مجله علمی- پژوهشی "پدافند الکترونیکی و سایبری"
سال سوم، شماره ۱، بهار ۱۳۹۴، ص ۴۰-۲۹

شناسایی وب‌گاه فیشینگ در بانک‌داری اینترنتی با استفاده از الگوریتم بهینه‌سازی صفحات شیب‌دار

نفیسه لنگری[1*]، مجید عبدالرزاق‌نژاد[2]
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 به‌منظور شناسایی وب‌گاه‌های فیشینگ در بانک‌داری الکترونیک براساس استخراج ویژگی های 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

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 به‌راحتی 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[3]، آشکارترین نوع حمله روی این نوع خدمات می‌باشد.

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 این‌که عواقب مالی داشته باشد. درهمین حال، خدمات اینترنتی تکمیل‌تر 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.

رایانامه نویسنده پاسخگو: nafise_langari@birjand.ac.ir

1- Obfuscation
2- Stealth
3- Phishing

ابزارهای ضدفیشینگ مبتنی بر لیست سیاه [۱]: استفاده از ابزارهای ضدفیشینگ در مرورگرها، یک روش برای تشخیص وب‌گاه‌های فیشینگ است. این ابزارها براساس ویژگی‌هایی از قبیل طول آدرس اینترنتی [۲]، محبوبیت وب‌گاه [۳]، طول مدتی که وب‌گاه ثبت‌شده و همچنین جستجوی وب‌گاه در لیست سیاه[1] [۱]، وب‌گاه فیشینگ را شناسایی و درصورت برخورد آن فعالیت‌های کاربر را مسدود کرده و به او هشدار می‌دهند. از ابزارهای ضدفیشینگ می‌توان به ابزار calling Id، Net Craft [۱]، EarthLink و Cloud mark [۴] اشاره کرد.

روش حل مبتنی بر تکنیک‌های داده‌کاوی [۵]: از روش‌های مبتنی‌بر داده‌کاوی می‌توان به روش داده‌کاوی فازی[2] و طبقه‌بندی انجمنی[3] اشاره کرد. روش داده‌کاوی فازی، Maher Aburrous [۲] از ترکیب تکنیک‌های داده‌کاوی و سیستم‌های فازی، به‌منظور بررسی بانک‌های اینترنتی که در معرض خطر وب‌گاه های فیشینگ هستند، استفاده کرده است و از طریق استخراج ۲۷ ویژگی، سایت فیشینگ را شناسایی می‌کند. در این روش از تعدادی تکنیک طبقه‌بندی داده کاوی مانند PART[4]، Prism و C4.5 استفاده شده است. در روش طبقه‌بندی انجمنی Moh'd Iqbal AL Ajlouni [۳] از تکنیک‌های CBA[5] و MCAR[6] برای شناسایی وب‌گاه فیشینگ در بانک‌داری اینترنتی استفاده شده است.

روش حل مبتنی بر تکنیک‌های ابتکاری[7] [۶-۷]: این روش‌ها میزان تشابه ویژگی‌های سایت فیشینگ و قانونی را آنالیز می‌کنند، درصورتی‌که میزان تشابه بیش از یک آستانه از پیش تعیین شده باشد، وب‌گاه، فیشینگ تشخیص داده می‌شود. از روش‌های ابتکاری، می‌توان به روش مبتنی بر تشابه بصری[8] [۸]، الگوریتم Link Guard [۹]، آنالیز تصویر و مشخصات سایت[9] [۱۰]، روش مبتنی بر تشابه طرح[10] [۱۱]، روش مبتنی بر[11] EMD [۱۲] اشاره کرد.

روش حل مبتنی بر تکنیک‌های یادگیری ماشین: روش‌های مبتنی بر یادگیری ماشین طبق تجربیاتی که به‌دست می‌آورند، وب‌گاه فیشینگ را شناسایی می‌کنند. از روش‌های مبتنی بر یادگیری ماشین می‌توان به روش مبتنی بر عصبی- فازی[12] [۱۳]، روش طبقه‌بندی رگرسیون لجیستیک [۱۵-۱۴]، روش پیشنهادی pagesafe [۱۶]، رگرسیون منطقی [۱۷]، ماشین‌بردار پشتیبان [۱۹ -۱۸]، درخت رگرسیون افزایشی بیزین [۱۸]، جنگل تصادفی [۱۷ و ۲۰] اشاره کرد.

