
Phishing Website Detection for e-Banking by Inclined Planes Optimization Algorithm

N. Langari*, M. Abdolrazzagh-Nezhad

*University of Birjand

(Received: 22/10/2014, Accepted: 01/09/2015)

ABSTRACT

One of the most important factors influencing the development of e-commerce and web-based commerce is security. However development of e-commerce leads to phishing and steal the customer information. So the various methods have been designed to detect phishing websites in the literature. Lacks of attention to the short lifetime of phishing website, and to reduce the amount of computation are the main gaps of these methods. In this paper, a new intelligent approach is proposed to detect phishing websites, in e-banking by extracting sensitive features of websites on phishing attacks and classifying candidate websites in three classes such as phishing, legitimate and suspicious websites based on inclined planes optimization algorithm. The comparison results of the new intelligent approach with the best available techniques, demonstrate the ability of this approach to detect phishing websites.

Keywords: Network Security; Service Security; Port Security; Authentication; Port-Knocking.

شناسایی وب‌گاه فیشینگ در بانک‌داری اینترنتی با استفاده از الگوریتم بهینه‌سازی

صفحات شیب‌دار

نفسه لنگری^{۱*}، مجید عبدالرزاق نژاد^۲

۱- کارشناسی ارشد، دانشکده فنی و مهندسی، دانشگاه بیرجند

۲- استادیار، دانشکده فنی و مهندسی، دانشگاه بزرگمهر قائنات

(دریافت: ۹۳/۰۷/۳۰؛ پذیرش: ۹۴/۰۶/۱۰)

چکیده

یکی از عوامل بسیار تأثیرگذار در توسعه تجارت الکترونیک و تجارت تحت وب، امنیت آن می‌باشد. اما متناسب با توسعه تجارت الکترونیک، مقوله فیشینگ و سرقت اطلاعات بانکی افراد به تهدید بسیار جدی در این حوزه بدل شده است. روش‌های متنوعی در شناسایی وب‌گاه فیشینگ مورد بررسی و تحلیل قرار گرفته‌اند. در اکثر روش‌ها توجهی به طول عمر کوتاه وب‌گاه فیشینگ و تلاش برای کاهش حجم محاسباتی صورت نگرفته است. از این جهت، در این مقاله یک رویکرد هوشمند جدید به منظور شناسایی وب‌گاه‌های فیشینگ در بانک‌داری الکترونیک براساس استخراج ویژگی‌های پراهمیت از جهت ارزیابی وب‌گاه فیشینگ و طبقه‌بندی وب‌گاه‌های کاندید در سه کلاس فیشینگ، قانونی و مشکوک توسط الگوریتم بهینه‌سازی صفحات شیب‌دار پیشنهاد و پیاده‌سازی گردیده است. مقایسه نتایج حاصله از این رویکرد هوشمند جدید با بهترین روش‌های موجود، اثبات‌کننده توانایی این رویکرد در شناسایی وب‌گاه‌های فیشینگ می‌باشد.

واژه‌های کلیدی: تشخیص وب‌گاه فیشینگ، الگوریتم بهینه

۱- مقدمه

فیشینگ، یک سرقت هویت برخط است که حمله‌کنندگان از ترکیب تکنیک‌های مهندسی اجتماعی و تکنیک‌های جعل سایت، برای آشکار شدن اطلاعات محرمانه کاربران از قبیل اطلاعات کارت اعتباری استفاده می‌کنند. تقریباً حدود یک دهه پیش اولین حملات فیشینگ صورت گرفت. اثرات این حملات روی قربانی‌ها بیشتر آزار دهنده بود، تا این که عواقب مالی داشته باشد. در همین حال، خدمات اینترنتی تکمیل تر شد و سیستم‌های بانک‌داری به‌طور گسترده‌ای پذیرفته شد. از این رو حملات فیشینگ که اوایل هدفی جز آزار دادن قربانی نداشت، با هدف قرار دادن خدمات مالی به یک کسب و کار جدی تبهکارانه تبدیل شد. متناسب با افزایش سطح تهدید فیشینگ، تنوع راه‌کارهای شناسایی وب‌گاه‌های فیشینگ و تقابل با این تهدید نیز افزایش یافت. لذا راه‌کارهای ارائه‌شده را می‌توان در ۵ گروه دسته‌بندی و تشریح نمود که در ادامه معرفی شده‌اند.

در حال حاضر با رشد شبکه جهانی اینترنت، فواصل جغرافیایی و اختلافات زمانی کم‌رنگ شده است و به‌صورت برخط می‌توان تبادلات تجاری و مالی را به راحتی انجام داد. یکی از بزرگترین موانع موجود بر سر راه توسعه تجارت الکترونیک، امنیت مربوط به آن و تضمین ایمنی دادوستد از طریق شبکه وب می‌باشد. یکی از حوزه‌های مطرح در تجارت الکترونیک، بانک‌داری اینترنتی است که به موجب آن تمام فعالیت‌های بانکی به‌صورت اینترنتی قابل اجرا است. از این جهت، امنیت خدمات بانک‌داری اینترنتی یک مسئله حیاتی است. فیشینگ^۲، آشکارترین نوع حمله روی این نوع خدمات می‌باشد.

رایانامه نویسنده پاسخگو: nafise_langari@birjand.ac.ir

- 1- Obfuscation
- 2- Stealth
- 3- Phishing

[۱۹-۱۸]، درخت رگرسینون افزایشی بیزین [۱۸]، جنگل تصادفی [۱۷ و ۲۰] اشاره کرد.

روش های مبتنی بر الگوریتم های فوق ابتکاری [۱۶]:

الگوریتم های فوق ابتکاری، یکی از انواع الگوریتم های بهینه سازی هستند که دارای راه کارهای برون رفت از بهینه محلی بوده و قابل کاربرد در طیف گسترده ای از مسائل هستند. رده های گوناگونی از این نوع الگوریتم ها در دهه های اخیر توسعه یافته است. از روش های فوق ابتکاری که در تشخیص فیشینگ مورد استفاده قرار گرفته اند، می توان به الگوریتم کلونی مورچه ها^{۱۳} (ACO) [۲۱ و ۲۳]، بهینه سازی گروه ذرات^{۱۴} (PSO) [۲۲-۲۱]، الگوریتم غذایابی باکتری^{۱۵} (BFOA) [۲۱] و الگوریتم خفاش بهبودیافته^{۱۶} (MBAT) [۲۳ و ۲۱] اشاره کرد.

مزیت اصلی استفاده از روش های مبتنی بر لیست سیاه، پیاده سازی راحت آن می باشد [۱]. این ابزارها اغلب وابسته به مرورگر خاصی هستند و نمی توانند بر روی چند مرورگر نصب شوند [۱]. روش های مبتنی بر لیست سیاه، معمولاً همه وبگاه های فیشینگ را شناسایی نمی کنند [۴]. در واقع لیست سیاه باید به روز رسانی شود، در غیر این صورت نمی تواند جدیدترین تهدیدات فیشینگ را شناسایی کند. روش های ابتکاری بر خلاف روش مبتنی بر لیست سیاه، وبگاه های فیشینگ جدید را تشخیص می دهند و مشکل روش های مبتنی بر لیست سیاه را ندارند [۸].

در روش های ابتکاری، در صورتی که فیشر، وبگاه فیشینگ را متفاوت از وبگاه قانونی ایجاد کند با شکست مواجه می شوند و نرخ دقت بسیار کاهش می یابد [۱۱]. همچنین روش های مبتنی بر یادگیری ماشین و داده کاوی دارای چالش های پیچیدگی در یادگیری و حجم محاسباتی بالا و به تبع عدم موفقیت در کاهش زمان عملیاتی هستند [۱۳]. روش های مبتنی بر الگوریتم های فوق ابتکاری دارای پیچیدگی کم می باشند و اغلب به سرعت به جواب می رسند [۱۶]. ساختار ساده، کاربرد آن ها در طیف گسترده ای از مسائل بهینه سازی و نیز امکان بررسی هم زمان حجم عظیمی از وبگاه ها در مسئله شناسایی وبگاه های فیشینگ، دلیل استقبال محققان از الگوریتم های فوق ابتکاری می باشد. از این جهت این روش ها مشکل روش های یادگیری ماشین و داده کاوی را نخواهد داشت و در زمان کوتاه تری به جواب می رسد. لذا در این تحقیق

ابزارهای ضد فیشینگ مبتنی بر لیست سیاه [۱]: استفاده از

ابزارهای ضد فیشینگ در مرورگرها، یک روش برای تشخیص وبگاه های فیشینگ است. این ابزارها براساس ویژگی هایی از قبیل طول آدرس اینترنتی [۲]، محبوبیت وبگاه [۳]، طول مدتی که وبگاه ثبت شده و همچنین جستجوی وبگاه در لیست سیاه [۱]، وبگاه فیشینگ را شناسایی و در صورت برخورد آن فعالیت های کاربر را مسدود کرده و به او هشدار می دهند. از ابزارهای ضد فیشینگ می توان به ابزار Net Craft [۱]، EarthLink و Cloud mark [۴] اشاره کرد.

