‌کاوی الگوهای رفتار

محمد یوسف نادری
دانشگاه شهید بهشتی
مرکز تحقیقات مخابرات
m.naderi@mail.sbu.ac.ir

رضا حسن زاده
مرکز تحقیقات مخابرات
rhsnzadeh@yahoo.com

حمید فرهادی
دانشگاه شهید بهشتی
h.farhady@mail.sbu.ac.ir

مقصود عباسپور
استادیار دانشگاه شهید بهشتی
maghsoud@sbu.ac.ir

چکیده: در این مقاله، یک سیستم تشخیص نفوذ پایگاه داده‌ها مبتنی بر داده کاوی ارائه شده است. روش ارائه شده بر مبنای یافتن ناهنجاری ها در رفتار کاربران پایگاه داده ها می باشد و از تکنیکهای استخراج الگوی رفتارترتیبی و زمانی استفاده می نماید. روش ارائه شده شامل دو فاز یادگیری وتشخیص می باشد. هدف فاز یادگیری ساختن پروفایلی دقیق از رفتار کاربران بوسیله کاویدن الگوی رفتار آنان در سه سطح تراکنش،دستور و عمل از مستندات ثبت وقایع سیستم می باشد. در فاز تشخیص هر فعالیت جدیدکاربر با پروفایل وی مقایسه شده تا ناهنجاری احتمالی شناسایی و اخطار مناسب تولید شود.

واژه های کلیدی: امنیت کامپیوتر، امنیت پایگاه داده ها، داده کاوی، تشخیص تهاجم، الگوی رفتاری زمانی، الگوی رفتاری ترتیبی.

مقدمه

اطلاعات هر سازمانی با ارزش ترین دارایی آن سازمان میباشد و لذا نیازمند مدیریت و حفاظت مناسب در مقابل تخریب ویا دسترسی های غیر مجاز است .مکانیسم های امنیتی موجود در سیستم های مدیریت پایگاه داده‌ها امکانات پایه ای امنیتی نظیر اهراز هویت، کنترل دسترسی، رمز نگاری داده ها و ثبت فعالیتهای کاربران را ارائه میدهند. اما وجود چنین مکانیسم ها یی در سیستم مدیریت پایگاه داده‌ها به معنای در امان بودن داده ها از آسیب پذیری های احتمالی نبوده و تنها امکاناتی محدود در قبال مقابله با حملات داخل سازمانی و یا حملاتی که از طریق نرم افزار های کاربردی ناامن صورت میگیرد ارائه می کنند.

کمبود امکانات مربوط به تشخیص نفوذ در سیستم های مدیریت پایگاه داده امروزی مسئله‌ای مهم است زیرا حملات نشانه گرفته شده به سمت پایگاه داده ها ممکن است توسط دیگر سیستم‌های تشخیص نفوذ در سطح شبکه و یا سیستم عامل شناسایی نشود. به عنوان نمونه تشخیص حملات داخل سازمانی امری دشوار است چرا که توسط کاربران مجازی که به سیستم و داده ها دسترسی دارند صورت می-پذیرد. مثالی دیگر از این دسته حملات، حملات تزریق SQL است که در آن مهاجم پرسش ارسالی به پایگاه را به طرز زیرکانه‌ای تغییر داده و بدین ترتیب به داده های مهم دست می یابد.

در سال‌های اخیر پژوهشگران روشها و دیدگاه‌های متفاوتی را در رابطه با تشخیص نفوذ بیان کرده اند. اما بیشتر این تلاشها بر تشخیص نفوذ در سطح شبکه و یا سیستم عامل تمرکز داشته است که برای حفاظت از پایگاه داده ها مناسب نیستند. آنچه این مشکل را حل می کند، سیستمی است که سوء استفاده از داده را شناسایی کرده و به مدیر سیستم گزارش دهد و یا درصورت نیاز تغییرات اعمال شده در داده را به عقب بازگرداند. اینگونه سیستمی می بایست قابل اعتماد و اطمینان بوده و همچنین از لحاظ مصرف منابع سیستم کم هزینه باشد. یک نمونه از چنین سیستمهایی، سیستم *تشخیص نفوذ پایگاه داده ها* می باشد .

در این مقاله روشی نوین برای شناسایی حملات به پایگاه داده ها با استفاده از تکنیکهای داده کاوی[1] ارائه شده است. این روش از توانایی تشخیص بر پایه الگوهای رفتاری ترتیبی[2] و زمانی[3] کاربر برخوردار است.

