تشخیص ناهنجاری روی وب از طریق ایجاد پروفایل کاربرد دسترسی

مریم السادات میرهادی تفرشی* و رضا عزمی
گروه کامپیوتر، دانشکده فنی مهندسی، دانشگاه الزهرا، تهران، ایران

چکیده

در پژوهش پیش رو با تمرکز روی شناسایی پیمایش‌های ناهنجار وب، سعی شده است تا از طریق مقایسه پروفایل‌های کاربرد وب با نشست فعلی کاربر رفتارهای بدخواهانه، مورد شناسایی قرار گیرند. در رویکرد پیشنهادی، ابتدا پروفایل های کاربرد وب از لاگ دسترسی وب سرور استخراج می‌شود؛ سپس با محاسبه شباهت هر نشست ورودی کاربر به پروفایل‌های اصلی و استخراج هشدارهای کنترل دسترسی متناظر با همان نشست یک شبکه عصبی فازی جهت تشخیص هنجار یا ناهنجاربودن پیمایش کاربر مورد استفاده قرار می‌گیرد. به دلیل فقدان داده استانداردی که هم شامل پیمایش‌های وب صفحات و هم شامل هشدارهای کنترل دسترسی متناظر با آن باشد، رویکردی نیز به منظور شبیه‌سازی پیمایش‌های یک کاربر عادی ارائه شد. ارزیابی‌های صورت گرفته نشان می‌دهد که روش ارائه‌شده در تشخیص پیمایش‌های ناهنجار توانمند عمل می‌کند.

واژگان کلیدی: پروفایل کاربرد وب، کنترل دسترسی، شبکه عصبی فازی، شناسایی حملات وب، تشخیص ناهنجاری وب

Web Anomaly Detection by Using Access Log Usage Profile

Maryam Sadat Mirhadi Tafreshi* & Reza Azmi
Department of Computer, Faculty of Engineering, University of Alzahra, Tehran, Iran

Abstract

Due to increasing in cyber-attacks, the need for web servers attack detection technique has drawn attentions today. Unfortunately, many available security solutions are inefficient in identifying web-based attacks.

The main aim of this study is to detect abnormal web navigations based on web usage profiles. In this paper, comparing scrolling behavior of a normal user with an attacker, and simultaneous use of the access control policy alarms provided in web pages crawling with high access level, leads to an attacker to be detected among ordinary users. Indeed, the proposed method in this research includes two main steps: firstly web usage profiles are extracted as web main patterns of users' behavior. In order to cluster similar web sessions we used a system inspired by artificial immune system. In the employed method, the rate at which a particular web page is visited as well as the time a user spends on the pages, is calculated so as to estimate how interesting a specific page is in a user's session. Therefore, the similarity in the web page is defined based on the combination of the similarity of web pages URLs and that of the users' level of interest in visiting them. Secondly, the difference between each current user session from the main profiles is calculated. Additionally, the access control logs are derived from corresponding sessions in this stage. Regarding the noisy nature of web server logs, a method was required so that a slight change in the data would not make a noticeable change in the results validity. Hence, a fuzzy neural network has been applied to distinguish normal and abnormal scrolling behavior in second step.

Due to the lack of a standard data that contains both web pages scrolling and access control logs corresponding to it, providing such a data was required. At first, those intended logs were produced. To do so, an Apache web server was run on the platform of a Centos machine. In order to create the logs

* نویسندهٔ عهده‌دار مکاتبات

* Corresponding author

completely similar to a real server's log, an e-commerce website was set up on Apache server. This website had about 160 different web pages to be visited by different users. At this point, a novel method is proposed to simulate the behavior of web users when they visit a website. Likewise, the abnormal data was generated by means of a large number of existing attack tools. It should also be noted that the access control policy has been used is SELinux and It has been added to Linux kernel.

As mentioned, web server access log varies greatly with changing user behaviors, the stability of the proposed method against noise should be evaluated. For this reason, the results has been investigated on noisy profiles created by making random changes on the main profiles, and only the testing phase is conducted again. Subsequently, the distance from the profiles having noise is compared with the main ones. To demonstrate the ability of this method, the results have been compared with a Support Vector Machine (SVM). The carried out evaluations show that our approach performs efficiently in identifying normal and abnormal scrolling.

Key words: fuzzy neural networks, web usage profile, anomaly detection, access control

۱- مقدمه

امروزه به‌دلیل رشد جرایم رایانه‌ای نیاز به رویکردهایی جهت امن‌سازی و حفاظت از سرورها و برنامه‌های تحت وب در برابر حملات بدخواهانه، مورد توجه قرارگرفته است. متأسفانه بسیاری از رویکردهای امنیتی موجود، در شناسایی حملات تحت وب ناکارآمدند. در بسیاری حملات وب تلاش برای حدس‌زدن رمز عبور، تزریق داده در پایگاه داده یا مواردی از این دست مانند انواع حملات injection، حملات brute force و غیره موجب آسیب جدی به وب‌سرور می‌شود. به نظر می‌رسد دسترسی به برخی صفحات مانند صفحات دسترسی و یا صفحاتی که به نحوی داده را بین کاربر وب و سرور مبادله می‌کنند و یا صفحاتی با سطح دسترسی مدیر سامانه که برای دسترسی کاربران طراحی نشده‌اند، از اهمیت بیشتری برخوردار است و تکرار دسترسی به این نوع از صفحات در مسیر پیمایش اهمیت بیشتری دارد. با استفاده از کنترل دسترسی به‌عنوان یک ویژگی مستقل و با تعریف سیاست‌های کنترل دسترسی مشخص الگوی دسترسی به چنین صفحاتی را می‌توان به‌دست آورد.

در این پژوهش به‌منظور تشخیص ناهنجاری به مدل‌سازی رفتارهای نرمال وب پرداخته می‌شود. به عبارت دیگر، مقایسه رفتار پیمایشی یک کاربر عادی در مقایسه با یک مهاجم و به‌کارگیری همزمان هشدارهای تولیدشده در پیمایش صفحات با سطح دسترسی بالا منجر به تشخیص یک مهاجم از کاربران عادی صفحات وب می‌شود. مقایسه رفتار پیمایشی کاربران بر اساس میزان شباهت الگوی دسترسی به صفحات با پروفایل‌های اصلی کاربرد وب استخراج می‌شود و منظور از هشدار، تعداد اخطارهای ثبت‌شده در لاگ کنترل دسترسی است که به‌عنوان یک ویژگی مستقل از نوع پیمایش در تشخیص ناهنجاری دخیل شده‌اند.

به‌دلیل دردسترس‌نبودن داده استاندارد در زمینه پیمایش‌های وب، که شامل حملات وب نیز باشد، در این پژوهش به تولید داده حقیقی جهت به‌کارگیری روش پیشنهادی پرداخته شده است. ارزیابی‌های صورت‌گرفته نشان می‌دهد که روش ارائه‌شده در تشخیص پیمایش‌های ناهنجار بسیار توانمند عمل می‌کند. در ادامه و در بخش دو به بیان پژوهش‌های انجام‌شده پرداخته و در بخش سوم مفاهیم اصلی مورد استفاده در این پژوهش بیان می‌شود. در بخش چهارم روش پیشنهادی به‌تفصیل آورده و درنهایت در بخش پنجم ارزیابی عملکرد روش مورد پیشنهادی مطرح شده است.

۲- کارهای مرتبط

سامانه‌های تشخیص نفوذ از دو رویکرد اصلی در تشخیص حملات استفاده می‌کنند. دسته نخست بر اساس مجموعه‌ای از قوانین یا بر اساس هر آنچه که به‌عنوان رفتار نرمال شناخته می‌شود، عمل می‌کنند. در این بخش از توضیح تمامی پژوهش‌های انجام‌شده در زمینه تشخیص ناهنجاری اجتناب شده و تنها به بیان مواردی اکتفا می‌شود که روی تشخیص ناهنجاری وب متمرکز بوده‌اند. رویکرد [39] سامانهٔ تشخیص نفوذی را که بر اساس پارامترها و مقادیری که به سرور منتقل می‌کنند، ارزیابی و تحلیل می‌کند. البته این سامانه‌ها قادر به شناسایی حملاتی که از قبل مدل‌سازی نشده‌اند، نیست.

در [41] بر اساس سامانه‌های ایمنی، مدلی برای شناسایی حملات وب ارائه شد. مدل پیشنهادی با نام IADMW، بر اساس انتخاب کلون شکل گرفته بود.