روش حل مبتنی بر تکنیک های داده کاوی [۵]: از روش های

مبتنی بر داده کاوی می توان به روش داده کاوی فازی^۲ و طبقه بندی انجمنی^۳ اشاره کرد. روش داده کاوی فازی، Maher Aburrous [۲] از ترکیب تکنیک های داده کاوی و سیستم های فازی، به منظور بررسی بانک های اینترنتی که در معرض خطر وبگاه های فیشینگ هستند، استفاده کرده است و از طریق استخراج ۲۷ ویژگی، سایت فیشینگ را شناسایی می کند. در این روش از تعدادی تکنیک طبقه بندی داده کاوی مانند PART^۴، Prism و C4.5 استفاده شده است. در روش طبقه بندی انجمنی Moh'd Iqbal AL Ajlouni [۳] از تکنیک های CBA^۵ و MCAR^۶ برای شناسایی وبگاه فیشینگ در بانک داری اینترنتی استفاده شده است.

روش حل مبتنی بر تکنیک های ابتکاری [۷-۶]: این

روش ها میزان تشابه ویژگی های سایت فیشینگ و قانونی را آنالیز می کنند، در صورتی که میزان تشابه بیش از یک آستانه از پیش تعیین شده باشد، وبگاه، فیشینگ تشخیص داده می شود. از روش های ابتکاری، می توان به روش مبتنی بر تشابه بصری^۸ [۸]، الگوریتم Link Guard [۹]، آنالیز تصویر و مشخصات سایت^۹ [۱۰]، روش مبتنی بر تشابه طرح^{۱۰} [۱۱]، روش مبتنی بر EMD^{۱۱} [۱۲] اشاره کرد.

روش حل مبتنی بر تکنیک های یادگیری ماشین: روش های

مبتنی بر یادگیری ماشین طبق تجربیاتی که به دست می آورند، وبگاه فیشینگ را شناسایی می کنند. از روش های مبتنی بر یادگیری ماشین می توان به روش مبتنی بر عصبی- فازی^{۱۲} [۱۳]، روش طبقه بندی رگرسینون لجیستیک [۱۴-۱۵]، روش پیشنهادی pagesafe [۱۶]، رگرسینون منطقی [۱۷]، ماشین بردار پشتیبان

9- Site Characteristics And Image Analysis

10- Layout-Similarity

11- Earth mover's distance

12- Neuro-fuzzy

13- Ant Colony Optimization

14- Particle Swarm Optimization

15- Bacteria Foraging Algorithm

16- Modified Bat algorithm

1- Black list

2- Fuzzy data mining

3- Associative classification

4- Projective adaptive resonance theory

5- Classification based on association rules

6- Multi-class classification based on association rules

7- Heuristic

8- Visual similarity

منبع و جاوا اسکریپت، سبک و محتوای صفحه و نوار آدرس وب‌گاه جستجو کرد. در جدول (۱) لیستی از ویژگی‌های قابل استنباط از معیارهای اشاره شده با توجه به تحقیقات انجام گرفته ارائه شده است.

جدول (۱). انواع ویژگی‌های یک وب‌گاه

معیارها	شاخص‌های فیشینگ
URL و هویت دامنه [۲۵-۲۷، ۲]	استفاده از آدرس IP، درخواست URL، URL غیر طبیعی، نام دامنه URL
امنیت و رمزگذاری [۲۵-۲۷، ۲]	استفاده از گواهی‌نامه، مجوز صدور گواهی‌نامه، کوکی غیرطبیعی، گواهی‌نامه‌های مهم (DN)
کد منبع و جاوا اسکریپت [۲۵-۲۷، ۲]	تغییر مسیر صفحات، استفاده از onmouseover برای مخفی کردن لینک، کنترل فرم سرور (SFH)
سبک و محتوای صفحه [۲۵-۲۷، ۲]	اشتباهات املایی، استفاده از فرم‌های ثبت نام، مرتبه صفحه، بلوک‌های عکس (ویژگی‌های آن شامل رنگ و اندازه تصویر می‌باشد)، فونت، رنگ پس‌زمینه، چیدمان متن و فاصله بین خطوط.
نوار آدرس وب‌گاه [۲۵-۲۷، ۲]	آدرس طولانی URL، جایگزینی شخصیت‌های مشابه برای URL، اضافه کردن یک پیشوند یا پسوند، استفاده از نماد @، استفاده از کد کاراکتر هگزادسیمال

۲-۲- مسئله طبقه‌بندی وب‌گاه‌ها

طبقه‌بندی یک مسئله داده‌کاوی برای پیدا کردن مدلی است که کلاس‌های داده‌ای را متمایز می‌کند، با این هدف که بتوان از این مدل برای پیش‌بینی کلاس یا اشیایی که برچسب کلاس آن‌ها ناشناخته می‌باشد استفاده نمود. این مسئله از نوع مسائل با نظارت^۱ می‌باشد. این بدان معناست که باید نوع کلاس داده‌های به‌کار گرفته‌شده در مسئله طبقه‌بندی، از قبل مشخص شده باشند. در فرایند این مسئله، داده‌های به دو بخش آموزش و آزمون تقسیم می‌شوند. داده‌های آموزش جهت شناسایی و تنظیم حدود طبقات به‌کار گرفته‌شده و از داده‌های آزمون جهت ارزیابی و محاسبه کیفیت طبقه‌بندی که در مرحله آموزش ترسیم شده است، استفاده می‌شود. نهایتاً، مدل آموزش‌یافته و آزمون‌شده با یک میزان دقت مشخص، می‌تواند جهت پیش‌بینی کلاس یا اشیایی که برچسب کلاس آن‌ها ناشناخته می‌باشد، مورد استفاده قرار گیرد. در این مقاله هدف طبقه‌بندی داده در سه کلاس وب‌گاه‌های قانونی، وب‌گاه‌های مشکوک و وب‌گاه‌های فیشینگ می‌باشد. این مهم درحالی پیگیری می‌شود که ۱۵ ویژگی تشکیل‌دهنده ابعاد فضای جواب بوده است.

الگوریتم بهینه‌سازی صفحات شیب‌دار [۲۴] به‌عنوان یکی از روش‌های فوق‌ابتکاری جدید، با توجه به مزایای قابل توجه خود، همچون هم‌گرایی سریع، عدم توقف در بهینه‌های محلی و کاهش حجم محاسباتی، برای اولین بار در شناسایی وب‌گاه فیشینگ و به‌منظور طبقه‌بندی وب‌گاه‌ها در یکی از طبقات فیشینگ، قانونی و مشکوک مورد توجه و استفاده قرار گرفته است.

در اکثر روش‌های ارائه‌شده، ویژگی‌هایی برای شناسایی وب‌گاه فیشینگ در نظر گرفته‌شده، که اثر چندانی در تشخیص فیشینگ ندارند و فیشرها به آن‌ها توجه چندانی نداشته‌اند [۳، ۵، ۲۱ و ۲۳]. این ویژگی‌ها، علاوه بر کاهش کارآمدی فرایند طبقه‌بندی، حجم محاسباتی را نیز بالا می‌برند. در این تحقیق، جهت رفع این مشکل، ویژگی‌هایی به‌منظور ارزیابی وب‌گاه فیشینگ در نظر گرفته شده‌اند، که بیشترین اثر را در تشخیص وب‌گاه فیشینگ دارند و فیشرها به آن‌ها توجه بسیاری داشته‌اند.