در ادامه مقاله در بخش ۲ بطور خلاصه تحقیقات انجام شده قبلی توضیح داده خواهد شد. سپس در بخش ۳ راه حل پیشنهادی بطور کامل توضیح داده می شود. در بخش ۴ نتایج آزمایشات انجام گرفته با بکارگیری ساز و کارهای ارائه شده در مقاله نشان داده شده است و نهایتاً در بخش ۵، نتیجه گیری مقاله آمده است.

۱- مروری بر کار های انجام شده

همانطور که گفته شد، تحقیقات متعددی در حوزه تشخیص نفوذ در سطح شبکه و سیستم عامل انجام شده است[8,10,15]. با این وجود تنها تعداد اندکی از تحقیقات بر حوزه‌ی تشخیص نفوذ پایگاه داده‌ها تمرکز داشته اند.

در [6] سیستم DEMIDS پیشنهاد شده که یک سیستم تشخیص سوء استفاده برای پایگاه داده های رابطه ای است. این سیستم با استفاده از مفهوم "صفت های ارجاء شده نزدیک به هم" ولاگ های موجود اقدام به تشکیل پروفایلی برای هر کاربر می‌نماید. دراین روش فرض بر اطلاع از شمای پایگاه داده مورد حفاظت است که خود ضعفی میباشد که عمومیت کاربرد را از این سیستم میگیرد. در [13] سیستم

تشخیص نفوذ بلادرنگی برای پایگاه داده ها با استفاده از برچسب های زمانی ارائه شده که پایگاه را در سطح تراکنش ها رسد می کند. ایده ی کلیدی این پژوهش بکارگیری خواص بلادرنگ داده ها برای انجام فرآیند تشخیص نفوذ است.

Barbara در [3] از Hidden Markov Model (HMM) و سری‌های زمانی برای تشخیص خراب کاری در داده ها استفاده میکند. در این روش یک مدل رفتارپایگاه داده به کمک اندازه گیری تغییرات رفتار در زمان ساخته میشود. الگوهای حمله آن الگوهایی هستند که توسط HMM با احتمال بالا به عنوان الگو شناسایی نشوند. Zhong[18] نیز از الگوریتمی استفاده کرده است که پروفایل کاربران را بر مبنای پرسشهای ارسالی کاربران به پایگاه تشکیل داده و از آن در تشخیص نفوذ بهره می جوید. Hu [9] ایده ای را مبنی بر استخراج وابستگی های بین داده ها ارائه نموده است و تراکنش هایی را که داده ها را بدون توجه به وابستگی های آنها دست کاری می کند حمله اعلام می نماید.

در [4] نیز یک سیستم تشخیص نفوذ بر مبنای کنترل دسترسی مبتنی بر Role ارائه شده است. در [7] Diaz-Gomez روشی برای تشخیص سوء استفاده در پایگاه داده ها مبتنی بر الگویتمهای ژنتیکی طراحی نموده است و [5] نیز با استفاده از یک شبکه نورو فازی بنام ART اقدام به مدیریت عملکرد امنیتی پایگاه داده ها نموده است.

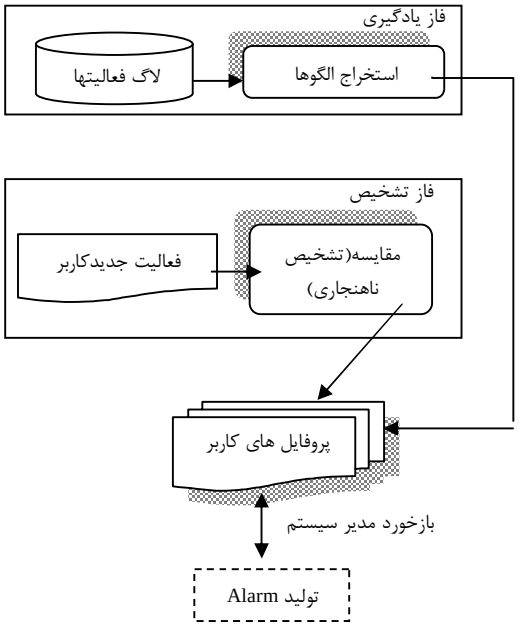
درنهایت، این مقاله بنا داردکارهای قبلی را گسترش داده و با ارائه مفهوم الگوی رفتاری زمانی و بکارگیری آن در کنار الگوی رفتاری ترتیبی، دقت شناسایی حملات را افزایش دهد.