درخواست‌های وب به‌عنوان آنتی‌ژن‌ها در نظر گرفته شده و توسط یک بردار ویژگی استخراج‌شده از کل درخواست نمایش داده می‌شدند. در این روش وزن اختصاص‌داده‌شده به هر ویژگی نشان‌دهندهٔ اهمیت آن ویژگی در بردار مشخصات است.

سامانهٔ ایمنی دسته‌بندی وب (WCIS)، با تمرکز بر تشخیص خودی و غیرخودی و استفاده از الگوریتم انتخاب منفی به بررسی درخواست‌های وب می‌پردازد [40]. این سامانه، طول URI، تعداد متغیرها و توزیع کاراکترها را به‌عنوان ویژگی در نظر می‌گیرد.

هر دو مدل IADMW و WCIS، بدون درنظرگرفتن نشست‌های وب به یک درخواست برچسب حمله می‌زدند. در پژوهش [43] مدلی بر اساس شبکه‌های ایمنی ارائه شد که از طریق کاهش ویژگی‌های ترافیک نرخ شناسایی و تشخیص حملات را بهبود بخشید. همچنین عملی‌بودن روش ایمنی مصنوعی اعمال‌شده در دسته‌بندی انواع حملات را مورد بررسی قرار داد.

[44] روشی ارائه کرد که با استفاده از مدل مارکف حملات برنامه‌های وب را شناسایی کند. این روش بر اساس پایش درخواست‌های وب واردشده عمل می‌کند. پژوهش‌گران در [49] از مدلی آماری جهت تحلیل داده جمع‌آوری‌شده از وب‌سرورها بهره می‌برند. نقطه‌ضعف رویکرد پیشنهادی وی بررسی یکی یکی درخواست‌های HTTP و ندیده‌گرفتن ارتباط معنایی میان درخواست‌های مختلف است. از آنجا که پروتکل HTTP بدون وضعیت است، درخواست‌های وب به‌تنهایی ارتباطی بین درخواست HTTP و رفتار دسترسی کاربر نمی‌توانند به‌دست دهند.

در [45] روشی ارائه شده است که ارتباط میان توالی درخواست‌ها و رفتار دسترسی کاربر را در مدل نشست‌های دسترسی کاربر در نظر می‌گیرد. در این روش درخواست‌های وب به نشست‌های وب شکسته می‌شوند تا الگوی مورد نظر میان درخواست‌ها شناسایی شود؛ و بر اساس تحلیل رفتار حملات وب، این مدل حملات وب را شناسایی می‌کند. عزمی و همکارانش در [38] رویکرد تشخیص ناهنجاری بر اساس روش ایمنی مصنوعی شکل گرفته است. در این روش درخواست‌های به‌دست‌آمده از لاگ‌های پیش‌پردازش‌شده به‌عنوان آنتی‌ژن در نظر گرفته می‌شوند. شبکه B-Cellها نمایانگر نسخه خلاصه‌شده‌ای از آنتی‌ژن‌هایی است که در شبکه دیده شده‌اند. AIS پیشنهادی نشست‌های وب را به خوشه‌های هنجار و ناهنجار دسته‌بندی می‌کند.

در [8] محاسبات آماری در نشست‌های وب انجام می‌گیرد که شامل تعداد درخواست‌ها در یک بازه زمانی خاص، طول درخواست‌ها در نشست‌ها، و طول هر نشست است. هر یک از این ویژگی‌ها با توجه به رفتارهای نرمال یادگیری می‌شوند و میزان این آماره‌ها در نشست‌های وب با مقادیر نرمال مقایسه شده و بدین ترتیب حملات تشخیص داده می‌شوند.

۳- مفاهیم اصلی

در این بخش مفاهیم اصلی و مهم که در درک بهتر روش پیشنهادی موثرند ارائه شده است. ابتدا مفاهیم مربوط به کنترل دسترسی و روش‌های مختلف آن و سپس مفاهیم مربوط به استخراج پروفایل‌های کاربرد وب از پیمایش‌های کاربران بیان خواهد شد.

۳-۱- مفاهیم کنترل دسترسی

هدف اساسی هر سازوکار کنترل دسترسی، فراهم‌کردن یک سامانهٔ قابل بررسی به‌منظور تضمین حفاظت از اطلاعات در برابر دسترسی‌های احراز هویت نشده و نامشخص است که مانند یک یا چند سیاست امنیتی از قبل تعیین شده باشد و شامل کنترل‌های زیر می‌شود [1]:

- محرمانگی[1]: کنترل افشای اطلاعات
- یکپارچگی[2]: کنترل تغییرات اطلاعات
- دسترس‌پذیری[3]

مدل‌های ارائه‌شده تاکنون را به سه دسته MAC، DAC و RBAC می‌توان تقسیم‌بندی کرد.

۳-۱-۱- کنترل دسترسی اجباری MAC

به هر نوع از مدل‌های کنترل دسترسی که سیاست‌های امنیتی را مستقل از عملیات کاربران اجرا می‌کند، کنترل دسترسی اجباری می‌گویند. تخصیص و اجرای برچسب‌های امنیتی توسط سامانه و تحت مدل MAC محدودیت‌هایی را روی فعالیت‌های کاربران اعمال می‌کند؛ درحالی‌که پیوستن به سیاست‌های امنیتی، مانع از اعمال تغییرات پویا در سیاست های زمینه‌ای می‌شود [46].

۳-۱-۲- کنترل دسترسی اختیاری (DAC)

گام نخست در توسعه سامانه کنترل دسترسی منطبق بر این مدل، شناسایی این سه مفهوم است: ۱) اشیایی که قرار است

[1] Confidentiality
[2] Integrity
[3] Availability

محافظت شوند. ۲) موضوعاتی که قرار است روی اشیا فعالیت‌هایی انجام دهند. ۳) عملیاتی که می‌توانند روی اشیا اجرا شوند و باید کنترل شوند. اشیا، موضوعات و عملیات ساماه‌های مختلف ممکن است متفاوت باشند. درواقع مدل DAC، به‌منظور پیاده‌سازی ماتریس‌های کنترل دسترسی توسعه داده شد [48]. در ماتریس دسترسی، وضعیت سامانه با سه تایی (S,O,A) ارائه می‌شود که در آن S، مجموعه‌ای از موضوعات، O مجموعه‌ای از اشیا و A ماتریس دسترسی سامانه است. درایه‌های این ماتریس تعیین می‌کنند که موضوع S بر شی O چه اختیاراتی دارد.

برخلاف چهارچوب MAC که در آن تصمیم‌گیری برای اجازه‌دادن یا ردکردن دسترسی توسط سامانه و بر اساس سیاست‌های از پیش تعریف شده اتخاذ می‌شود، DAC به موضوعات اختیار تصمیم‌گیری درباره دسترسی به اشیا تحت مالکیت خودشان را می‌دهد.

سیاست‌ها و مدل‌های اختیاری را در مقابل فرآیندهایی که برای اجرای برنامه‌های بداندیش[1] خود، از اختیارات کاربر مربوطه استفاده می‌کنند، آسیب‌پذیر می‌کند. به عبارتی دیگر مدل اختیاری را به وسیله اجرای برنامه Trojan به‌صورت توکار[2] می‌توان دور زد.

۳-۱-۳- کنترل دسترسی مبتنی بر نقش[3] RBAC

سیاست کنترل دسترسیِ اختیاری مشکلاتی مانند افشای اطلاعات و عدم ضمانت جامعیت دارد. سیاست دسترسیِ اجباری نیز که یک مدل امنیتی چندسطحی است، از نظر پیچیدگی و دست‌وپاگیربودن مطلوبیت و کاربرد فراگیری ندارد[48].

مدل مبتنی بر نقش یک مدل غیراختیاری است و درحقیقت گونه‌ای از سیاست کنترل دسترسیِ اجباری یا MAC به‌شمار می‌آید، با این تفاوت که بر اساس نیازمندی‌های امنیتی چندسطحی نیست. در مدل کنترل دسترسی مبتنی بر نقش، نقش مجموعه‌ای از تراکنش‌هایی است که کاربر یا گروهی از کاربران در سازمان خود می‌توانند انجام دهند. منظور از تراکنش دسترسی به یک مخزن داده یا تابع گذار است. این تراکنش‌ها توسط مدیر سامانه به نقش‌ها اختصاص داده می‌شوند.