برای تحقق هدف شناسایی وب‌گاه فیشینگ، سازماندهی مقاله به‌صورت زیر ارائه گردیده است: تشریح مسئله شناسایی وب‌گاه فیشینگ در بخش ۲ مورد بحث قرار گرفته است. در بخش ۳ به بیان روش پیشنهادی پرداخته شده است. در بخش ۴ بر روی بررسی نتایج به‌دست‌آمده از پیاده‌سازی روش پیشنهادی برای شناسایی وب‌گاه‌های فیشینگ تمرکز شده است. در بخش ۵، خلاصه و نتیجه‌گیری این تحقیق تشریح شده است.

۲- تشریح مسئله شناسایی وب‌گاه فیشینگ

مسئله شناسایی وب‌گاه‌های فیشینگ در این تحقیق، نوعی مسئله طبقه‌بندی با معیارهای خاص خود می‌باشد. طبقه‌بندی نیز خود وابسته به ویژگی‌های انتخابی وب‌گاه‌ها تعریف می‌گردد. لذا در این بخش، نخست بر روی تشریح ویژگی‌های یک وب‌گاه به‌عنوان بستر تعریف مسئله تحقیق تمرکز کرده و سپس مسئله طبقه‌بندی وب‌گاه‌ها، و درنهایت، انواع معیارهای ارزیابی تشریح گردیده است.

۲-۱- ویژگی‌های یک وب‌گاه

هر وب‌گاه دارای یک سری ویژگی‌های منحصربه‌فرد است که باعث تمایز آن از سایر وب‌گاه‌ها و حتی تمایز با وب‌گاه‌های بسیار متشابه خود می‌شود. لذا همواره فیشر در ساخت یک وب‌گاه جعلی، به‌صورت آگاهانه یا ناآگاهانه، اقدام به تغییر برخی از ویژگی‌های وب‌گاه‌های قانونی، از حالت اولیه‌شان می‌نماید. این ویژگی‌ها را می‌توان در پنج پارامتر، URL و هویت دامنه، امنیت و رمزگذاری، کد

بوده است. حال به کمک این روش کدگذاری، به هر ناحیه یک کد باینری منحصر به فرد اطلاق خواهد شد. با قرار دادن داده‌های مختلف در معادله (۱) هر طبقه‌بندی، می‌توان به آسانی عضویت آن داده را به یکی از کدهای تعریف شده به طول k را شناسایی کرد. فرض کنید تعداد کل داده‌های به کار گرفته شده در بخش آموزش طبقه‌بندی برابر با N باشند. آن‌گاه مشخص شود که مثلاً $N_{011...0}$ داده در ناحیه $011...0$ قرار دارد. اگر از تعداد $N_{011...0}$ داده، φ_j داده متعلق به کلاس C_j باشد ($j=1,2,\dots,r$)، آن‌گاه ناحیه $011...0$ متعلق به کلاسی است که بیشترین فراوانی را داشته باشد. یعنی، اگر $\varphi_n = \max(\varphi_1, \varphi_2, \dots, \varphi_r)$ ، ناحیه $011...0$ متعلق به کلاس h خواهد بود. حال، برای محاسبه Miss طبقه‌بندی ایجاد شده توسط k طبقه‌بند، کافی است، Miss هر ناحیه یا تعداد داده‌هایی که کلاس آن‌ها به کلاس اطلاق داده شده به آن ناحیه همخوانی ندارد را محاسبه کرد. نهایتاً، جمع Miss کلیه نواحی، مقدار تابع تناسب تعریف شده در معادله (۲) را نتیجه خواهد داد.

۲-۳- معیارهای ارزیابی

به منظور ارزیابی کیفی طبقه بندی وب‌گاه‌ها، هفت معیار ارزیابی در این پژوهش مورد توجه قرار گرفته که شامل معیارهای زیر می‌باشند:

۲-۳-۱- نرخ دقت تشخیص: نرخ دقت تشخیص وب‌گاه فیشینگ از طریق معادله (۳) محاسبه می‌شود.

$$\text{Acc}\% = 100 - \frac{\text{Miss}}{N_{\text{total}}} \times 100 \quad (3)$$

در معادله (۳)، N_{total} ، تعداد کل وب‌گاه‌ها می‌باشد.

۲-۳-۲- نرخ خطا: نرخ خطا در تشخیص وب‌گاه‌ها از طریق معادله (۴) محاسبه می‌شود

$$\text{Err}\% = \frac{\text{Miss}}{N_{\text{total}}} \times 100 \quad (4)$$

در معادله (۴)، N_{total} ، تعداد کل وب‌گاه‌ها می‌باشد.

۲-۳-۳- زمان تشخیص: زمان تشخیص، زمان متوسط برای شناسایی هر وب‌گاه به‌عنوان یک معیار برای سرعت در نظر گرفته شده است.

$$T_{\text{detect}} = \frac{T}{N_{\text{total}}} \quad (5)$$

در معادله (۵)، T ، زمان تشخیص همه وب‌گاه‌ها و N_{total} ، تعداد کل وب‌گاه‌هاست.

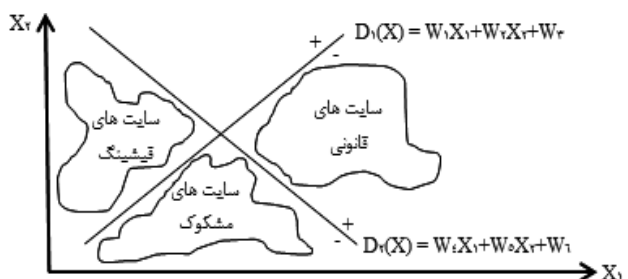
با مفروض قرار دادن n ویژگی در فضای جواب، تابع تصمیم‌گیری خطی طبقه‌بندی وب‌گاه‌ها به صورت زیر فرموله می‌گردد [۲۸].

$$D(X) = W_1 X_1 + W_2 X_2 + \dots + W_n X_n + W_{n+1} \quad (1)$$

در معادله (۱)، بردار ویژگی در فضای n بعدی و $(W_1, W_2, \dots, W_{n+1})$ بردار ضرایب می‌باشد. وظیفه یک طبقه‌بندی کننده محاسبه بردار ضرایب ابرصفحه‌ها می‌باشد. در شکل ۱، نمونه‌ای از طبقه‌بندی وب‌گاه‌ها در سه کلاس براساس ۲ ویژگی و ۲ ابر صفحه نمایش داده شده است. جهت طبقه‌بندی وب‌گاه‌ها، فضای ویژگی جستجو می‌شود تا یک مجموعه از ابر صفحات با حداقل کردن تابع تناسب تعریف شده در معادله (۲) به دست آید.

$$F(W) = \text{Miss} \quad (2)$$

Miss، تعداد نقاط داده به اشتباه طبقه‌بندی شده می‌باشد. اما با افزایش تعداد ابرصفحه‌های طبقه‌بند، تعداد نواحی تولید شده از تلاقی این ابرصفحه به صورت نمایی افزایش یافته و محاسبه Miss با مشکل روبرو می‌شود. برای عبور از این مشکل، راه حل تکنیکی، کدگذاری معنایی نواحی تولید شده می‌باشد که به شرح ذیل طرح‌ریزی می‌شود.



شکل (۱). نمونه‌ای از طبقه‌بندی وب‌گاه‌ها با ۲ ویژگی و ۲ ابرصفحه

فرض کنید، تعداد k طبقه‌بند (C_1, C_2, \dots, C_k) ، وظیفه طبقه‌بندی r کلاس C_1, C_2, \dots, C_r را برعهده داشته باشند. تعداد نواحی تولید شده توسط k طبقه‌بند برابر با 2^k می‌باشد. همچنین، مقدار k باید در نامساوی $2^k \leq r$ ، صادق باشد. از طرفی، کرانه‌ها یا دیواره‌های هر ناحیه توسط یکی از نامساوی‌های $D_i(X) \geq 0$ و یا $D_i(X) < 0$ در آن $i=1, 2, \dots, k$ تعیین می‌شوند. اگر برای هر ناحیه، به‌ازاء هر کرانه $D_i(X) \geq 0$ کد ۱ و به‌ازاء هر کرانه $D_i(X) < 0$ کد صفر در نظر گرفته شود، آن‌گاه برای هر ناحیه یک کد باینری به طول k خواهیم داشت. برای مثال، کد $011...0$ بدین معناست که $D_1(X) < 0, D_2(X) \geq 0, D_3(X) \geq 0, \dots, D_k(X) < 0$

ویژگی‌های مبتنی بر ناهنجاری^۲ و ویژگی‌های مبتنی بر جاوا اسکریپت و HTML بخش‌بندی می‌شوند که در (جدول ۲) قابل مشاهده می‌باشند. ویژگی‌های استخراج شده از اهمیت خاصی برخوردار هستند و مقالات [۲، ۶، ۲۵ و ۳۱] به پراهمیت بودن آن‌ها در تشخیص وب‌گاه فیشینگ اشاره کرده‌اند. از این‌رو، به ارزیابی وب‌گاه مبتنی بر این تعداد ویژگی اکتفا شده است.