۲- راه حل پیشنهادی

ایده ی ارائه شده بر پایه ی شناسایی ناهنجاریها[۴] در رفتار کاربران پایگاه داده ها می باشد . در این روش با بهره جویی از روشهای داده کاوی، الگوی رفتار هر کاربر از دو منظر ترتیبی و زمانی و در سه سطح انتزاع استخراج میگردد. طبق آزمایشات صورت پذیرفته در روش ارائه شده، به علت جامعیت الگوهای رفتاری استخراج شده نتایج تشخیص ناهنجاری، دقیقتر خواهد بود. در ادامه روش پیشنهادی به تفصیل توضیح داده خواهد شد.

۲-۱ شرح کلی سیستم

روند کلی کار شامل دو فاز است: فاز یادگیری و فازتشخیص. در فاز یادگیری لاگ فعالیتهای کاربر که شامل تمامی پرسش ها و تراکنش های وی است پردازش شده و الگوهای ترتیبی و زمانی آن استخراج گردیده و نهایتا از این الگوها پروفایل کاربر ساخته می‌شود. سپس در فاز تشخیص فعالیتهای جدید کاربر با پروفایل های بدست آمده از فاز یادگیری و یا پروفایلهای بروز شده مقایسه می‌شود. در صورت عدم تطابق رفتار جدید با رفتار مجاز کاربر، اخطاری به مدیر سیستم ارسال می گردد. ضمناً در طول زمان کارکرد سیستم از طریق دریافت بازخورد مدیر، پروفایل‌های موجود بروز رسانی می‌شوند. شکل(1) فرایند کلی کار سیستم را نشان میدهد. توجه به این نکته لازم است که در روش-های تشخیص ناهنجاری همواره فرض بر این است که درفاز یادگیری تمام فعالیت‌های کاربر مجاز هستند.



شکل(1): بلوک دیاگرام کلی سیستم

۲-۲ استخراج الگوها

الگوی رفتاری فعالیت هایی است که توسط کاربر در فاز یادگیری تکرار می شوند و قابل استخراج از لاگ فعالیتهای کاربر هستند. این الگوها از دو بعد ترتیبی و زمانی کاویده میشوند که در ادامه هر یک توضیح داده خواهد شد.

۲-۲-۱ کاویدن الگوی رفتار ترتیبی

رفتار ترتیبی نشان دهنده ی رفتار های عادی کاربر است که در کل فاز یادگیری تکرارمی شوند. این الگوهای رفتاری در سه سطح انتزاعی کاویده می شوند: ترتیبی از تراکنش ها، ترتیبی از دستورات و ترتیبی از عملیات.

تعریف1: الگوی تراکنش عبارتست از لیستی منظم از تراکنشهای انجام شده توسط کاربر در یک نشست، با حداقل آستانه‌ی X.

تعریف2: الگوی دستور عبارتست از لیستی منظم از دستورات INSERT، DELETE، SELECT، UPDATE که از تراکنشهای کاربر با حداقل آستانه ی Y استخراج شده اند.

تعریف3: الگوی عملیات عبارتست از لیستی منظم از عمل های Read و Write اجرا شده در یک دستور[9] با حداقل آستانه ی Z.

نمونه ای از تراکنش ها، دستورات و عملیات در جدول ۱ نشان داده شده است که در آن T و C به ترتیب نشان دهنده ی تراکنش و دستور؛ U,S,D,I به ترتیب نشان دهنده ی دستورات INSERT، DELETE، SELECT، UPDATE و W,R نشان دهنده ی عمل های Read و Write هستند. بدون از دست رفتن جامعیت، از اعداد صحیح برای نشان دادن فقره داده های پایگاه استفاده میکنیم. جدول (۲) حاوی نمونه ای از نشست ها و تراکنشهای انجام گرفته توسط یک کاربر است.