مهمترین مزیت مدل مبتنی بر نقش به‌عنوان یک مدل غیراختیاری، سهولت در استفاده و قدرت در مدیریت مقیاس بالای اشیا و موضوعات است .این مدل نه پیچیدگی سطح‌بندی در سیاست اجباری را دارد و نه مبنای غیر واقعی ارتباط مالکیت بین شیء و ماهیت در سیاست اختیاری است.

۲-۳- استخراج پروفایل کاربرد وب

در این قسمت رویکرد استفاده‌شده جهت استخراج پروفایل‌های اصلی از نشست‌های وب آورده شده است. روش به‌کارگرفته‌شده به‌منظور تولید پروفایل‌ها به‌طورتقریبی مشابه کار انجام شده توسط عزمی و عظیم‌پور [32] است. در اینجا تنها به ارائه بخشی از روش مذکور که مورد استفاده قرار گرفته است، بسنده خواهد شد. در این پژوهش از روش مذکور به‌منظور جداسازی الگوهای اصلی پیمایش وب استفاده شده است. درواقع، پروفایل‌های اصلی با استفاده از این روش استخراج شده‌اند که به‌عنوان الگوهای پیمایش اصلی در محاسبه میزان شباهت مورد استفاده قرار می‌گیرند. رویکرد استفاده شده جهت خوشه‌بندی نشست‌های وب مشابه، سامانه‌ای الهام‌گرفته‌شده از سامانهٔ ایمنی مصنوعی است. روش‌های ایمنی مصنوعی از مهم‌ترین الگوریتم‌های فراابتکاری[4] به‌منظور حل مسائل پیچیده هستند و یکی از مهم‌ترین کاربردهای این نوع الگوریتم‌ها را خوشه‌بندی داده‌ها می‌توان بیان کرد [47]. در این پژوهش نیز به‌منظور دسته‌بندی داده‌های وب از روشی بر مبنای همین الگوریتم‌ها استفاده شده است. در این روش ترکیبی از سلول‌های B وزن‌دار پویا[5] D_W_Bcells خلاصه‌ای از مدل یادگیری‌شده را نمایش می‌دهند؛ به‌علاوه تعاملاتی از نوع سرکوب و تحریک نیز میان DWB-cellsها در جریان است. فرآیند یادگیری با ارائه مجموعه داده ورودی (آنتی‌ژن‌ها) به شبکه‌ای از سلول‌های B آغاز می‌شود. سامانه، شبکه بهینه‌ای از DWB cellهای متصل بهم را با عملیات cloning یاد می‌گیرد. هر DWB cell نمایه یک الگوی یادگیری‌شده است که با یک آنتی‌ژن یا DWB cell دیگر مطابقت دارد؛ به‌علاوه هر DWB cell یک ناحیه تأثیر دارد که با تابع وزن توصیف می‌شود. مقدار این تابع وزن با افزایش فاصله از آنتی‌ژن و زمانی که آنتی‌ژن به شبکه ارائه شده است، کاهش می‌یابد. قدرت اتصالات میان DWB cellها رابطه مستقیم با میزان شباهت آن‌ها دارد. تابع فعال‌سازی i امین، D-W-B-cell توسط j امین آنتی‌ژن در شبکه پس از ارائه J آنتی‌ژن به شبکه در رابطه (۱) آورده شده است. در این رابطه d_{ij}^2

[1] Malware
[2] built in
[3] Role Based Access Control
[4] Meta-heuristic
[5] Dynamic weigted Bcell

فاصله آنتی‌ژن j ام تا D-W-B-cell، i ام است و σ_{ij}^2 نیز ضریبی است که اندازهٔ ناحیه تأثیر پیرامون یک خوشه را تعیین می‌کند، و ثابت τ نرخ فراموشی در شبکه ایمنی است.

$$w_{ij} = e^{-\left(\frac{{d_{ij}}^2}{2{\sigma_{i,j}}^2}+\frac{(J-j)}{\tau}\right)} \qquad (۱)$$

سطح تحریک یک D-W-B-cell پس از ارائه J آنتی‌ژن به شبکه، و ضریب بهینه i امین D-W-B-cell به‌ترتیب بر اساس رابطه (۲) و (۳) است.

$$\sigma_{i,J}^2 = \frac{\sum_{j=1}^{J} w_{ij}{d_{ij}}^2}{2\sum_{j=1}^{J} w_{ij}} \qquad (۲)$$

$$\delta_{iJ} = \frac{\sum_{j=1}^{J} w_{ij}}{\sigma_{i,J}^2} + \alpha(t)\frac{\sum_{l=1}^{N_B^j} w_{il}}{\sigma_{i,J}^2} - X.\beta \qquad (۳)$$

در صورتی‌که یک آنتی‌ژن به اندازه کافی یک B-cell را تحریک کند ($w_{ij} \geq w_{min}$)، عمر این B-cell با عدد صفر به‌روزرسانی می‌شود. در غیر اینصورت عمر B-cell یک واحد افزایش می‌یابد. با کاهش عمر B-cell ضرایب افزایش یافته و درنتیجه B-cellهایی که به‌تازگی فعال شده‌اند، تأثیر بیشتری بر شبکه دارند. با استفاده از الگوریتم خوشه‌بندی KNN به‌طور متناوب، B-cellهای مشابه دسته‌بندی و درنهایت محاسبه تحریک و مقیاس تغییر مقادیر از روابط ۴و۵ محاسبه می‌شود. در این روابط N_B^j تعداد سلول‌های B مجموعه کاری است که در حضور j امین آنتی‌ژن و در KNN خود فعال شده‌اند.

$$\delta_{iJ} = \frac{\sum_{j=1}^{J} w_{ij}}{\sigma_{i,J}^2} + \alpha(t)\frac{\sum_{l=1}^{N_B^j} w_{il}}{\sigma_{i,J}^2} - X.\beta(t)\frac{\sum_{l=1}^{N_B^j} w_{il}}{\sigma_{i,J}^2} \qquad (۴)$$

$$\sigma_{i,J}^2 = \frac{e^{\frac{1}{\tau}\sigma_{i,J-1}^2 w_{i,J-1}+w_{iJ}{d_{iJ}}^2+\alpha(t)\sum_{l=1}^{N_B^j} w_{il}{d_{il}}^2 - X.\beta(t)\sum_{l=1}^{N_B^j} w_{il}{d_{il}}^2}}{2\left(e^{\frac{1}{\tau}w_{i,J-1}+w_{iJ}+\alpha(t)\sum_{l=1}^{N_B^j} w_{il} - X.\beta(t)\sum_{l=1}^{N_B^j} w_{il}}\right)} \qquad (۵)$$

شکل (۱) گام‌های الگوریتم ایجاد پروفایل کاربرد وب را نشان می‌دهد.

با توجه به ماهیت روش مذکور ورود داده جدید (لاگ‌های جدید دسترسی کاربران) نیازی به اعمال مجدد الگوریتم به کل داده موجود نیست و تنها لازم است تا ورودی جدید به‌عنوان آنتی‌ژن جدید به الگوریتم تزریق شود تا با سایر پروفایل‌های اصلی مقایسه شود. بنابراین به‌روزرسانی روش برای دسترسی‌های جدید به سرعت و با افزودن داده جدید به ورودی الگوریتم صورت می‌گیرد.

گام های الگوریتم ایجاد پروفایل‌های کاربرد وب

۱- N_g را حداکثر اندازه جمعیت قرار دهید
۲- جمعیت D-W-B-cell و $\sigma_i^2 = \sigma_{ave}$ را با استفاده از یک آنتی ژن رندم مقداردهی کنید.
۳- گام های زیر برای هر آنتی ژن تکرار شود:
۱- ارائه آنتی ژن به هر D-W-B-cell
۲- در صورت فعال شدن آنتی ژن D-W-B-cell ($w_{ij} \geq w_{min}$)
i. عمر با مقدار ($t = 0$) بروز رسانی شود.
ii. D-W-B-cell جاری وKNN آن برای فعال کردن زیر شبکه افزوده شود.
۳- در غیر اینصورت
i. عمر D-W-B-cell یک واحد اضافه شود.
۴- اگر برای تمام D-W-B-cell ها ($w_{ij} \leq w_{min}$) باشد
i. یک D-W-B-cell جدید برابر با آنتی ژن ایجاد کن(D-W-B-cell = antigen)
۵- در غیر اینصورت
i. برای هر در زیر شبکه فعال گام های زیر تکرار شود
۱- تحریک D-W-B-cell با استفاده از رابطه(۴) محاسبه شود
۲- D-W-B-cell σ_i^2 با استفاده از رابطه(۵) بروزرسانی شود.
۶- انجام عمل تولید مثل (clonning) بر اساس سطح تحریک D-W-B-cells
۷- اگر سایز جمعیت بیشتر از N_g باشد
i. مازاد حداقل D-W-B-cell های تحریک شده، حذف شود.