جدول (۲). ویژگی‌های استخراج شده جهت طبقه‌بندی وب‌سایت‌ها و شناسایی وب‌سایت‌های فیشینگ

معیارهای در نظر گرفته شده	معیارهای در نظر گرفته شده
تغییر مسیر صفحات، استفاده از On Mouse Over جهت پنهان کردن لینک، غیر فعال کردن کلیک راست، استفاده از pop-up	ویژگی‌های مبتنی بر جاوا اسکریپت و HTML
درخواست URL، URL of Anchor، کنترل فرم سرور (SFH) ^۳ ، URL غیر طبیعی	ویژگی‌های مبتنی بر آبنرمال
آدرس IP، طول URL، استفاده از @، پسوند یا پیشوند مجزا شده توسط " - دامنه و زیر دامنه، HTTPS (پروتکل انتقال ابر متن)، SSL (لایه سوکت امن)	ویژگی‌های مبتنی بر نوار آدرس

۲-۳- الگوریتم بهینه‌سازی صفحات شیب‌دار

بر روی سطح زمین هر جسم مرتفع دارای یک انرژی پتانسیل متناسب با ارتفاع خود از سطح زمین است که متاثر از نیروی گرانش زمین مایل است ضمن از دست دادن این انرژی، به حداقل انرژی ممکن برسد. با در نظر گرفتن این واقعیت فیزیکی، الگوریتم بهینه‌سازی صفحات شیب‌دار (IPO)^۴ [۲۴]، بر اساس دینامیک حرکت اجسام گرد بر روی سطوح شیب‌دار بدون اصطکاک بیان شده است. هر عامل هوشمند این الگوریتم، یک گوی کوچک^۵ می‌باشد که در زمان t در موقعیت x و با ارتفاع h شناسایی می‌شود. سطح شیب‌دار بدون اصطکاک به‌عنوان فضای جستجوی عامل شامل تپه‌ها، دره‌ها و شانه‌های متعددی می‌باشد که گوی‌ها با شتاب‌های ثابت، براساس موقعیت شان نسبت به گوی‌های دیگر در ارتفاع پایین‌تر، به‌دنبال کاهش ارتفاع خود می‌باشند شکل (۲). در این الگوریتم منظور از زمان، اشاره به فرایند تکراری و تکاملی الگوریتم داشته و موقعیت عامل (x)، نمایش‌دهنده جزئیات کمی و پارامتریک عامل می‌باشد. کیفیت عامل در این الگوریتم پارامتر ارتفاع (h) می‌باشد. لذا با پیاده‌سازی این عامل بر روی یک مسئله کمینه‌سازی می‌توان این‌گونه بیان داشت که موقعیت عامل در زمان t، معادل جزئیات یک

۲-۳-۴- نرخ مثبت کاذب، نرخ منفی کاذب، نرخ مثبت درست، نرخ منفی درست: نرخ مثبت کاذب، نرخ منفی کاذب، نرخ مثبت قانونی که به نادرستی به‌عنوان وب‌گاه فیشینگ طبقه‌بندی می‌شود و نرخ منفی کاذب، نرخ منفی قانونی طبقه‌بندی می‌شوند. نرخ مثبت کاذب توسط معادله (۶)، نرخ منفی کاذب توسط معادله (۷)، نرخ مثبت درست توسط معادله (۸) و نرخ منفی درست توسط معادله (۹) به‌دست می‌آیند.

$$FP = \frac{N_{LP}}{N_{LL} + N_{LP}} \quad (۶)$$

$$FN = \frac{N_{PL}}{N_{PP} + N_{PL}} \quad (۷)$$

$$TP = \frac{N_{PP}}{N_{PP} + N_{PL}} \quad (۸)$$

$$TN = \frac{N_{LL}}{N_{LL} + N_{LP}} \quad (۹)$$

در معادلات (۶) و (۹)، N_{LP} ، تعداد وب‌گاه قانونی که به‌عنوان فیشینگ طبقه‌بندی شده است و N_{LL} ، تعداد وب‌گاه قانونی که به‌درستی طبقه‌بندی شده است، می‌باشند. در معادلات (۷) و (۸)، N_{PL} ، تعداد وب‌گاه فیشینگ که به‌عنوان قانونی طبقه‌بندی شده است و N_{PP} ، تعداد وب‌گاه فیشینگ که به‌درستی طبقه‌بندی شده است، می‌باشند.

۳- روش پیشنهادی

در این مقاله یک رویکرد شناسایی هوشمند وب‌گاه فیشینگ طی مراحل استخراج ویژگی‌های وب‌گاه و طبقه‌بندی با الگوریتم بهینه‌سازی صفحات شیب‌دار، طراحی شده است. لذا در ۳ زیربخش، ابتدا به تشریح نحوه استخراج ویژگی‌ها پرداخته شده است و سپس تشریح فرایند الگوریتم بهینه‌سازی صفحات شیب‌دار و نهایتاً نحوه طبقه‌بندی با بهره‌گیری از این الگوریتم قسمت‌های پایانی این بخش را در برخواهند گرفت.

۳-۱- استخراج ویژگی

جهت استخراج ویژگی‌های وب‌گاه از ابزار Googbar [۳۱] که به زبان جاوا اسکریپت و PHP ایجاد شده، استفاده شده است. این ابزار بر روی مرورگر موزیلا نصب می‌شود و با فراخوانی هر وب‌گاه، ویژگی‌های آن را استخراج می‌کند. ویژگی‌هایی که از طریق این ابزار استخراج شدند به سه معیار از ویژگی‌های مبتنی بر نوار آدرس،

2- Abnormal based features

3- Server form handler

4- Inclined Planes system Optimization

5- Tiny balls

1- Perl Hypertext Preprocessor

$$U(w) = \begin{cases} 1 & w > 0 \\ 0 & w \leq 0 \end{cases} \quad (13)$$

اگر گویی که دارای کمترین ارتفاع ($\min_{1 \leq i \leq K} f(X_i^t)$) در زمان t ام باشد با $X_{best}^t = (x_{best,1}^t, x_{best,2}^t, \dots, x_{best,n}^t)$ نمایش داده شود، آن گاه سرعت ثابت گوی t ام در بعد d ام در زمان t ام به کمک فرمول زیر تعیین می گردد:

$$v_{id}^t = \frac{x_{best,d}^t - x_{id}^t}{\Delta t}, \quad d=1,2,\dots,n, \quad i=1,2,\dots,K \quad (14)$$

حال با هدف کمینه نمودن ارتفاع گوی ها، موقعیت جدید گوی i در بعد d ام که با x_{id}^{t+1} نمایش داده می شود بر اساس حرکت با شتاب ثابت a_{id}^t و سرعت v_{id}^t به صورت زیر محاسبه می گردد:

$$x_{id}^{t+1} = \theta_1 R_1 a_{id}^t \Delta t^2 + \theta_2 R_2 v_{id}^t \Delta t + x_{id}^t \quad (15)$$

$$, \quad d=1,2,\dots,n, \quad i=1,2,\dots,K$$

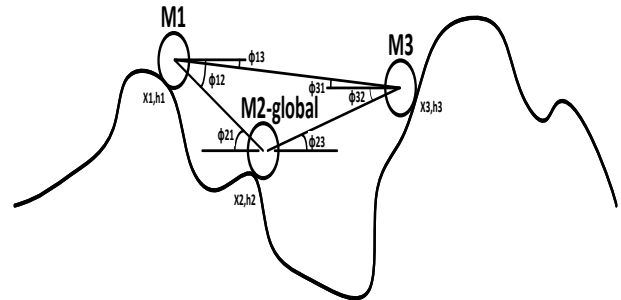
در معادله (۱۵)، R_1 و R_2 ضرایب وزنی به صورت تصادفی در بازه $[0,1]$ بوده و ضرایب θ_1 و θ_2 ، ضرایب متغیر با زمان برای کنترل بهتر عملکرد الگوریتم IPO می باشند. با تغییر موقعیت هر گوی، مجدداً میزان ارتفاع آن تعیین و فرایند فوق الذکر تکرار می شود. این روند برای T بار تکرار الگوریتم اجرا می گردد، در نهایت بهترین گوی در تکرار T ام به عنوان پاسخ الگوریتم استخراج می شود.