جدول(۱): نمونه ای از تراکنش ها، دستورات و عملیات کاربر در یک نشست

T	C	Operation	T	C	Operation
T1	I	W(1),W(3),W(7)	T4	D	R(3),W(2),W(5)
	I	W(4),W(3),W(5)		D	R(1),W(5)
	U	R(4),R(5),W(3),W(1)		I	W(2),W(3), W(4)
	I	W(2),W(4)		U	R(5), W(3),W(1)
				U	W(3), W(6) ,W(5)
T2	S	R(1),R(4), R(1),R(3)	T5	S	R(4),R(7)
	S	R(3)			
	S	R(5), R(2),R(4)		S	R(1),R(3),R(7)
	I	W(4),W(1),R(3)			
	I	W(7),W(2),W(4), W(6),W(5)		I	W(3),W(6)
	S	R(7), R(2),R(1),R(4), R(3)			
T3	S	R(4),R(5)	T6	I	W(2),W(5)
	S	R(4),R(1),R(3)			

الگوهای تراکنش ها، دستورات و عملیات در فرم الگوهایی ترتیبی هستند و به همین دلیل برای استخراج آنها می توان از تکنیکهای داده کاوی ترتیبی[۵] استفاده کرد. تعداد قابل توجهی الگوریتم داده کاوی ترتیبی موجود است که می توان از میان آنها به الگوریتمهای SPAM [2]و AprioriAll[1] ,PerfixSpan[16] ,GSP[17] اشاره کرد. که ما در فاز آزمایش جهت کاهش فضای جستجو در فرآیند استخراج الگوها از الگوریتم SPAM بهره گرفته ایم.

با استفاده از الگوریتم‌های داده‌کاوی ترتیبی وبا توجه به تعاریف انجام شده، (به عنوان مثال) الگوی دستور <S,S,I> و الگوی تراکنش <T3,T5,T4> از جداول (۱) و (۲) قابل استخراج هستند و برای دستور Delete الگوی عملیات <R(3), W(2),W(5)> استخراج می‌شود. همچنین نمونه‌های دیگری از الگوهای عملیات استخراج شده برای هر دستور در جدول(۳) نمایش داده شده‌اند.

جدول(۲): نمونه ای از نشست های یک کاربر

Session	Transaction
S2	T1,T2,T3,T5,T4,T6,T7,T8
S3	T2
S4	T6,T5,T4
S5	T3,T5,T4,T1
S6	T2,T5,T4,T1,T3,T7,T5,T7,T5

جدول(۳): نمونه ای از الگوهای ترتیبی عملیات

دستور	حداقل آستانه	الگوی عملیات کاویده شده
Insert	> =20%	< W(2),W(5)>
Delete	> =50%	< R(3),W(2),W(5)>
Update	> =30%	< R(5),W(3)>
Select	> =20%	< R(4), R(1),R(3)>

۲-۲-۲ کاویدن الگوی رفتاری زمانی

تکنیک کاویدن الگوی رفتاری ترتیبی به تنهایی نمی‌تواند تمامی جنبه های رفتار کاربر را استخراج نماید. در عمل بسیاری از الگوهای مفید رفتار کاربر تنها در یک بازه‌ی زمانی مشخص(برای مثال هر جمعه صبح)، تکرار می شوند نه در زمانهای دیگر. چنین الگوهایی را "الگوی رفتار زمانی" می نامیم که در سه سطح انتزاع تراکنش، دستور و عمل بطور متناوب در بازه هایی رخ می دهند. برای نمایش الگوی رفتار مفهوم Time-signature بصورت زیر استفاده می نماییم:

تعریف۴: $TS_t = (D_1T_1, D_2T_2, ...D_nT_n)$ نشان‌دهنده ی یک Time-signature است که در آن D_i یک واحد زمان است مانند سال، ماه، هفته، روز و ساعت. همچنین T_i یک بازه ی زمانی بسته است که حاوی تمامی مقادیر معتبر زمانی D_i می باشد. برای مثال یک Time-signature می تواند بصورت زیر باشد:

(سال {۱۳۸۶-۱۳۸۰},ماه {۵-۳},روز {۳۱-۱}) یا (*,*,روز {۳-۵},ساعت {۱۵-۹}).

می‌توان برای توصیف مفاهیم زمانی روزمره از بازه های زمانی مانند "صبح" یا "بعد از ظهر" استفاده می کنیم.

تعریف۵: الگوی زمانی بر مبنای محدودیت زمانی T عبارتست از زوج مرتب <BP,TS> که BP همان الگوی رفتار و TS نشان دهنده ی Time-signature آن می باشد. در اینجا BP که متناوباً در زمان های TS و در محدودیت زمانی T تکرار می شود شامل الگو های تراکنش، دستور و عمل می باشد.