(شکل-۱): الگوریتم تولید پروفایل کابرد وب
(Figure-1): Web usage profile algorithm

۳-۳- معیار اندازه‌گیری شباهت میان نشست‌های وب

در روش استفاده‌شده علاوه‌بر زمان سپری‌شده توسط یک کاربر در صفحه وب، نرخ مورد مشاهده قرارگرفتن یک صفحه خاص، به‌منظور تخمین میزان مورد علاقه‌بودن آن صفحه نیز در یک نشست کاربر نیز محاسبه می‌شود. بنابراین شباهت در صفحه وب براساس ترکیب شباهت URLهای صفحات وب و شباهت میزان علاقه کاربران به مشاهده آنها تعریف می‌شود.

با درنظرگرفتن شباهت میان دو صفحه بر اساس URLها (Token_Sim) و علاقه‌مندی کاربران به آن‌ها (Interest_Sim)، شباهت دو صفحه از رابطه (۶) به‌دست می‌آید. در این رابطه مقدار ثابت α ضریب کمتری است که از ۲ تا ۱ مقدار می‌گیرد[32]

$$Similarity(P_i, P_j) = \alpha.Token\_Sim(P_i, P_j) + (1-\alpha).Interest\_sim(P_i, P_j) \qquad (۶)$$

با استفاده از روش توالی دنباله‌ها (SA[1]) [29].و یافتن بهترین تطابق میان دو دنباله شباهت دو نشست وب محاسبه می‌شود. تابع امتیازدهی مورد استفاده نیز، مطابق رابطه (۷) است. همان‌طورکه دراین رابطه نیز دیده می‌شود، به‌ازای هر تطابق کامل (دو صفحه با میزان شباهت ۱)، امتیاز شباهت

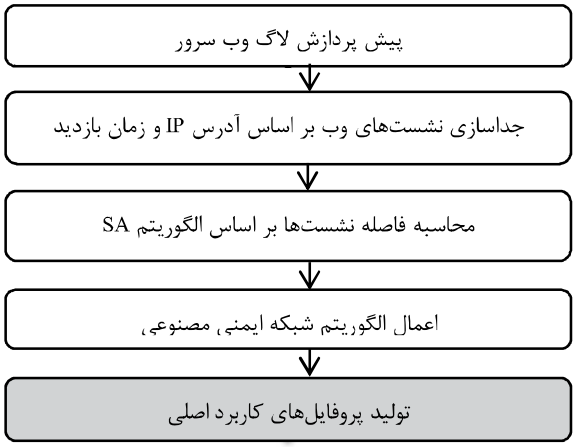
[1] Sequence Alignment

عدد بیست بوده و برای هر عدم تطابق (دو صفحه با میزان شباهت ۰ شباهت با عدد ۱۰- امتیاز دهی می‌شود. برای صفحاتی با میزان شباهت بین صفر و یک، امتیازدهی مابین ۱۰- و ۲۰ است. همان‌طورکه اشاره شد روش هم‌ترازی دنباله‌ها از رویکرد برنامه‌نویسی پویا استفاده می‌کند. در شکل روند تولید پروفایل‌های اصلی آورده شده است.

$$Score(P_i,P_j) = -10 + \left(30 \times Similarity(P_i,P_j)\right) \quad (۷)$$

$$Similarity(S_i,S_j) = \frac{Optimal\_SA\_Score-(-10\times L_l)}{(20\times L_s)-(-10\times L_l)} \quad (۸)$$

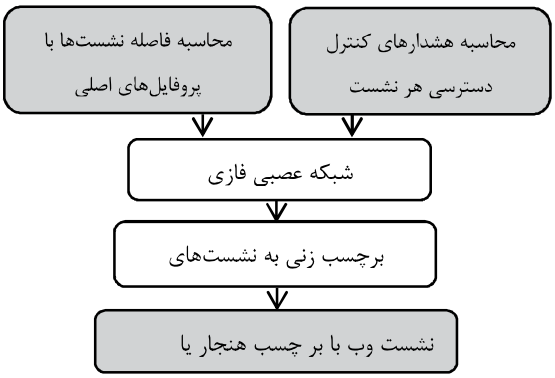
شکل (۲) گام‌های استخراج پروفایل کاربردی وب از لاگ دسترسی وب‌سرور را نشان می‌دهد.



(شکل-۲): روند استخراج پروفایل از داده اصلی
(figure-2): Steps to extarct profiles from input data

۴- چارچوب پیشنهادی تشخیص ناهنجاری

در بخش ۳ مفاهیم کنترل دسترسی و کلیات مربوط به استخراج پروفایل کاربرد وب بیان شد. در این بخش رویکرد پیشنهادی جهت تشخیص ناهنجاری در نشست‌های وب مطرح می‌شود. روش پیشنهادی دو مرحلهٔ اصلی دارد، مرحلهٔ نخست استخراج پروفایل‌های کاربرد وب به‌عنوان الگوی پیمایش کاربران وب است. جزییات این روش در بخش قبل آورده شد. مرحله دوم محاسبه فاصله هر نشست ورودی کاربر با پروفایل‌های اصلی و همچنین استخراج هشدارهای کنترل دسترسی متناظر با همان نشست و تزریق این داده‌ها به شبکه عصبی فازی جهت تشخیص هنجار یا ناهنجار بودن پیمایش کاربر است. شکل (۳) نمایی کلی از روش مورد استفاده در تشخیص ناهنجاری را نشان می‌دهد.



(شکل-۳): مراحل روش پیشنهادی تشخیص ناهنجاری وب
(figure-3): Anomaly detection procedure in proposed method

تعداد پروفایل‌های اصلی بایستی با دقت انتخاب شود چراکه این پروفایل‌ها نمایان‌گر الگوی پیمایش وب توسط کاربران هستند. درصورت افزایش تعداد، بار محاسباتی روی سامانه بالا رفته و حتی ممکن است منجر به تعادل نتایج سامانه کنترل دسترسی شود. درصورت کم‌بودن تعداد، پروفایل‌ها نمی‌توانند الگوی پیمایش وب را به‌طورکامل پوشش دهند. شکل مراحل رویکرد پیشنهادی این پژوهش را نمایش می‌دهد.

با توجه به ماهیت نوفه‌ای لاگ‌های دسترسی وب‌سرور نیاز به استفاده از روشی بود که با تغییر اندک داده‌ها در صحت نتایج تغییر محسوسی ایجاد نشود. بنابراین از یک شبکه عصبی فازی جهت یادگیری الگوی پیمایش‌های نرمال از پیمایش‌های ناهنجار استفاده شد. ورودی این شبکه عصبی ماتریسی از میزان شباهت‌های محاسبه‌شده نسبت به پروفایل‌های اصلی و نیز تعداد هشدارهای کنترل دسترسی در نشست متناظر است و خروجی آن برچسب هنجار یا ناهنجار بر روی داده ورودی است که با استفاده از یک طبقه‌بند سیگموید در لایه خروجی شبکه انجام می‌شود. در شبکه عصبی مورد استفاده جهت مقداردهی اولیه به قوانین از روش خوشه‌بندی FCM[1] استفاده شد. مقادیر اولیه وزن‌ها نیز با مقادیر بسیار کوچک مقداردهی شد. یادگیری وزن‌ها شبکه مذکور از روش back propagation و با روش gradient descent انجام شد.