۳-۳- طبقه بندی با الگوریتم بهینه سازی صفحات

شیب دار

مسئله طبقه بندی وبسایت ها با ۱۵ ویژگی تشریح شده در بخش ۱-۳ و ۳-۳ کلاس وبسایت های قانونی، مشکوک و فیشینگ که توسط ۴ طبقه بندی به عنوان مناسب ترین تعداد طبقه بندی به صورت سعی و خطا به دست آمده، طراحی می گردد. منظور از سعی و خطا، انتخاب تصادفی یک نمونه کوچک از مسئله تحقیق (در این مورد، انتخاب تصادفی ۱۰۰ وب گاه با ۸ ویژگی) می باشد، که با مقادیر مختلف پارامترهای الگوریتم اجرا می شود. بهترین نتایج با توجه به کیفیت جواب و زمان محاسبه، به منظور پیاده سازی بروری مسئله اصلی انتخاب می شوند. در این تحقیق، به منظور تنظیم مقادیر پارامترهای مسئله و الگوریتم مانند: تعداد ابرصفحه ها، تعداد جمعیت اولیه و شرط خاتمه الگوریتم، مقادیر متعددی در پیاده سازی مسئله نمونه و الگوریتم مورد نظر قرار گرفتند. در نهایت، نتایج مناسب مربوط به ۴ طبقه بندی، برای طبقه بندی مسئله شناسایی وب گاه فیشینگ، ۵۰ گویی برای اندازه جمعیت الگوریتم و ۱۰۰۰ تکرار به عنوان شرط خاتمه الگوریتم انتخاب شدند. همان طور که در بخش های قبلی

جواب مسئله n ($X^t = (x_1^t, x_2^t, \dots, x_n^t)$) بعدی در تکرار t ام بوده و ارتفاع عامل در زمان t ، نمایش دهنده کیفیت ارزیابی جواب ($f(X^t)$) مسئله در تکرار t ام الگوریتم می باشد.



شکل (۲) - مثالی از فضای جستجو به همراه ۳ گوی

برای توصیف دقیق تر الگوریتم IPO، تعداد K گوی به صورت تصادفی بر روی فضای جستجو (سطح شیب دار) n بعدی را در نظر بگیرید. موقعیت جاری گوی i ام در زمان t ام به صورت زیر نمایش داده می شود:

$$X_i^t = (x_{i1}^t, x_{i2}^t, \dots, x_{in}^t), \quad 1 \leq i \leq K \quad (10)$$

هدف الگوریتم کمینه نمودن ارتفاع این گوی ها ($f(X_i^t)$) می باشد. برای رسیدن به این هدف، در گام نخست زاویه بین دو گوی i و j که زاویه خط راست واصل بین این دو گوی با خط افق می باشد به کمک فرمول زیر محاسبه می شود:

$$\phi_d^t(i, j) = \tan^{-1} \left(\frac{f(X_i^t) - f(X_j^t)}{x_{id}^t - x_{jd}^t} \right), \quad d=1,2,\dots,n \quad (11)$$

$$, \quad i, j=1,2,\dots,K, \quad i \neq j$$

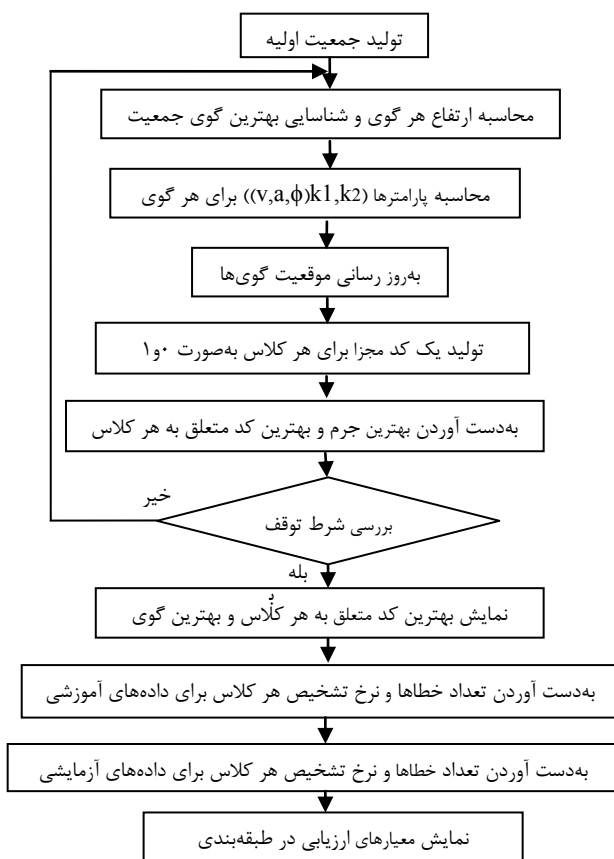
هر گوی در بعد d ام خود ($d=1,2,\dots,n$)، متناسب با زاویه اش با بعد d ام سایر گوی ها و نیز اختلاف ارتفاع آن گوی با سایر گوی ها، یک شتاب برای لغزش روی سطح شیب دار (فضای جستجو) می گیرد که به صورت زیر محاسبه می شود:

$$a_{id}^t = \sum_{\substack{j=1 \\ j \neq i}}^K U(f(X_i^t) - f(X_j^t)) \sin(\phi_d^t(i, j)), \quad (12)$$

$$d=1,2,\dots,n, \quad i=1,2,\dots,K$$

که در معادله (۱۲)، تابع $U(f(X_i^t) - f(X_j^t))$ تابعی پله ای به صورت معادله (۱۳) بوده و اجازه می دهد گوی t ام در بعد d ام خود شتابی متناسب با زاویه این گوی با سایر گوی ها بگیرد، اگر ارتفاع گوی t ام از سایر گوی ها بیشتر باشد.

صفحه $D_1(X)$ ، $D_2(X)$ ، $D_3(X)$ و $D_4(X)$ می‌باشند که در صورتی که مقادیر ویژگی‌های یک وب‌گاه مانند X' در معادلات این ابرصفحه‌ها قرار گیرد، مقادیر، بزرگ‌تر یا کوچک‌تر از صفر خواهند شد. در صورتی که مقدار بزرگ‌تر از صفر یک ابرصفحه را ۱ و مقدار کوچک‌تر از صفر یک ابرصفحه به‌ازای مقادیر ویژگی‌های یک وب‌گاه صفر کدگذاری کنیم، نواحی ۱۶ گانه با کدهای باینری به طول ۴ کدگذاری خواهند شد. مانند شکل (۵) که ناحیه ۵ با کد ۰۱۱۱ کدگذاری شده است. هر ناحیه متعلق به کلاسی خواهد بود که بیشترین سایت از آن کلاس در ناحیه مذکور قرار گیرد. مثلاً اگر ناحیه ۱۰۰۱ شامل ۸۰ وب‌گاه باشد که از این تعداد ۶۰ وب‌گاه قانونی، ۱۵ وب‌گاه مشکوک و ۵ وب‌گاه باشند، آن‌گاه این ناحیه متعلق به کلاس سایت‌های قانونی خواهد بود. همچنین Miss این ناحیه نیز برابر با ۲۰ می‌شود. به عبارت دیگر، تعلق این ناحیه به کلاس سایت‌های قانونی، بدین معناست که در صورتی که ویژگی‌های یک سایت مقدار ابرصفحه اول و چهارم را بزرگ‌تر از صفر و مقدار ابرصفحه دوم و سوم را کوچک‌تر از صفر کرد به ناحیه مذکور تعلق داشته و باید سایتی قانونی باشد، در غیر این صورت خطای طبقه‌بندی صورت گرفته است.