روشهای بسیاری برای استخراج الگوهای ترتیبی[1,17]، قواعد وابستگی زمانی[11]، الگوهای ترتیبی چندبعدی[12] و الگوهای متناوب دسترسی وب[19] ارائه شده اند. از آنجا که اکثرکارهای قبلی

به طور کامل مفهوم الگوی زمانی ارائه شده در این مقاله را پوشش نمیدهند برای استخراج این الگوها ما الگوریتم TBSP را ارائه می نماییم که در آن محدودیت زمانی در فرمت Time-signature، حداقل آستانه و لاگ فعالیت‌ها به عنوان ورودی داده می‌شود و الگوی رفتار زمانی کاربر به عنوان خروجی بدست می‌آید. الگوریتم ارائه شده بدین شرح است:

Input: 1-Time-signature **TS**: indicate periodic time constraint
2-Minsupport X: support thresholds
3-Audit log: include Sessions (s_{DB}), Transactions (T_{DB}),Commands (C_{DB}) and Operations (O_{DB}) with their Time (t_i)

Output: The set of temporal behavior sequence pattern over **TS**

Method:

1: Create empty sets of Transaction-TBSP, Command-TBSP, Operation-TBSP and TBSP;
2: For each $S_i \in S_{DB}$,
For each $< T_j, t_w > \in S_i$,
if t_w is covered by **TS**
insert: T_j into Transaction-TBSP,
C_j into Command-TBSP
O_j into Operation-TBSP;
3: TBSP [] = Generating temporal behavior sequence pattern using existing sequential pattern mining algorithm for Transaction-TBSP, Command-TBSP and Operation-TBSP with minimum support X
4: Return TBSP [];

(1)

درمرحله اول، مجموعه های تهی از transaction-TBSP، command-TBSP، operation-TBSP و TBSP ایجاد می‌گردد. در مرحله دوم مجموعه های زمانی که رفتار عادی کاربردر محدوده زمانی ورودی را نشان میدهد، ساخته می شوند. در سومین مرحله با بکار گیری هر کدام از الگوریتم های کاویدن الگوی ترتیبی روی مجموعه های زمانی بدست آمده از مرحله دوم، الگوهای زمانی مورد نظر قابل تولید هستند. حاصل، الگوهای زمانی در محدوده زمانی داده شده در ورودی است.

در الگوریتم ارائه شده، t_w زمان وقوع تراکنش T_jوهمچنین C_j و O_j به معنای ترتیب دستورات و ترتیب عملیات درتراکنش T_j است.

در جدول (۴) نشست های یک کاربر به همراه تراکنشهای اجرا شده و زمان آنها در آن نشست نشان داده شده است. برای توصیف مفاهیم زمانی روزمره ، زمان اجرای تراکنش را به پنج قسمت مطابق جدول (۵) تقسیم میکنیم.

جدول(۴): نمونه ای از نشست‌های کاربر جهت کاویدن الگوهای رفتاری زمانی

Session	T	Time	Session	T	Time
S9	T3 T6 T2 T4	(Mon,1:14:13) (Mon, 1:16:20) (Mon, 1:16:48) (Tue, 1:19:10)	S12	T5 T1 T3	(Thu, 1:16:48) (thu, 21:52:15) (wed, 1:19:10)
S10	T4 T3 T1 T5	(Fri,1:14:13) (fri, 2:07:20) (sat, 13:25:48) (sun, 11:29:10)	S13	T1 T6 T3	(sat,10:16:14) (sat, 17:30:27) (sun, 11:49:16)
S11	T2 T4	(sat, 15:16:11) (sun, 18:17:59)	S14	T6	(wed, 5:29:07)

جدول(۵): بازه های زمانی تعریف شده

بازه زمانی	محدوده
صبح	(*,[7-12])
عصر	(*,[12-18])
شب	(*,[18-24])
نیمه شب	(*,[24 -7])
تعطیلات	(*,{جمعه})