روش‌های مبتنی بر الگوریتم‌های فوق ابتکاری [۱۶]: الگوریتم‌های فوق ابتکاری، یکی از انواع الگوریتم‌های بهینه‌سازی هستند که دارای راه‌کارهای برون رفت از بهینه محلی بوده و قابل کاربرد در طیف گسترده‌ای از مسائل هستند. رده‌های گوناگونی از این نوع الگوریتم‌ها در دهه‌های اخیر توسعه یافته است. از روش‌های فوق ابتکاری که در تشخیص فیشینگ مورد استفاده قرارگرفته‌اند، می‌توان به الگوریتم کلونی مورچه‌ها[13] (ACO) [۲۱ و ۲۳]، بهینه‌سازی گروه ذرات[14] (PSO) [۲۲ -۲۱]، الگوریتم غذایابی باکتری[15] (BFOA) [۲۱] و الگوریتم خفاش بهبودیافته[16] (MBAT) [۲۱ و ۲۳] اشاره کرد.

مزیت اصلی استفاده از روش‌های مبتنی بر لیست سیاه، پیاده‌سازی راحت آن می‌باشد [۱]. این ابزارها اغلب وابسته به مرورگرخاصی هستند و نمی‌توانند بر روی چند مرورگر نصب شوند [۱]. روش‌های مبتنی بر لیست سیاه، معمولاً همه وب‌گاه‌های فیشینگ را شناسایی نمی‌کنند [۴]. در واقع لیست سیاه باید به‌روز رسانی شود، درغیر این‌صورت نمی‌تواند جدیدترین تهدیدات فیشینگ را شناسایی کند. روش‌های ابتکاری بر خلاف روش مبتنی‌بر لیست سیاه، وب‌گاه‌های فیشینگ جدید را تشخیص می‌دهند و مشکل روش‌های مبتنی بر لیست سیاه را ندارند [۸].

در روش‌های ابتکاری، درصورتی‌که فیشر، وب‌گاه فیشینگ را متفاوت از وب‌گاه قانونی ایجاد کند با شکست مواجه می‌شوند و نرخ دقت بسیار کاهش می‌یابد [۱۱]. همچنین روش‌های مبتنی بر یادگیری ماشین و داده کاوی دارای چالش‌های پیچیدگی در یادگیری و حجم محاسباتی بالا و به‌تبع عدم موفقیت در کاهش زمان عملیاتی هستند [۱۳]. روش‌های مبتنی بر الگوریتم‌های فوق ابتکاری دارای پیچیدگی کم می‌باشند و اغلب به سرعت به جواب می‌رسند [۱۶]. ساختار ساده، کاربرد آن‌ها در طیف گسترده‌ای از مسائل بهینه‌سازی و نیز امکان بررسی هم‌زمان حجم عظیمی از وب‌گاه‌ها در مسئله شناسایی وب‌گاه‌های فیشینگ، دلیل استقبال محققان از الگوریتم‌های فوق ابتکاری می‌باشد. از این جهت این روش‌ها مشکل روش‌های یادگیری ماشین و داده‌کاوی را نخواهد داشت و در زمان کوتاه‌تری به جواب می‌رسد. لذا در این تحقیق

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
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

الگوریتم بهینه‌سازی صفحات شیب‌دار[۲۴] به‌عنوان یکی از روش‌های فوق ابتکاری 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 طبقه بندی، 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 ارایه 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 برروی 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 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 [۲، ۲۷-۲۵]	استفاده از آدرس IP، درخواست URL، URL غیر طبیعی، نام دامنه URL
امنیت و رمزگذاری [۲، ۲۷-۲۵]	استفاده از گواهی‌نامه، مجوز صدور گواهینامه، کوکی غیرطبیعی، گواهی‌نامه‌های مهم (DN)
کد منبع و جاوا اسکریپت [۲، ۲۷-۲۵]	تغییر مسیر صفحات، استفاده از onMouseOver برای مخفی کردن لینک، کنترل فرم سرور (SFH)
سبک و محتوای صفحه [۲، ۲۷-۲۵]	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 [۲، ۲۷-۲۵]	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[1] 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. نهایتا، 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.