شکل (۴). روند طبقه‌بندی با الگوریتم بهینه‌سازی صفحات شیب‌دار

اشاره شد، وظیفه الگوریتم بهینه‌سازی صفحات شیب‌دار در طبقه‌بندی وب‌سایت‌ها تخمین مناسب بردار ضرایب معادله (۱) به‌عنوان ضرایب ابرصفحات طبقه‌بندی می‌باشد، به طوری که میزان خطا وب‌گاه‌های به اشتباه طبقه‌بندی‌شده، حداقل شود. هر گوی در فضای ۶۴ بعدی مانند شکل (۳) تعریف می‌گردد که در خود، ضرایب ۴ ابرصفحه را جای داده است. در ادامه فرایند طبقه‌بندی با کمک الگوریتم بهینه‌سازی صفحات شیب‌دار در ۱۱ گام بیان شده است. همچنین، فلوچارت این فرایند نیز در شکل (۴) قابل مشاهده می‌باشد.

W	...	W	W	...	W	W	...	W	W	..	W
1	...	16	17	...	32	33	...	48	49	..	64

شکل (۳). تعداد ابعاد یک گوی در فضای جستجو

گام ۱ (تولید جمعیت اولیه): تعداد ۵۰ گوی تولید شده است که هر کدام از گوی‌ها یک آرایه به طول ۶۴ دارد و با مقادیری که به‌طور تصادفی بین ۱- و ۱ انتخاب شده‌اند، تکمیل می‌شوند.

گام ۲ (به‌دست آوردن بهترین گوی): هر گوی در جمعیت اولیه، نمایانگر ضرایب چهار ابرصفحه طبقه‌بندی‌کننده وب‌گاه‌ها می‌باشد. لذا ارتفاع (کیفیت) هر گوی، میزان خطای حاصل از طبقه‌بندی آن گوی می‌باشد که توسط معادله (۴) محاسبه می‌گردد. گویی که کمترین خطا (ارتفاع) را دارد، بهترین گوی جمعیت اولیه انتخاب می‌شود.

گام ۳ (محاسبه پارامترهای سرعت (v)، شتاب (a)، زاویه (phi) برای هر گوی): این پارامترها را می‌توان از طریق معادلات (۱۱، ۱۲ و ۱۴) که در بخش ۳-۲ تشریح شد، به‌دست آورد.

گام ۴ (محاسبه موقعیت جدید هر گوی): براساس شتاب و سرعت هر گوی، موقعیت جدید آن محاسبه می‌شود. این امر را می‌توان از طریق معادله (۱۵) که در بخش ۳-۲ ارائه شد، به‌دست آورد.

شایان ذکر است که گویی که طبقه‌بندی متناظر آن، خطای کمتری داشته باشد از کیفیت بهتری (ارتفاع کمتری) برخوردار خواهد بود. برای محاسبه خطای طبقه‌بندی هر گوی (محاسبه ارتفاع) گام ۵ طراحی گردیده است.

گام ۵ (ارزیابی موقعیت هر گوی): هر گوی در برگیرنده ۴ ابرصفحه می‌باشد که فضای مسئله را حداکثر به ۱۶ ناحیه تقسیم می‌کند. علت بیان حداکثر، وجود احتمال تقاطع بیش از دو ابرصفحه در یک نقطه در فضای ویژگی‌ها می‌باشد. ضرایب یک گوی، ضرایب چهار ابر

جدول (۳). نتایج به دست آمده از شناسایی وب گاه فیشینگ توسط الگوریتم بهینه سازی صفحات شیب دار

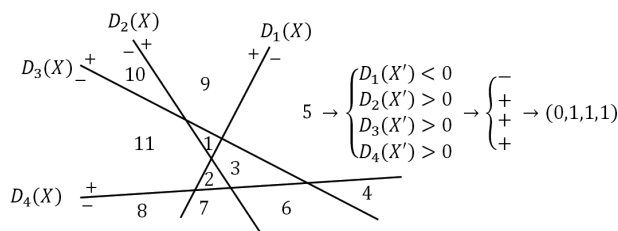
ارزیابی وب گاه مبتنی بر ۱۵ ویژگی	معیار ارزیابی
٪۹۷/۰۸	نرخ تشخیص وب گاه فیشینگ
٪۹۸	نرخ تشخیص وب گاه قانونی
٪۹۳/۰۸	نرخ تشخیص وب گاه مشکوک
۱۰۰۰	تعداد کل داده ها
۴۴	تعداد داده هایی که اشتباه طبقه بندی شده اند
۹۵۶	تعداد داده هایی که درست طبقه بندی شده اند
۰/۷۹۰۷	زمان تشخیص الگوریتم

جدول (۴). نتایج پیاده سازی الگوریتم IPO در تشخیص وب گاه های

معیار ارزیابی	مبتنی بر ۲۷ ویژگی			
	MBAT	BFOA	ACO	PSO
نرخ دقت تشخیص	٪۹۷/۹۸	٪۹۱/۳۸	٪۸۸/۹۲	٪۹۲/۱۳
نرخ خطا	٪۲/۰۲	٪۸/۶۲	٪۱۱/۰۸	٪۷/۸۷
زمان تشخیص (بر حسب ثانیه)	۰/۸۲۲۸۰	۰/۸۲۱۱	۱/۲۴۵۴۲	۱/۱۶۹۰۳
شرط خاتمه	۱۰۰	۱۰۰	۱۰۰	۱۰۰

به منظور پیاده سازی فرایند طبقه بندی داده های محک، ۱۰۰۰ وب گاه کاندید برای طبقه بندی به دو دسته داده های آموزشی، مشتمل بر ۶۰۰ وب گاه و داده های آزمون، مشتمل بر ۴۰۰ وب گاه تقسیم بندی شده اند. نتایج به دست آمده از داده های آموزشی و آزمایشی به طور جداگانه محاسبه شده است و متوسط این نتایج به عنوان خروجی الگوریتم در جدول (۳) قابل مشاهده است.

مقایسه ارزیابی نتایج به دست آمده اجرای الگوریتم IPO در شناسایی وب گاه های فیشینگ با سایر الگوریتم های فوق ابتکاری معتبر در جدول (۴) صورت پذیرفته است. در این مقایسه نتایج روش های فوق ابتکاری معتبری همچون الگوریتم بهینه سازی ازدحام ذرات (PSO)، الگوریتم بهینه سازی کلونی مورچگان (ACO)، الگوریتم بهینه سازی غذاییابی باکتری (BFOA) و الگوریتم اصلاح شده خفاش (MBAT) در شناسایی وب گاه های فیشینگ مورد توجه قرار گرفته است. نتایج مقایسه شده الگوریتم IPO از میانگین ارزیابی داده های آموزشی و داده های آزمون به دست آمده است. لازم به ذکر است که در طبقه بندی وب گاه ها توسط الگوریتم IPO، پارامتر



شکل (۵). نحوه کدگذاری کلاس ها

گام ۶ (محاسبه بهترین گوی): گویی که کمترین خطا در نتیجه طبقه بندی ایجاد شده از ضرایب آن را تولید کند، بهترین گوی (کم ارتفاع ترین گوی) و طبقه بندی متناظر آن، بهترین طبقه بندی انتخاب می شود.

گام ۷ (بررسی شرط توقف): در گام ۷ شرط توقف الگوریتم که تعداد تکرارهای الگوریتم در نظر گرفته شده بررسی می شود.