برای مثال، جدول (۱) و (۴) و محدودیت زمانی ({شنبه,یکشنبه},صبح) به عنوان ورودی به الگوریتم (۱) برای کاویدن الگوی رفتاری زمانی داده می شود. فرض می‌کنیم که میخواهیم الگوی تراکنش، دستور و عملی را که دارای آستانه بیشتر از ۶۰٪ هستند را استخراج کنیم. نتایج به ترتیب شامل <R(5),W(3)> و <I,U> و <T1,T3> می باشد. ناگفته پیداست که برخی الگوهای زمانی تنها با استفاده از تکنیک کاویدن زمانی قابل استخراج هستند. برای مثال در جدول (۶) نمونه هایی از الگوهای زمانی عملیات با آستانه های نشان داده شده کاویده شده اند. بامقایسه‌ی این جدول با جدول (۳) روشن می‌شود که الگوی عمل <W(2),W(4)> برای دستور INSERT با آستانه ی بزرگتر از ۴۰٪ تنها با بهره گیری از کاویدن الگوی زمانی قابل استخراج می باشد.

جدول(۶): نمونه ای از الگوهای زمانی عملیات

دستور	حداقل آستانه	الگوی عملیات کاویده شده
Insert	> =40%	< W(2),W(4)>
Delete	> = 50%	< R(3),W(2),W(5)>
Update	> =60%	< R(5),W(3)>
Select	100%	< R(7)>

۲-۳ تولید و بروزرسانی پروفایل

هر پروفایل دارای پارامترهای شناسه کاربر, الگوهای رفتاری ترتیبی, الگوهای رفتاری زمانی و فیلدهای "معتبر از" و "معتبر تا" می‌باشد که در شکل ۲ نمایش داده شده اند.

شناسه کاربر	الگوهای رفتاری ترتیبی	الگوهای رفتاری زمانی	معتبر از	معتبر تا

شکل(۲) -پارامترهای پروفایل

شناسه کاربر, شناسه واحدی است که هنگام کار با پایگاه داده در لاگ فعالیت‌های او ذخیره می‌شود.الگوهای رفتاری ترتیبی وزمانی الگوهای قابل قبولی از رفتار نرمال کاربر در ۳ سطح انتزاع هستند که به عنوان رفتار مجاز کاربر استخراج می‌شوند.فیلدهای "معتبر از" و "معتبر تا" نیز برای ثبت تغیرات و زمان بروزرسانی ها استفاده میشود.

حال اگرالگوی فعالیت جدیدی در فازتشخیص با توجه به الگوهای مجاز درون پروفایل,ناهنجار شناخته شود به مدیر سیستم اخطاری داده می‌شود.ادمین رفتار مشکوک رابررسی کرده ودر صورت غلط بودن اخطار,با اعلام مجاز بودن فعالیت جدید باعث بروزرسانی پروفایل واضافه شدن الگوی رفتار جدید به الگوهای قبلی می‌شود. همان طور که گفته شد برای ثبت نمودن این تغیرات وبروزرسانی ها از فیلدهای "معتبر از" و "معتبر تا" استفاده می‌نماییم که درفیلد "معتبر تا" مقادیری غیر از ∞ نشان دهنده معتبر نبودن پروفایل است.هم چنین اگر کاربرجدیدی به سیستم اضافه شود که لاگ فعالیتی از اوموجود نباشد می‌توان مقادیری پیش‌فرض را بر اساس سیاست‌های ادمین به پروفایلش اختصاص داد.

۲-۴ مقایسه(تشخیص ناهنجاری)

پس از شکل گیری پروفایل ها می‌توان از آنها در تشخیص رفتار ناهنجار در پایگاه داده ها بهره گرفت. برای این منظور ابتدا هر فعالیت جدیدی از سوی کاربر با الگوی موجود در پروفایل وی مقایسه شده تا میزان شباهت بین آن دو مشخص گردد.

برای تعیین میزان شباهت دو الگو، از تابعی تحت عنوان تابع شباهت سنجی رفتاری[10] استفاده می کنیم که دو الگو را به عنوان ورودی دریافت کرده و میزان شباهت را در خروجی تحویل میدهد.