سیاست کنترل دسترسی مورد استفاده در این روش SELinux است. SELinux توسط United States National Security Agency (NSA) به‌عنوان یک ویژگی به کرنل لینوکس اضافه و بر اساس کنترل دسترسی اجباری[2] (MAC)

[1] Fuzzy Cmean
[2] Mandatory Access Control

پیاده‌سازی شده است؛ درواقع سازوکار SELinux به این صورت است که در ابتدا کرنل به بررسی کنترل دسترسی اختیاری (DAC) می‌پردازد. درصورتی‌که این بررسی با موفقیت انجام شود، پس از آن به بررسی MAC که پایه و اساس SELinux است، پرداخته می‌شود. درواقع به این مورد باید توجه کنیم که MAC پس از DAC به بررسی سطح دسترسی‌ها خواهد پرداخت، درصورتی‌که اجازه دسترسی را صادر نکند کرنل هرگز به سراغ MAC نخواهد رفت؛ SELinux همچنین از کنترل دسترسی مبتنی بر نقش نیز پشتیبانی می‌کند و بنابراین نسبت به هر سه مدل مطرح‌شده در بخش ۳ برتری دارد. با توجه به محدودیت‌های ایجادشده در صورت اجرای کنترل دسترسی SELinux در مد اجباری[1]، سیاست کنترل دسترسی در مد مجاز[2] اجرا شد. در این مد از اجرای هیچ فرآیندی ممانعت به‌عمل نمی‌آید و تنها عدم پذیرش‌ها جهت تحلیل در فایل لاگ ثبت می‌شوند.

۵- راه‌اندازی آزمایشی

علاوه‌بر روش پیشنهادی بخش ۴ به‌منظور تشخیص ناهنجاری در این پژوهش، رویکردی نیز به‌منظور شبیه‌سازی پیمایش‌های یک کاربر عادی ارائه شد. به‌دلیل فقدان داده استانداردی که هم شامل پیمایش‌های وب صفحات و هم شامل هشدارهای کنترل دسترسی متناظر با آن باشد، در دسترس نبود، نیاز به تولید چنین داده‌ای وجود داشت. مرحله نخست تهیه داده، تولید لاگ دسترسی و لاگ کنترل دسترسی است. به همین منظور یک وب سرورApache در بستر یک ماشین مجازی Centos راه‌اندازی شد. برای ایجاد لاگ‌هایی که به‌طورکامل مشابه لاگ یک سرور واقعی باشند، روی سرور Apache، یک وب سایت خرید کالا مانند نمونه‌های واقعی آن راه‌اندازی شد. وب‌سایت مورد نظر حدود ۱۶۰ صفحه مختلف برای بازدید کاربران داشت و شامل صفحات انتخاب نوع کالا، انتخاب نام شرکت تولیدکننده(برند)، صفحات مربوط به مشخصات یک کالای خاص، صفحات خرید، login، ارتباطات و سفارش‌های کاربران است.

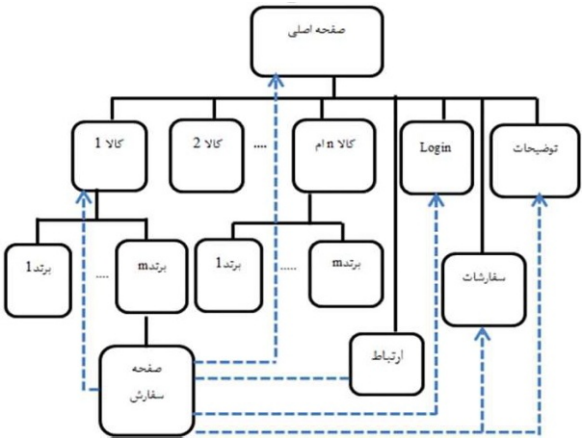
پس از ایجاد بستر مناسب برای ثبت لاگ‌های مورد نیاز، نوبت تولید لاگ مورد نظر می‌شود و بایستی از ابزارهایی استفاده می‌شد که بتوانند مشاهده صفحات وب توسط کاربران مختلف را شبیه‌سازی کنند.

[1] Enforcing
[2] Permissive

۵-۱- شبیه‌سازی رفتار کاربر

با توجه به نیاز شبیه‌سازی رفتار چندین کاربر مختلف و تنظیم زمان مشاهده صفحات، ترتیب آن‌ها و نشانی کاربر مشاهده‌کنندهٔ صفحه، بایستی از ابزار متن‌بازی استفاده می‌شد که این تغییرات در پیکربندی آن قابل اعمال باشند. بنابراین پس از بررسی ابزارهای موجود در این زمینه ابزاری به نام curl-loader انتخاب شد. فعالیت هر کلاینت مجازی شبیه‌سازی‌شده توسط این ابزار قابل جمع‌آوری و ثبت لاگ در سرور وب است. لاگ ثبت‌شده مانند لاگ مربوط به دسترسی یک کاربر واقعی شامل اطلاعات مربوط به تجزیهIP، برپایی ارتباط، ارسال درخواست و دریافت پاسخ، داده‌ها و فرآیندهای ارسال و دریافت‌شده، خطاهای شبکه و رویدادها و خطاهای لایه کاربرد هستند.

به منظور شبیه‌سازی رفتار یک کاربر نرمال، ابتدا روند پیمایش یک کاربر معمولی مطابق با شکل (۴) ارائه شد. همان‌طور که در شکل (۴) دیده می‌شود، مشاهده صفحات اصلی و در ادامه صفحات مربوط به نوع کالا و سپس برند کالای مورد نظر و درنهایت خرید، یک روال عادی مشاهده سایت است؛ اما از آنجا که کاربر می‌تواند در هر زمان روی سایر پیوندهای موجود در صفحه نیز کلیک کرده و از آن‌ها نیز بازدید کند، روالی پیشنهاد شد که بتواند تمامی احتمالات ممکن را در نظر بگیرد. بنابراین از هر صفحه به تمامی صفحاتی که به آن لینکی وجود دارد، احتمال انتقال نیز با ضریب کوچکتری نسبت به حالت قبل وجود دارد.



(شکل-۴): پیمایش‌های ممکن برای کاربر در مشاهده صفحات وب
(figure-4): Possible navigation paths for user in web pages scrolling

پیوندهای موجود در صفحه اصلی همه به‌طور یکسان شانس انتخاب‌شدن دارند؛ اما با پیش‌رفتن در مسیر پیمایش،

فضاهای انتخاب مستقل از هم پدید می‌آیند. ورود به هریک از این فضاهای مستقل شانس برابر ندارد به این معنی که شانس انتخاب یک برند در صفحه یک کالای خاص و ورود به فضای کالاها، با احتمال ورود به صفحه ارتباط یکسان نیست؛ اما در صورت ورود به هر فضا، احتمال انتخاب با استفاده از تابع توزیع احتمال نرمال صورت می‌گیرد درواقع احتمال انتخاب برندهای مختلف یک کالا، برابر در نظر گرفته شده است. همچنین احتمال انتخاب پیوندی در فضای موجود بیشتر از احتمال انتخاب یک پیوند از فضایی دیگر در نظر گرفته شده است.

با درنظرگرفتن تمامی موارد مذکور و استخراج تمام مسیرهای پیمایش ممکن روی ۱۶۰ صفحه وب سایت نوشته شده، یک اسکریپت bash جهت پیاده‌سازی عملی نوشته شد. پیمایش صفحات در این اسکریپت، بر اساس توضیحات داده‌شده، صورت می‌گیرد و برای مشاهده هر صفحه، برنامه curl-loader فراخوانی می‌شود. همچنین به‌منظور واقعی‌بودن زمان مشاهدات مربوط به هر صفحه، از یک تابع وقفه در اسکریپت مذکور استفاده شده است که با توجه به اهمیت صفحه، هم مطالب آن و محل قرارگیری صفحه در کل پیمایش زمان سپری‌شده روی هر صفحه نیز متغیر است.

با اجرای اسکریپت مذکور، دسترسی کاربران مختلف به وب‌سرور شبیه‌سازی می‌شود و دسترسی‌های صورت‌گرفته در فایل access.log مربوط به سرورApache ثبت می‌شود.

در هر اجرای اسکریپت مذکور کلاینت‌ها از آدرس‌های IP مجزا از صفحات وب بازدید می‌کنند. با اجرای متوالی اسکریپت بیش از ۱۲۰۰۰ دسترسی مختلف در فایل لاگ ثبت شد. پس از انجام مراحل پیش‌پردازش، نشست‌های وب بر اساس فاصله زمانی جدا شدند. سپری‌شدن زمان بیش از یک ساعت و تغییر آدرسIP به اتمام نشست فعلی منجر می‌شود. همچنین درخواست‌هایی که هم‌زمان در یک ثانیه رخ می‌دهند نیز از نشست‌ها حذف می‌شوند. استخراج نشست‌های وب با استفاده از نرم‌افزار مطلب صورت گرفته است. پس از استخراج نشست‌های وب (حدود ۹۱۰ نشست جداگانه) از لاگ‌های تولیدشده، الگوریتم بخش۴-۲ روی این داده اجرا شد و ۱۰ پروفایل اصلی داده به‌عنوان الگوهای اصلی رفتار یک کاربر عادی استخراج شدند. نمونه‌ای از پروفایل تولید شده در جدول (۱) آمده است.