گام ۸: نمایش بهترین کد متعلق به هر کلاس و بهترین گوی

گام ۹ و ۱۰: به دست آوردن توابع هدف برای داده های آموزشی و آزمایشی

گام ۱۱: نمایش معیارهای ارزیابی در طبقه بندی وب گاه ها

۴- نتایج شبیه سازی روش پیشنهادی

پیاده سازی این ایده به کمک نرم افزار متلب^۱ در محیط سیستم عامل ویندوز ۷ و پردازشگر Intel Core i3 و RAM 4GB انجام گرفته است. نتایج حاصله جهت ارزیابی وب گاه فیشینگ در جدول (۳) ارائه شده است. این نتایج، با نتایج روش های فوق ابتکاری دیگر که توسط [۲۱، ۲۲، ۲۳] به دست آمده، در جدول (۴)، مقایسه شده است. پایگاه داده مورد استفاده در [۲۱، ۲۲، ۲۳] برای الگوریتم های PSO، ACO و MBAT شامل ۱۶۵۲ وب گاه بوده است، که ۱۰۸۰ وب گاه برای آموزش و ۵۷۲ وب گاه برای آزمون به کار گرفته شده اند. این داده های محک برای BFOA برابر ۱۰۰۰ وب گاه در نظر گرفته شده که ۱۰۰ وب گاه برای آموزش و ۹۰۰ وب گاه ی دیگر برای آزمون انتخاب می شوند. داده های محک این مقاله از مجموعه داده های <http://www.phishtank.com> که شامل لیست url وب گاه های فیشینگ بوده و به طور منظم به روز رسانی می شوند مستخرج گردیده است. تعداد وب گاه های فیشینگ، وب گاه های قانونی و وب گاه های مشکوک در داده های محک به ترتیب ۵۰۰ و ۳۰۰ و ۲۰۰ می باشد. این داده ها در تاریخ ۱۱ جولای ۲۰۱۴ تا ۱۸ جولای ۲۰۱۴ و ۲۱ و ۲۲ آگوست ۲۰۱۴ جمع آوری شده است.

جدول (۶). نمونه‌ای از طبقه‌بندی وب‌گاه‌ها

طبقه‌بندی	مشکوک	قانونی	فیشینگ	آدرس سایت
True			✓	http://66.45.253.74/~fnwi/padyyme/pnconlinenowjamesjames/pnconlinenowjamesjames/Homepage.do.htm
False			✓	http://bp.@secure.com.ru/ubanking.ch/deu.html?login&locale=de-CH
True			✓	https://69.57.245.62/css/US_Bank/hhi/U_S_%20Bank%20Internet%20Banking.htm
False			✓	https://s3rvmail.com/baneco/ebank_login/
True		✓		https://www.central-bank.org.tt/
False	✓			http://198.100.148.13/cgi-sys/suspendedpage.cgi?login
True	✓			https://www.bbt.com/bbt.com/online-services/online-banking/online-banking-overview.page
False		✓		https://creditcardforum.com/blog/american-express-prepaid-card-review/
True		✓		https://prepaid.bankofamerica.com/EddCard/Pages/Home.aspx
True			✓	http://update-confirmation-4c106bb@338610fa0a332778de9303b6c.luizlozano.net/dd68b9dc0940101aef7c60b95f42acd3/
True			✓	http://www.techbrasil.net/bankofamerica.com.update.sys.login.in/updat.sys.hey.bro.here/Sitkey.Signon.do/prospect.php?_nfpb=login&_pageLabel=page_logonform
True			✓	https://ebanking.capital-g.com.coluccio.ca/CorporateBankingWeb44/Core/Login.aspx.html?EsetProtoscanCtx=9f4db50
True	✓			https://fb@ecard.bsl.org.au/qw/2121signinebyeswseBayISAPIdllSignInUsingSSL1pUserldco_partnerId2siteid186ru=http3A2F2Fmyebyes2Fws2FeBayISAPIdll3F.php
True		✓		https://www.bank-of-algeria.dz/
True	✓			http://www.eblinko.com/yep/index.html
False			✓	http://ecard.bsl.org.au/qw/2121signinebyeswseBayISAPIdllSignInUsingSSL1pUserldco_partnerId2siteid186ru=http3A2F2Fmyebyes2Fws2FeBayISAPIdll3F.php
True	✓			http://www.ezonebanking.com/category/banking/

زمان تشخیص، بیان‌کننده زمان آموزش مدل طبقه‌بندی وب‌گاه‌ها بعد از ۱۰۰۰ بار تکرار الگوریتم (شرط توقف الگوریتم) است، به طوری این پارامتر در الگوریتم‌های دیگر، بعد از ۱۰۰ بار تکرار الگوریتم (شرط توقف الگوریتم) حاصل شده است. با عنایت به این که شرط توقف الگوریتم IPO، ده برابر بزرگ‌تر از الگوریتم‌های مقایسه‌شده می‌باشد، انتظار می‌رود که زمان تشخیص این الگوریتم نیز عددی حدود ۱۰ برابر بزرگ‌تر را نشان دهد. اما کمتر بودن زمان تشخیص الگوریتم IPO با وجود دارا بودن تعداد تکرارهای بیشتر، عاملی بر طرح این ادعا می‌گردد که زمان تشخیص الگوریتم IPO به مراتب از دیگر الگوریتم‌ها کمتر می‌باشد.

میانگین نرخ دقت تشخیص داده‌های آموزشی و آزمون، در الگوریتم IPO براساس ۱۵ ویژگی به‌طور قابل توجهی از نتایج روش‌های PSO، ACO و BFOA با وجود در نظر گرفتن ۲۷ ویژگی، بهتر و دقیق‌تر بوده و تنها اختلاف کوچکی با الگوریتم MBAT مبتنی بر ۲۷ ویژگی دارد. این برتری برای نرخ خطا نیز مشاهده می‌شود ولی در پیچیدگی زمانی، برتری متعلق به الگوریتم IPO با در نظر گرفتن ۱۵ ویژگی، در فرایند طبقه‌بندی وب‌گاه‌ها می‌باشد.

نتایج به‌دست آمده از ارزیابی پارامتر نرخ مثبت کاذب، نرخ منفی کاذب، نرخ مثبت درست، نرخ منفی درست، در تشخیص وب‌گاه فیشینگ توسط الگوریتم IPO در (جدول ۵) آمده است. برای ارزیابی این پارامترها ۲۰۰ وب‌گاه فیشینگ و ۱۲۰ وب‌گاه قانونی مورد بررسی قرار گرفته‌اند. مطالعه جداول سه‌گانه این بخش، نشان می‌دهد که اگرچه شناسایی وب‌گاه‌های فیشینگ با تعداد ویژگی زیاد، نرخ دقت تشخیص را افزایش می‌دهد ولی الگوریتم IPO با در نظر گرفتن تعداد ویژگی‌های کمتر، نرخ دقت تشخیص مناسبی را در زمان به مراتب کمتری داشته باشد. در (جدول ۶) نمونه‌ای از طبقه‌بندی وب‌گاه‌ها قابل مشاهده است.

جدول (۵). نرخ مثبت کاذب و منفی کاذب در تشخیص وب‌گاه فیشینگ توسط الگوریتم بهینه‌سازی صفحات شیب‌دار

به‌عنوان طبقه‌بندی شده	به‌عنوان وب‌گاه فیشینگ طبقه‌بندی شده	به‌عنوان طبقه‌بندی شده
۱۱ FN = ۰/۰۵۵	۱۸۹ TN = ۰/۹۴۵	وب‌گاه فیشینگ
۱۱۵ TP = ۰/۹۵۸	۵ FP = ۰/۰۴۱	وب‌گاه قانونی

۵- نتیجه گیری

یکی از جدیدترین تهدیدات امنیتی در فضای مجازی، سرقت اطلاعات شخصی و مالی افراد توسط فیشرفرمان می‌باشد. روش‌های متنوعی در شناسایی وب‌گاه فیشرفرمان مورد بررسی و تحلیل قرار گرفته‌اند. در روش‌های موجود، به‌طور هم‌زمان به طول عمر کوتاه وب‌گاه‌های فیشرفرمان، کاهش حجم محاسبات و امکان تحلیل و کنترل حجم گسترده‌ای از وب‌گاه‌ها توجه نشده است. لذا در این مقاله به‌منظور تحقق هم‌زمان سه پارامتر یادشده و ایجاد ابزاری کارآمد برای مجموعه‌های نظارتی در حوزه بانک‌داری الکترونیکی، رویکرد جدیدی ارائه و پیاده‌سازی شد.