1- Supervise

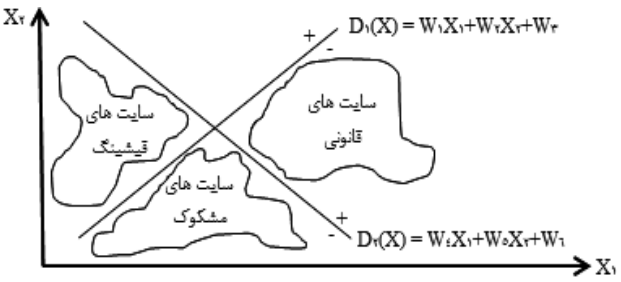
با مفروض قرار دادن n ویژگی در فضای جواب، تابع تصمیم‌گیری خطی طبقه‌بندی وب‌گاه‌ها به‌صورت زیر فرموله می‌گردد [۲۸].

$$D(X)=W_1X_1+W_2X_2+\ldots+W_nX_n+W_{n+1} \quad (۱)$$

در معادله (۱)، $(X_1,X_2,\ldots,X_n)$ بردار ویژگی در فضای n بعدی و $(W_1,W_2,\ldots,W_{n+1})$ ، بردار ضرایب می‌باشد. وظیفه یک طبقه‌بندی‌کننده 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) به‌دست آید.

$$F(W)=Miss \quad (۲)$$

Miss، تعداد نقاط داده به اشتباه طبقه‌بندی‌شده می‌باشد. اما با افزایش تعداد ابرصفحه‌های طبقه‌بند، تعداد نواحی تولیدشده از تلاقی این ابرصفحه به‌صورت 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 $D_1(X),D_2(X),\ldots,D_k(X)$، وظیفه طبقه‌بندی r کلاس $C_1,C_2,\ldots,C_r$ را برعهده داشته باشند. تعداد نواحی تولید شده توسط k طبقه‌بند برابر با 2^k می‌باشد. همچنین، مقدار k باید در نامساوی $r \le 2^k$، صادق باشد. از طرفی، کرانه‌ها یا دیواره‌های هر ناحیه توسط یکی از نامساوی‌های $D_i(X)\ge 0$ و یا $D_i(X)<0$ در آن $i=1,2,\ldots,k$، تعیین می‌شوند. اگر برای هر ناحیه، به‌ازاء هر کرانه $D_i(X)\ge 0$ کد ۱ و به‌ازاء هر کرانه $D_i(X)<0$ کد صفر در نظر گرفته شود، آن‌گاه برای هر ناحیه یک کد باینری به‌طول k خواهیم داشت. برای مثال، کد ۰۰...۰۱۱ بدین معناست که $D_1(X)<0$، $D_2(X)\ge 0$، $D_3(X)\ge 0$، ... و $D_k(X)<0$ بوده است. حال به کمک این روش کدگذاری، به هر ناحیه یک کد باینری منحصربه‌فرد اطلاق خواهد شد. با قرار دادن داده‌های مختلف در معادله (۱) هر طبقه‌بندی، می‌توان به آسانی عضویت آن داده را به یکی از کدهای تعریف‌شده به طول k را شناسایی کرد. فرض کنید تعداد کل داده‌های به‌کارگرفته‌شده در بخش آموزش طبقه‌بندی برابر با N باشند. آن‌گاه مشخص شود که مثلا $N_{011\ldots0}$ داده در ناحیه ۰...۰۱۱ قرار دارد. اگر از تعداد $N_{011\ldots0}$ داده، φ_j داده متعلق به کلاس C_j باشد $(j=1,2,\ldots,r)$، آن‌گاه ناحیه ۰...۰۱۱ متعلق به کلاسی است که بیشترین فراوانی را داشته باشد. یعنی، اگر $\varphi_h=\max(\varphi_1,\varphi_2,\ldots,\varphi_r)$ ، ناحیه ۰...۰۱۱ متعلق به کلاس h خواهد بود. حال، برای محاسبه Miss طبقه‌بندی ایجادشده توسط k طبقه‌بند، کافی است، Miss هر ناحیه یا تعداد داده‌هایی که کلاس آن‌ها به کلاس اطلاق داده‌شده به آن ناحیه همخوانی ندارد را محاسبه کرد. نهایتا، جمع Miss کلیه نواحی، مقدار تابع تناسب تعریف شده در معادله (۲) را نتیجه خواهد داد.