(جدول-۱): سه نمونه از پروفایل‌های استخراج شده از لاگ دسترسی وب

(table-1): Three samples of extracted web usage profiles

پروفایل‌های استخراج شده
/home/mobilepage /home/Pcpage /home/PCpage/dell /home/PCpage/dell/Dell-INSPIRON-3541 /home/PCpage/dell/Dell-INSPIRON-3542 /home/PCpage/asus /home/mobilepage /home/mobilepage/Samsung /home/mobilepage/Samsung/Samsung-Galaxy-Note-4
/home/mobilepage /home/mobilepage/HTC /home/mobilepage/iphone /home/mobilepage/iphone/iPhone-6-16GB /home/mobilepage/HTC/HTC-Desire-200 /home/mobilepage/HTC/HTC-Desire-310 /home/mobilepage/HTC/HTC-Desire-610 /home/mobilepage/HTC/HTC-Desire-816 /home/mobilepage/Samsung /home/camerapage /home/camerapage/nikon /home/camerapage/nikon/Nikon-COOLPIX-L8 /home/Pcpage
/home/tabletpage /home/tabletpage/samsungtab /home/tabletpage/samsungtab/Samsung-Galaxy-Note-101-2014-Edition-3G-16i /home/PCpage/asus /home/PCpage/asus/ASUS-X550LD /home/PCpage/asus/ASUS-X550L /home/PCpage/asus/ASUS-X200MA /home/PCpage/asus/ASUS-KN551LN /home/PCpage/asus/ASUS-N56JN /home/PCpage/asus/ASUS-X452 /home/PCpage/apple /home/camerapage

۲-۵- تولید داده حملات

همان‌طور که بیان شد، کاربر عادی به‌طورمعمول به صفحات اصلی سایت مراجعه می‌کند، اما رفتار مهاجمان از الگوی متفاوتی پیروی می‌کند. به‌عنوان نمونه، دسترسی به صفحات، در پوشه‌هایی که محتوای نمایشی مرتبط با کاربران ندارند مانند پوشه cgi-bin، تکرار دسترسی به صفحات با سطح دسترسی مدیر سامانه و یا دسترسی مکرر به پایگاه داده به‌طور مشخص رفتار یک کاربر عادی نیست و با پروفایل‌های اصلی تشخیص داده‌شده فاصله رفتاری دارد. علاوه‌براین سیاست کنترل دسترسی تعریف‌شده در زمان بروز این موارد هشداری در فایل لاگ ذخیره می‌کند که معیاری جهت سنجش پیمایش ناهنجار است.

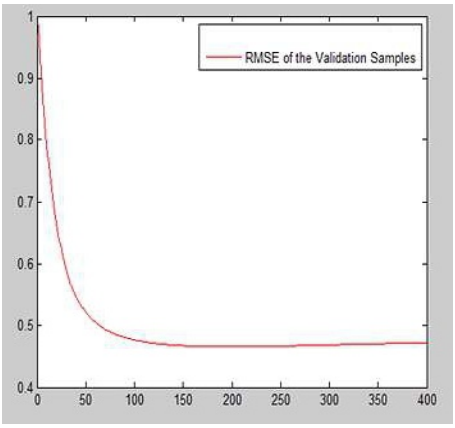
همان‌طورکه بیان شد برای آموزش شبکه عصبی به داده با برچسب هنجار و ناهنجار نیاز است. داده هنجار با استفاده از اسکریپت بیان‌شده در بخش قبل تولید شده است؛ اما برای تولید پیمایش‌های ناهنجار از ابزار حمله استفاده شد.

حملات انجام‌شده شامل DOS وDDOS، انواعfuzzer و حملاتDictionary بوده و همچنین طیف بزرگی از انواع ابزارهای پویش آسیب‌پذیری وب را دربرمی‌گیرد. فهرست ابزارهای استفاده‌شده جهت پیاده‌سازی حملات به این شرح است: DirBuster, Webshag, Wfuzzer, Netsparker, Webscarb, wapati, Rfuzz, W3af, Skipfish. از ابزارهای مذکور با انواع پیکربندی و با زمان‌بندی‌های مختلف استفاده

شده تا داده حملات متنوع باشد. با توجه به اینکه تعداد نشست‌های وب استخراج‌شده از داده حمله ۴۵۰ نشست است. حدود چهارصد نشست وب نیز با استفاده از ابزار شبیه‌سازی رفتار کاربر تولید شد تا شبکه عصبی فازی به سمت یکی از برچسب‌های مورد نظر بایاس نشود. لازم به توضیح است که داده ورودی به شبکه عصبی فازی یک ماتریس به ابعاد ۸۵۰×۱۱ است، که ۸۵۰ مجموع نشست‌های حمله و هنجار بوده و ۱۱ میزان شباهت به هرکدام از ۱۰ پروفایل اصلی و تعداد اخطارSELinux در بازه زمانی نشست‌های خاص است.

۳-۵- آزمایش و ارزیابی

پس از انجام مرحله یادگیری شبکه عصبی فازی نوبت به مرحله تست می‌رسد. به‌منظور آزمایش این شبکه از روش K-cross fold و به طور خاص از 3-fold استفاده شد. نتایج نهایی از میانگین‌گیری بر روی نتایج مربوط به k مرحله آزمایش به دست می‌آید. در شکل (۵) نمودار یادگیری شبکه عصبی فازی پیشنهادی بر اساس خطای RMSE آورده شده است.



(شکل-۵): نمودار یادگیری شبکه عصبی فازی پس از ۴۰۰ تکرار بر حسب خطای RMSE

(figure-5): Fuzzy neural network learning diagram after 400 repetitions in terms of RMSE error

از آنجا که رویکرد پیشنهادی و داده مورد استفاده از جنسی نیستند که مشابه آن‌ها در سایر روش‌های تشخیص ناهنجاری وب آورده شده باشد، روشی را مبنای مقایسه با آن نمی‌توان قرار داد. در اینجا هدف، ارزیابی قدرت برچسب‌زنی الگوریتم است.

۴-۵- ارزیابی پایداری الگوریتم نسبت به نوفه

با توجه به ماهیت نوفه‌ای لاگ‌های دسترسی وب‌سرور روش پیشنهادی در برابر نوفه مورد ارزیابی قرار گرفت. به این منظور نتایج روی پروفایل‌های نوفه‌ای بررسی شد. پروفایل‌های نوفه‌ای با ایجاد تغییرات رندوم روی پروفایل‌های اصلی تولید می‌شوند و تنها مرحله آزمایش روش دوباره انجام می‌گیرد. در ادامه فاصله از پروفایل‌های دارای نویز با پروفایل‌های اصلی مقایسه می‌شود. به منظور نشان دادن توانایی روش پیشنهادی، در این قسمت نتایج با یک SVM[1] مقایسه شده است.

معیارهای نرخ مثبت اشتباه[2] (FPR)، نرخ مثبت صحیح[3] (TPR)، معیار نرخ منفی اشتباه[4] (FNR)، نرخ پیش‌گویی مثبت[5] (PPV)، نرخ پیش‌گویی منفی[6](NPV) و معیار F برای نتایج به‌دست‌آمده محاسبه شده و در جدول (۲) آورده شده است. معیار[7]F [49]که با عنوان F1 نیز استعمال می‌شود، درحقیقت نوعی میانگین هارمونیک[8] برای دو معیار PPV و TPR است. این معیار، مقداری بین صفر و یک دارد و بهترین مقدار آن، یک است.

ابتدا یک شبکه SVM با داده مشابه داده ورودی روش پیشنهادی آموزش داده می‌شود و نتایج هر دو روش مشخص و سپس به شکل رندوم به پروفایل‌های اصلی نوفه افزوده شده و دوباره عملکرد هر دو روش در برابر پروفایل‌های نوفه‌ای بررسی می‌شود. برای افزودن نوفه به شکل رندوم بخشی از پروفایل‌ها حذف و یا اضافه شدند. جدول (۲) نتایج مربوط به خروجی رویکرد پیشنهادی و نتایج مربوط به خروجی SVM را نشان می‌دهد.