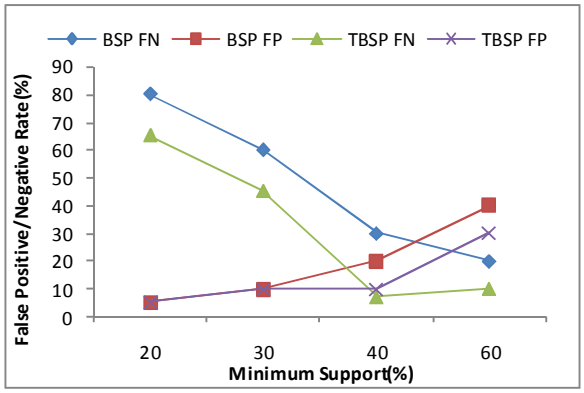
$$\mathrm{BMatch}(BS_1, BS_2) = \left(\frac{\sum_{i=0}^{z-1} SM(i, BS_1, BS_2)}{\sum_{i=0}^{z-1} K^i}\right)^{\frac{1}{z-1}} \qquad (۱)$$

$$\mathrm{SM}(i, BS_i, BS_2) = \begin{cases} SM(i, BS_1, BS_2)*k & i>0 \,\&\, BS_1[i]=BS_2[i] \,\&\, SM(i-1, BS_1, BS_2)=0 \\ 0 & BS_1[i]\neq BS_2[i] \\ 1 & i=0 \,\&\, BS_1[i]=BS_2[i] \\ 1 & i>0 \,\&\, BS_1[i]\neq BS_2[i] \,\&\, SM(i-1, BS_1, BS_2)=0 \end{cases} \qquad (۲)$$

مقدار بازگردانده شده توسط تابع مقداری بین صفر تا یک است. هرچه شباهت دو الگو بیشتر باشد، مقدار خروجی تابع بزرگتر خواهد بود. در تساوی (۱) و (۲) مقدار k یک عدد ثابت دلخواه بین یک تا دو بوده که طبق [10] بهتر است بزرگتر از یک مقدار دهی شود. اگر مقدار شباهت محاسبه شده توسط تابع از مقدار مشخصی که براساس سیاستهای امنیتی سیستم تعیین گردیده کمتر باشد، الگوی مقایسه شده به عنوان یک رفتار ناهنجار شناسایی شده و اخطاری به مدیر سیستم داده می شود. در بخش آزمایشات این مقدارمشخص معادل 0.5 در نظر گرفته شده است.

۳- نتایج آزمایشات

برای آزمون روش ارائه شده دو مجموعه معنا دار پرسش تولید شد .یک مجموعه به عنوان رفتار طبیعی کاربر که تمامی پرسش های موجود در آن معتبر و مجاز تلقی می گردد و مجموعه دیگر شامل پرسیشهایی تصادفی و پرسشهایی شامل حملات رایج مانند SQL Injection که قرار است سیستم تشخیص نفوذ پایگاه داده ها توسط آنها آزمایش شود. جهت ارائه نتایج سنجش از دو معیار رایج False Positive و False Negative استفاده شده است . روش استفاده از الگوی ترتیبی که آن را به اختصار BSP مینامیم و روش استفاده از الگوی ترتیبی و زمانی که TBSP مینامیم با هم مقایسه شده اند و نتایج حاصل در شکل (3) آورده شده اند.



شکل (3): مقایسه FP و FN با بکارگیری الگوهای ترتیبی و زمانی

همانطور که از نمودار قابل دریافت است با دخیل نمودن مفهوم الگوهای زمانی در تشکیل الگوهای رفتاری کاربر شاهد رشد قابل ملاحظه ای در

دقت تشخیص سیستم هستیم. در حوزه FN با کاربست الگوهای زمانی رفتار کاربر حدود ۱۵٪ و در مورد FP نیز، استفاده از الگوهای زمانی منجر به افت حدود ۵٪ میگردد.

۴- نتیجه گیری

روش تشخیص نفوذ پایگاه دادهی ارائه شده شامل دو بخش یادگیری و تشخیص است. در فاز یادگیری پروفایلی از رفتارکاربر ساخته می شود و در فاز تشخیص با استفاده از این پروفایل معین می گردد که آیا رفتار کاربر طبیعی است یا خیر. نکته کلیدی در این میان در نظر گرفتن دسته ای از الگوهای مفید رفتار کاربر است که تنها در بازه های زمانی مشخصی، تکرار می‌شوند وآنها را "الگوهای رفتاری زمانی"نامیدیم. نتایج آزمایشات صورت گرفته کارایی روش ارائه شده را در افزایش دقت شناسایی و کاهش False Positive نشان می دهد.برای کارهای آینده قصد داریم بابکارگیری شبکه های عصبی شناسایی حملات را دقیقتر نماییم.