۲-۳- معیارهای ارزیابی

به‌منظور ارزیابی کیفی طبقه بندی وب‌گاه ها، 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.

$$\mathrm{Acc\%}=100-\frac{\mathrm{Miss}}{\mathrm{N_{total}}}\times 100 \quad (۳)$$

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

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

$$\mathrm{Err\%}=\frac{\mathrm{Miss}}{\mathrm{N_{total}}}\times 100 \quad (۴)$$

در معادله (۴)، $\mathrm{N_{total}}$، 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.

$$\mathrm{T_{detect}}=\frac{\mathrm{T}}{\mathrm{N_{total}}} \quad (۵)$$

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

۲-۳-۴- نرخ مثبت کاذب، نرخ منفی کاذب، نرخ مثبت درست، نرخ منفی درست: نرخ مثبت کاذب، نرخی از وب‌گاه قانونی که به نادرستی به‌عنوان وب‌گاه فیشینگ طبقه‌بندی می‌شود و نرخ منفی کاذب، نرخی از وب‌سایت‌های فیشینگ که اشتباهاً به‌عنوان وب‌گاه قانونی طبقه‌بندی می‌شوند، می‌باشند. 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 با زبان جاوا اسکریپت و PHP[1] ایجادشده، 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 نوارآدرس، ویژگی‌های مبتنی بر ناهنجاری[2] 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
ویژگی‌های مبتنی بر جاوا اسکریپت و HTML	تغییر مسیر صفحات، استفاده از On Mouse Over جهت پنهان کردن لینک، غیر فعال کردن کلیک راست، استفاده از pop- up
ویژگی‌های مبتنی بر آبنرمال	درخواست URL، URL of Anchor، کنترل فرم سرور (SFH)[3]، URL غیر طبیعی
ویژگی‌های مبتنی بر نوار آدرس	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)

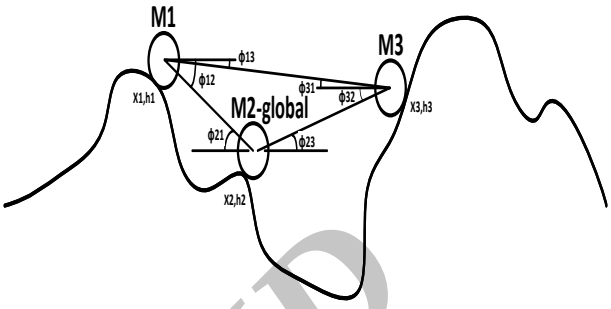
۳-۲- الگوریتم بهینه‌سازی صفحات شیب‌دار

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)[4] [۲۴]، بر اساس دینامیک حرکت اجسام گرد بر روی سطوح شیب‌دار بدون اصطکاک بیان شده است. هر عامل هوشمند این الگوریتم، یک گوی کوچک[5] 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 می باشد. 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

1- Perl Hypertext Preprocessor

2- Abnormal based features
3- Server form handler
4– Inclined Planes system Optimization
5– Tiny balls

جواب مسئله n ($X^t=(x_1^t,x_2^t,\ldots,x_n^t)$) بعدی در تکرار t ام بوده و ارتفاع عامل در زمان t، نمایش‌دهنده کیفیت ارزیابی جواب ($f(X^t)$) مسئله در تکرار t ام الگوریتم می‌باشند.



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

برای توصیف دقیق‌تر الگوریتم IPO، تعداد K گوی به‌صورت تصادفی بر روی فضای جستجو (سطح شیبدار) n بعدی را در نظر بگیرید. موقعیت جاری گوی i ام در زمان t ام به‌صورت زیر نمایش داده می شود:

$$X_i^t=\left(x_{i1}^t,x_{i2}^t,\ldots,x_{in}^t\right) \quad ,1\le i\le K \qquad (۱۰)$$

هدف الگوریتم کمینه نمودن ارتفاع این گوی‌ها ($f(X_i^t)$) می‌باشد. 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 می‌باشد به کمک فرمول زیر محاسبه می شود:

$$\phi_d^t(i,j)=\tan^{-1}\left(\frac{f\left(X_i^t\right)-f\left(X_j^t\right)}{x_{id}^t-x_{jd}^t}\right) \ , \ d=1,2,\ldots,n$$
$$, \ i,j=1,2,\ldots,K \quad , \ i\ne j \qquad (۱۱)$$