همان‌طور که در جدول دیده می‌شود هر دو روش در برچسب‌های مثبت عملکرد مشابه و خوبی دارند و تعداد برچسب‌های زده‌شدهٔ اشتباه مثبت یکسان است. تفاوت اصلی در زدن برچسب‌های منفی است. همچنین دیده می‌شود که با وجود عملکرد بهتر روش SVM در حالت معمولی، با افزودن نوفه به پروفایل‌ها نسبت به رویکرد پیشنهادی نتایج ضعیف‌تری دارد و می‌توان نتیجه گرفت که روش پیشنهادی در برابر نوفه به مراتب بهتر عمل می‌کند.

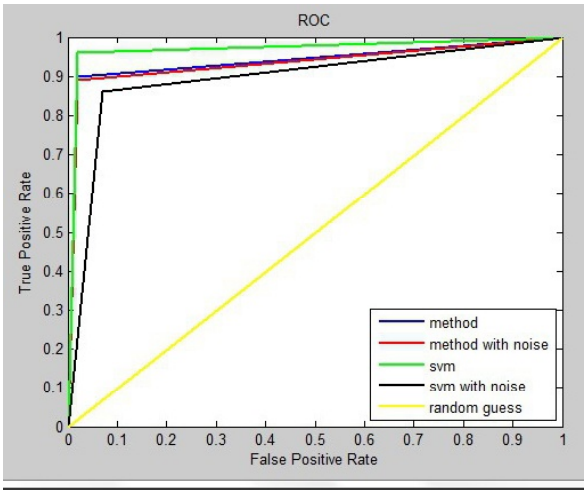
[1] Support vector Machine
[2] False Positive Rate (FPR)
[3] True Positive Rate (TPR)
[4] True Positive Rate (TPR)
[5] Positive Predicted Value (PPV)
[6] Negative Predicted Value (NPV)
[7] F-measure, F-score
[8] Harmonic Mean

(جدول-۲): مقایسه نتایج آزمایش روی پروفایل‌های نوفه‌ای با روش پیشنهادی و روش SVM

(table-2): Comparison of test results on noise profiles with proposed method and SVM method

معیار	روش	SVM	روش با نویز	SVM با نویز
FPR	۹/۲	۳/۷	۹/۲	۱۲/۹
TPR	۹۷	۹۷	۹۷	۹۷
FNR	۳	۳	۳	۳
TNR	۹۰	۹۶	۹۰	۸۷
PPV	۸۸	۹۵	۸۸	۸۴
NPV	۹۸	۹۸	۹۸	۹۷
F	۹۲	۹۴	۹۲	۹۰

نمودار ROC مربوط به ارزیابی نتایج در شکل (۶) نمایش داده شده است. بر اساس نمودار ROC روش پیشنهادی در حالت معمولی نسبت به SVM ضعیف‌تر عمل می‌کند؛ اما در حالتی که پروفایل‌ها دارای نوفه باشند، این روش نتایج بهتری را نشان می‌دهد.



(شکل-۶): نمودار ROC روش پیشنهادی در حضور نوفه در مقایسه با روش SVM

(figure-6): ROC Chart of proposed method in the presence of noise compared to the SVM method

۶- نتیجه‌گیری

در این مقاله به ارائه یک رویکرد جدید در تشخیص ناهنجاری در پیمایش‌های وب پرداخته شد. در این روش ابتدا پروفایل‌های اصلی بازدید کاربران استخراج و سپس با استفاده از میزان شباهت پیمایش کاربر با این پروفایل‌ها و استخراج هشدارهای سامانهٔ کنترل دسترسی، یک شبکه عصبی فازی آموزش داده می‌شود تا به پیمایش‌های ناهنجار و مخرب صفحات وب را تشخیص دهد. داده ناهنجار با استفاده از بسیاری از ابزارهای تولید حمله موجود صورت گرفت. به‌منظور دست‌یابی به نتایج واقعی‌تر توصیه می‌شود، رویکردی مشابه روش پیشنهادی روی داده‌های وب‌سرور واقعی مورد بررسی قرار گیرد. درصورتی‌که بازه زمانی جمع‌آوری داده روی یک سرور واقعی زیاد باشد، روش پیشنهادی می‌تواند پیمایش‌های ناهنجار مربوط به مهاجمان و یا رباط‌ها را با دقت مناسبی تشخیص دهد.

۷- مراجع 7-References

[1] J. Daniel E. Geer. The shrinking perimeter: Makingthe case for data-level risk case management. Veradsys White Paper, January 2004.

[2] C.Squicciarini, Elisa Bertino. Lorenzo D.Martino.Fedrica Paci.Anna .Security for Web Services and Service-Oriented Architectures .Springer, 2010.

[3] R. Azmi, B. Pishgoo, H. Nemati, "Hypervisor-based Intrusion Detection Using Artificial Immune Systems", 8th International Iranian ISC Conference on Information Security and Cryptology, pp. 147-153, (2011).

[4] S. S. Anand and B. Mobasher, "Intelligent Techniques for Web Personalization", *LNAI 3169*, Springer-Verlag ,2005, 1–37.

[5] B. Mobasher, "Web Usage Mining and Personalization", Practical Handbook of Internet Computing, Chapman Hall and CRC Press, 2004.

[6] Selma Elsheikh.2008. Web Usage Data for Web Access Control (WUDWAC). World Congress on Engineering, Jul 2008.

[7] Priyanka V. Patil, Dharmaraj Patil , 2013,Preprocessing Web Logs for Web Intrusion Detection, IJAIS Proceedings on International Conference and workshop on Advanced Computing 2013.

[8] Grant panel, Helen Ashman.2010, Anomaly Detection Over User Profiles for intrusion detection, Originally published in the Proceedings of the 8th Australian Information Security Mangement Conference, Edith Cowan University, Perth Western Australia.

[9] Yi Xie, Shensheng Tang. 2012,online anomaly detection based on web usage minig, IEEE 26th international parallel abd Distributed Processing Symposiom.

[10] Hamid Bagheri,Fereidoon Shams, 2011, "An Auto-Delegation Mechanism for Role Based Access Control model" 2nd World Conference on Information Technology", Antalya.

[11] Suganyadevi Janani Manimozhi Mirdula, 2002, "Preprocessing in Web Usage Mining" .

[12] R. Kosala and H. Blockeel, "Web mining research: a survey," ACM SIGKDD Explorations Newsletter, vol. 2, no. 1, pp. 1–15, Jun. 2000.

[13] P. R. Kumar and A. K. Singh, "Web Structure Mining: Exploring Hyperlinks and Algorithms for Information Retrieval," *American Journal of applied sciences*, vol. 7, no. 6, pp. 840–845, 2010.

[14] J. Sivaramakrishnan and V. Balakrishnan, "*Web Mining Functions in an Academic Search Application,*" *Informatica*, vol. 13.

[15] J. Srivastava, R. Cooley, M. Deshpande, and P.-N. Tan, "Web usage mining: discovery and applications of usage patterns from Web data," *ACM SIGKDD Explorations Newsletter*, vol. 1, no. 2, pp. 12–23, Jan. 2000.

[16] L. K. Grace, V. Maheswari, and D. Nagamalai, "*Analysis of Web Logs and Web User in Web Mining,*" International Journal of Network Security & Its Applications, Jan. 2011.

[17] D. Dixit and M. Kiruthika, "Preprocessing of Web Logs," *International Journal on Computer Science and Engineering*, vol. 2, pp. 2447-2452, 2010.

[18] V. Sathiya Moorthi and V. Murali Bhaskaran, "Data preparation Techniques for Web Usage Mining in World Wide Web–an approach," *International Journal of Recent Trends in Engineering*, vol. 2, no. 4, 2009.

[19] B. Mobasher, H. Dai, T. Luo, and M. Nakagawa, "Effective personalization based on association rule discovery from web usage data," in *Proceedings of the 3rd international workshop on Web information and data management*, Atlanta, Georgia, USA, 2001, pp. 9–15.