۵- مراجع

[1] Agrawal, R. and Srikant, R. Mining Sequential Patterns. In Proceedings of the 1995 Int. Conf. Data Engineering, Taipei,Taiwan, March 1995.Pages 3 -14.

[2] Ayres, J., Flannick, J., Gehrke, J., and Yiu, T. "Sequential Pattern Mining using A Bitmap Representation", In ACM SIGKDD Conference, pp.429-435, 2002.

[3] Barbara, D., Goel, R., and Jajodia, S. Mining Malicious Data Corruption with Hidden Markov Models. In Proceedings of the 16th Annual IFIP WG 11.3 Working Conference on Data and Application Security, Cambridge, England, July 2002.

[4] Bertino, E., Kamra, A.,Terzi, E. and Vakali, A. "Intrusion Detection in RBAC-Administered Databases", In Proceedings of Annual Computer Security Applications Conference(ACSAC), 2005

[5] Chen, R.C., An Anomaly Intrusion Detection on Database Operation by Fuzzy ART Neural Network,int.Computer Symposium, Dec 2004

[6] Chung, C., Gertz M., and Levitt, K. DEMIDS: A Misuse Detection System for Database Systems. In Third Annual IFIP TC-11 WG 11.5 Working Conference on Integrity and Internal Control in Information Systems, Kluwer Academic Publishers, pages 159-178, November 1999.

[7] Diaz-Gomez, A., A Genetic Algorithm Approach for Doing Misuse Detection in Audit Trail Files,"IEEE proceedings og the 15th international conference on computing"(2006)

[8] Fawcett, T. and Provost F. Adaptive Fraud Detection. Data Mining and Knowledge Discovery, pages 177-181, 1997

[9] Hu, Y., and Panda, B. A Data Mining Approach for Database Intrusion Detection, Proceedings of the ACM Symposium on Applied Computing, pp. 711-716 (2004).

[10] Hofmoyer, A., Forrest, S. and Somayaji, A. Intrusion Detection Using Sequence of System Calls, Journal of computer Security Vol. 6,1998

[11] Lane, T., and Brodley, C.E. Temporal sequence learning and data reduction for anomaly detection. ACM Trans. Inf. Syst. Secur., 2(3):295–331, 1999.

[12] Lee, C.H. "IMSP: An information theoretic approach for multi-dimensional sequential pattern mining", Applied Intelligence, Vol. 26, No. 3. (June 2007), pp. 231-242

[13] Lee, V., Stankovic, J., Son, S.: Intrusion detection in real-time databases via time signatures. In: Proceedings of the IEEE Real- Time Technology and Applications Symposium (RTAS) (2000)

[14] Lee, S.Y., Low, W.L., Wong, P.Y. Learning fingerprints for a data- base intrusion detection system. In: ESORICS '02: Proceedings of the 7th European Symposium on Research in Computer Security London. pp. 264–280, Springer-Heidelburg (2002)

[15] Lunt, T., Tamaru, A., Gilham, F., Jagannathan, R., Neumann,P.,Javitz, H., Valdes, A., Garvey, T.: A real-time intrusion detection expert system (ides)—final technical report. Technical Report, Computer Science Laboratory, SRI International (1992)

[16] Pei, J., Han, J., Pinto, H., Chen, Q, Dayal, U., and Hsu, M-C. PrefixSpan: Mining Sequential Patterns Efficiently by Prefix-Projected Pattern Growth. In Proceeding of 2001 International Conference on Data Engineering (ICDE'01),Heidelberg, Germany, April 2001.

[17] Srikant, R. and Agrawal, R. Mining Sequential Patterns: Generalizations and Performance Improvements. In EDBT, pages 3–17, March 1996

[18] Zhong, Y., and Qin, Y. Research on Algorithm of User Query Frequent Itemsets Mining, Machine Learning Cybernetics, pp. 1671-1676 (2004).

[19] Zhou, B.Y., Hui, S.C. and Fong, A.C.M., "An Efficient Approach for Mining Sequential Access Patterns", 8th Pacific Rim International Conference on Artificial Intelligence (2004), Auckland, New Zealand, LNAI 3157, pp.485-493.

[1] Data Mining
[2] Behavior Sequential Pattern
[3] Temporal Behavior Pattern
[4] Anomalies
[5] Sequence Mining