در این رویکرد، ابتدا ۱۵ ویژگی اثرگذار و حساس به جعل، که خواسته یا ناخواسته توسط فیشرفرمان دستخوش تغییر می‌گردند، انتخاب و استخراج می‌گردد. گروه وب‌گاه‌های هدف مطالعه، براساس ویژگی‌های مستخرج‌شده توسط الگوریتم بهینه‌سازی صفحات شیب‌دار به سه طبقه وب‌گاه‌های قانونی، مشکوک و فیشرفرمان طبقه‌بندی شدند. همان‌طور که اشاره شد، الگوریتم بهینه‌سازی صفحات شیب‌دار برای اولین بار در حل مسئله تحقیق مورد ارزیابی قرار گرفت. این مهم براساس جستجو برای تعیین حدود ۴ ابرصفحه طبقه‌بند در ابعاد ۱۶ بعدی صورت پذیرفت. براساس نتایج قابل مشاهده در جدول (۳) و مقایسه با بهترین روش‌های موجود جدول (۴)، الگوریتم IPO با نرخ تشخیص ۹۶/۵۶٪ و در زمان کوتاه ۰/۷۹۰۷ ثانیه قادر به طبقه‌بندی گروه وب‌گاه‌های کاندید مطالعه در سه طبقه قانونی، فیشرفرمان، مشکوک شد. این تحقیق راه را برای ورود به شناسایی خودکار وب‌گاه فیشرفرمان بر روی مقیاس بزرگی از سایت‌ها به‌طور آگاه به زمینه می‌تواند باز نماید. از جمله افق‌های پیش روی این تحقیق می‌توان به ایجاد بستر شناسایی پویا جهت انتخاب انعطاف‌پذیری ویژگی‌های گروه وب‌گاه‌های هدف مختلف، و پیاده‌سازی طبقه‌بندی فازی توسط الگوریتم‌های فوق‌ابتکاری جدید، جهت افزایش نرخ دقت تشخیص در مسائل دنیای واقعی اشاره نمود.

۶- منابع

- [4] L. F. Cranor, S. Egelman, J. I. Hong, and Y. Zhang, "Phishing Phish: An Evaluation of Anti-Phishing Toolbars," in NDSS, 2007.
- [5] M. Sirajuddin, "Data Mining Approach for Deceptive Phishing Detection System," ijsret, vol. 2, pp. 337-334, 2013.
- [6] E. Medvet, E. Kirida, and C. Kruegel, "Visual-similarity-based phishing detection," in Proceedings of the 4th international conference on Security and privacy in communication networks, p. 22, 2008.
- [7] W. Zhang, H. Lu, B. Xu, and H. Yang, "Web phishing detection based on page spatial layout similarity," Informatica, vol. 37, pp. 231-244, 2013.
- [8] L. Wenyin, G. Huang, L. Xiaoyue, Z. Min, and X. Deng, "Detection of phishing webpages based on visual similarity," in Special interest tracks and posters of the 14th international conference on world wide web, pp. 1060-1061, 2005.
- [9] S. T. Kumar, V. Kumar, and A. Kumar, "Detection and Prevention of Phishing Attacks Using Linkguard Algorithm," 2008.
- [10] J. S. White, J. N. Matthews, and J. L. Stacy, "A method for the automated detection phishing websites through both site characteristics and image analysis," in SPIE Defense, Security and Sensing, pp. 84080B-84080B-11, 2012.
- [11] A. P. Rosiello, E. Kirida, C. Kruegel, and F. Ferrandi, "A layout-similarity-based approach for detecting phishing pages," in Security and Privacy in Communications Networks and the Workshops, 2007. Secure Comm 2007. Third International Conference on, pp. 454-463, 2007.
- [12] N. R. T. Guhan, "Analyzing and Detecting Phishing Webpages with Visual Similarity Assessment Based on Earth Movers Distance with Linear Programming Model," International Journal of Advanced Engineering Technology, vol. III, pp. 327-330, 2012.
- [13] P. Barraclough, M. Hossain, M. Tahir, G. Sexton, and N. Aslam, "Intelligent phishing detection and protection scheme for online transactions," Expert systems with applications, vol. 40, pp. 4697-4706, 2013.
- [14] A. Demaris and S. H. Selman, "Logistic regression," in Converting Data into Evidence, ed: Springer, pp.115-136, 2013.
- [15] S. Garera, N. Provos, M. Chew, and A. D. Rubin, "A framework for detection and measurement of phishing attacks," in Proceedings of the 2007 ACM workshop on Recurring malware, pp. 1-8, 2007.
- [16] P. Sengar and V. Kumar, "Client-side defense against phishing with pagesafe," International Journal of Computer Applications, vol. 4, pp. 6-10, 2010.
- [17] S. S. Abu-Nimeh, D. Nappa, X. Wang, and S. Nair, "A comparison of machine learning techniques for phishing detection," in Proceedings of the anti-phishing working groups 2nd annual crime researchers summit, pp. 60-69, 2007.
- [1] Y. Zhang, S. Egelman, L. Cranor, and J. Hong, "Phishing phish: Evaluating anti-phishing tools," 2006.
- [2] M. Aburrous, M. A. Hossain, K. Dahal, and F. Thabtah, "Intelligent phishing detection system for e-banking using fuzzy data mining," Expert systems with applications, vol. 37, pp. 7913-7921, 2010.
- [3] M. D. I. A. Ajlouni, W. E. Hadi, and J. Alwedyan, "Detecting Phishing Websites Using Associative Classification," European Journal of Business and Management, vol. 5, pp. 36-40, 2013.

- [26] M. Aburrou, M. A. Hossain, K. Dahal, and F. Thabatah, "Modelling Intelligent Phishing Detection System for e-Banking using Fuzzy Data Mining," in *Cyber Worlds, 2009. CW'09. International Conference on*, pp. 265-272, 2009.
- [27] P. Barraclough, M. Hossain, M. Tahir, G. Sexton, and N. Aslam, "Intelligent phishing detection and protection scheme for online transactions," *Expert Systems with Applications*, 2013.
- [28] S. H. Zahiri and S. A. Seyedin, "Intelligent Particle Swarm Classifiers," *Iranian journal of electrical and computer engineering*, vol. 4, p. 63, 2015.
- [29] M. Aburrou, M. A. Hossain, K. Dahal, and F. Thabtah, "Experimental case studies for investigating e-banking phishing techniques and attack strategies," *Cognitive Computation*, vol. 2, pp. 242-253, 2010.
- [30] A. Y. Fu, L. Wenyn, and X. Deng, "Detecting phishing web pages with visual similarity assessment based on earth mover's distance (EMD)," *Dependable and Secure Computing, IEEE Transactions on*, vol. 3, pp. 301-311, 2006.
- [31] R. Mohammad, T. McCluskey, and F. A. Thabtah, "Intelligent Rule based Phishing Websites Classification," *IET Information Security*, 2013.
- [18] J. M. De-Sa, "Pattern recognition: concepts, methods, and applications," Springer, 2001.
- [19] H. M. Deylami and Y. P. Singh, "Cybercrime detection techniques based on support vector machines," *Artificial Intelligence Research*, vol. 2, 2013.
- [20] L. Breiman, "Random forests," *Machine learning*, vol. 45, pp. 5-32, 2001.
- [21] D. M. L. V. Radha Damodaram, "Experimental Study on Meta Heuristic Optimization Algorithms for Fake Website Detection," *International Association of Scientific Innovation and Research (IASIR)*, vol. 2, pp. 43-53, 2012.
- [22] M. Radha Damodaram and M. Valarmathi, "Phishing Website Detection and Optimization Using Particle Swarm Optimization Technique," *International Journal of Computer Science and Security (IJCSS)*, vol. 5, p. 477, 2011.
- [23] M. R. Damodaram and M. Valarmathi, "Bacterial Foraging Optimization for Fake Website Detection," *International Journal of Computer Science & Applications (TIJCSA)*, vol. 1, 2013.
- [24] M. H. Mozaffari, H. Abdy, and S. H. Zahiri, "Application of inclined planes system optimization on data clustering," in *Pattern Recognition and Image Analysis (PRIA), 2013 First Iranian Conference on*, pp. 1-3, 2013.
- [25] M. Aburrou, M. Hossain, K. Dahal, and F. Thabtah, "Associative classification techniques for predicting e-banking phishing websites," in *Multimedia Computing and Information Technology (MCIT), 2010 International Conference on*, pp. 9-12, 2010.