[20] H.Malek,M.M.Ebadzadeh,M.Rahmati, Threenew fuzzy neural networks learning algorithms based on clustering, training error and genetic algorithm,ApplIntell.35(2011)1–

[21] S.L. Chiu,Fuzzy model identification based on cluster estimation,J.Intell. Fuzzy Syst.2(1994)209–219.

[22] R.R.Yager,D.P.Filev,Learning of fuzzy rules by mountain clustering,in: Proceeding ofSPIEConferenceonAppliedFuzzyLogicTechnology,1993, pp. 246–254.

[23] A. Salimi,M.M.Ebadzadeh, CFNN: Correlated fuzzy neural network, Neurocomputing148(2015)430–444.

[24] G. Leng,Th.McGinnity,Design for self-organizing fuzzy neural network based on geneticalgorithm,IEEETrans.FuzzySyst.14(2006)755–766.

[25] B.Pizzileo,K.Li,G.W.Irwin,W.Zhao, Improved structure optimization for fuzzy-neuralnetworks,IEEETrans.FuzzySyst.20(2012)1076–1089.

[26] T. W. Yan, M. Jacobsen, H. Garcia-Molina, and U. Dayal, "From user access patterns to dynamic hypertext linking," *Computer Networks and*

[27] R. Forsati, M. R. Meybodi, and A. Rahbar, "An efficient algorithm for web recommendation systems," presented at the IEEE/ACS International Conference on Computer Systems and Application-s, AICCSA 2009, 2009, pp. 579-586.

[28] N. C. Jones and P. Pevzner, *An introduction to bioinformatics algorithms*. The MIT Press, 2004.

[29] W. Wang and O. R. Zaïane, "Clustering Web Sessions by Sequence Alignment," in Proceedings of 13th International Workshop on Database and Expert Systems Applications, Los Alamitos, CA, USA, 2002, vol. 0, p. 394.

[30] C. Li and Y. Lu, "Similarity Measurement of Web Sessions by Sequence Alignment," presented at the IFIP International Conference on Network and Parallel Computing Workshops, NPC Workshops, 2007, pp. 716-720.

[31] B. Hay, G. Wets, and K. Vanhoof, "Segmentation of visiting patterns on web sites using a sequence alignment method," *Journal of Retailing and Consumer Services*, vol. 10, no. 3, pp. 145-153, May 2003.

[32] R.Azmi,M.Azimpour-kivi, "Applying Sequence Alignment in Tracking Evolving Clusters of Web-Sessions Data:an Artificial Immune Network Approach", 2011 Third International Conference on Computational Intelligence, Communication Systems and Networks.

[33] B. Hay, G. Wets, and K. Vanhoof, "Segmentation of visiting patterns on web sites using a sequence alignment method," *Journal of Retailing and Consumer Services*, vol. 10, no. 3, pp. 145-153, May 2003.

[34] B. H. Helmi and A. T. Rahmani, "An AIS algorithm for Web usage mining with directed mutation," in *IEEE Congress on Evolutionary Computation, CEC 2008 (IEEE World Congress on Computational Intelligence)*, 2008, pp. 3122-3127.

[35] T. Zhang, R. Ramakrishnan, and M. Livny, "BIRCH: an efficient data clustering method for

very large databases," ACM SIGMOD Record, vol. 25, no. 2, pp. 103–114, Jun. 1996.

[36] O. Nasraoui, C. C. Uribe, C. R. Coronel, and F. Gonzalez, "TECNO-STREAMS: tracking evolving clusters in noisy data streams with a scalable immune system learning model," presented at the Third IEEE International Conference on Data Mining, ICDM, 2003, pp. 235- 242.

[37] S.Alam,G.Dobbie,P.Riddle,"Particle Swarm Optimization basedClustering Of Web Usage Data",2008 IEEE/WIC/ACM International Conference on web Intelligent and Intelligent Agent Technology.

[38] R.Azmi,M.Raji,V.Derhami," Web Anomaly Detection Using Arti_cial Immune System and Web Usage Mining Approach "2012, ICIC,Zanjan

[39] C. Kruegel and G. Vigna, Anomaly detection of web-based attacks, in Proceedings of the 10th ACM Conference on Com-puter and Communications Security (2003), 251-261

[40] L. Guangminl, Modeling Unknown Web Attacks in Network Anomaly Detection, International Conference on Conver-gence and Hybrid Information Technology (2008).

[41] M. Danforth, Towards a Classifying Arti_cial Immune Sys-tem for Web Server Attacks: Department of Computer andElectrical Engineering and Computer Science, Interna-tional Conference on Machine Learning and Applications (2009).

[42] M. A. Rassam, M. A. Maarof, and A. Zainal, Intrusion De-tection System Using Unsupervis-ed Immune Network Cluster-ing with Reduced Features, Int. J. Advance. Soft Comput. Appl. 2/2010 (2010).

[43] Valeur, F., Mutz, D., Vigna, G.: A learning-based approach to the detection of SQL attacks. In: Julisch, K., Krügel, C. (eds.) DIMVA 2005. LNCS, vol. 3548, pp. 123–140. Springer, Heidelberg (2005).

[44] Kantardzic, M.: Data Mining Concepts, Models, Methods and Algorithm. IEEE Press, New York (2002).

[45] L. Jie, S. Jianwei, H.Changzhen," A Novel Framework for Active Detection of HTTP Based Attack", Communication Systems and Information Technology,.Springer-Verlag Berlin Heidelberg 2011.

[46] R. Anderson. Security Engineering: A Guide to Building Dependable Distributed Systems. WileyComputer Publishing, New York, New York, 2001.

[۴۷] جعفریان مقدم. احمد رضا، برزین‌پور. فرناز، فتحیان. محمد، "روش نوین خوشه‌بندی ترکیبی با استفاده از سامانه ایمنی مصنوعی و سلسله مراتبی"، فصلنامه علمی پژوهشی پردازش علائم و داده‌ها دوره ۱۳ شماره ۴ (۱۳۹۵-۱۲).

[47] A.Jafarian-Moghaddam,F. Barzinpour,M. Fathian, new clustering Technique us-ing Artificial Immune System and Hierarchical technique,Quarty journal Signal and Data Processing, Volume 13, Issue 4 (3-2017).

[48] B. W. Lampson. Protection. ACM SIGOPS Operating System Review, 8(1):18–24, January 1974.

[49] Wu, S. X., Banzhaf, W.,"The use of computational intelligence in intrusion detectionsystems: A review", Applied Soft Computing, vol. 10, pp. 1–35, (2010).

مریم السادات میرهادی تفرشی مدرک کارشناسی خود را در رشته مهندسی کامپیوتر از دانشگاه شاهد و مدرک کارشناسی ارشد خود را در سال ۱۳۹۴ در رشته هوش مصنوعی از دانشگاه الزهرا دریافت کرده است. او از سال ۱۳۹۳ عضو آزمایشگاه امنیت سیستم عامل (OSSL) و آزمایشگاه تشخیص ناهنجاری مبتنی بر وب (WADL) دانشگاه الزهرا بوده و در پروژه‌های انجام‌شده همکاری داشته است. زمینه‌های علاقه‌مندی ایشان تشخیص ناهنجاری وب، متن‌کاوی و داده‌کاوی، تشخیص الگو، یادگیری ماشین و امنیت شبکه است.

نشانی رایانامه ایشان عبارت است از :

m.s.mirhadi@gmail.com

رضا عزمی مدرک کارشناسی خود را از دانشگاه صنعتی امیرکبیر در رشته مهندسی برق الکترونیک دریافت کرد. ایشان مدرک کارشناسی ارشد و درجه دکترای خود را به‌ترتیب در سال‌های ۱۳۷۲و ۱۳۷۸ از دانشگاه تربیت مدرس در رشته مهندسی برق الکترونیک اخذ کرد. وی هم‌اکنون عضو هیئت علمی گروه مهندسی کامپیوتر در دانشگاه الزهرا است. ایشان بنیان‌گذار آزمایشگاه امنیت سیستم عامل (OSSL)، آزمایشگاه تشخیص ناهنجاری مبتنی بر وب (WADL)، آزمایشگاه پردازش تصاویر پزشکی (MIPL)، آزمایشگاه

تشخیص حالت صورت و چهره (FFERL) در دانشگاه الزهرا است. وی مدیر پروژه و عضو فنی بسیاری از پروژه‌های صنعتی نیز بوده است.
نشانی رایانامه ایشان عبارت است از:
azmi@alzahra.ac.ir