هر گوی در بعد d ام خود ($d=1,2,\ldots,n$)، متناسب با زاویه‌اش با بعد d ام سایر گوی‌ها و نیز اختلاف ارتفاع آن گوی با سایر گوی‌ها، یک شتاب برای لغزش روی سطح شیب‌دار (فضای جستجو) می‌گیرد که به‌صورت زیر محاسبه می‌شود :

$$a_{id}^t=\sum_{\substack{j=1\\ j\ne i}}^{K}U\left(f\left(X_i^t\right)-f\left(X_j^t\right)\right)\sin\left(\phi_d^t(i,j)\right),$$
$$d=1,2,\ldots,n \quad , \quad i=1,2,\ldots,K \qquad (۱۲)$$

که در معادله (۱۲)، تابع $U(f(X_i^t)-f(X_j^t))$، تابعی پله‌ای به‌صورت معادله (۱۳) بوده و اجازه می‌دهد گوی i ام در بعد d ام خود شتابی متناسب با زاویه این گوی با سایر گوی‌ها بگیرید، اگر ارتفاع گوی i ام از سایر گوی‌ها بیشتر باشد.

$$U(w)=\begin{cases}1 & w>0\\ 0 & w\le 0\end{cases} \qquad (۱۳)$$

اگر گویی که دارای کمترین ارتفاع ($\min_{1\le i\le K} f(X_i^t)$) در زمان t ام باشد با $X_{best}^t=\left(x_{best,1}^t,x_{best,2}^t,\ldots,x_{best,n}^t\right)$ نمایش داده شود، آن‌گاه سرعت ثابت گوی i ام در بعد d ام در زمان t ام به کمک فرمول زیر تعیین می‌گردد:

$$v_{id}^t=\frac{x_{best,d}^t-x_{id}^t}{\Delta t} \quad , \ d=1,2,\ldots,n \quad , \ i=1,2,\ldots,K \qquad (۱۴)$$

حال با هدف کمینه نمودن ارتفاع گوی‌ها، موقعیت جدید گوی i در بعد d ام که با x_{id}^{t+1} نمایش داده می‌شود بر اساس حرکت با شتاب ثابت a_{id}^t و سرعت v_{id}^t به‌صورت زیر محاسبه می‌گردد:

$$x_{id}^{t+1}=\theta_1R_1a_{id}^t\,\Delta t^2+\theta_2R_2v_{id}^t\,\Delta t+x_{id}^t$$
$$, \ d=1,2,\ldots,n \quad , \ i=1,2,\ldots,K \qquad (۱۵)$$

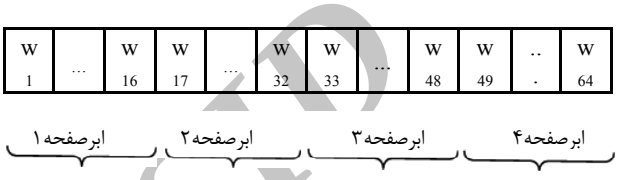
در معادله (۱۵)، R_1 و R_2 ضرایب وزنی به‌صورت تصادفی در بازه [۰،۱] بوده و ضرایب θ_1 و θ_2، ضرایب متغیر با زمان برای کنترل بهتر عملکرد الگوریتم IPO می‌باشند. با تغییر موقعیت هر گوی، مجدداً میزان ارتفاع آن تعیین و فرایند فوق الذکر تکرار می‌شود. این روند برای T بار تکرار الگوریتم اجرا می‌گردد، در نهایت بهترین گوی در تکرار T ام به‌عنوان پاسخ الگوریتم استخراج می‌شود.

۳-۳- طبقه‌بندی با الگوریتم بهینه‌سازی صفحات شیب‌دار

مسئله طبقه‌بندی وب‌سایت‌ها با ۱۵ ویژگی تشریح‌شده در بخش ۳-۱ و ۳ کلاس وب‌سایت‌های قانونی، مشکوک و فیشینگ که توسط ۴ طبقه‌بند[1] به‌عنوان مناسب‌ترین تعداد طبقه‌بند به‌صورت سعی و خطا به‌دست‌آمده، طراحی می‌گردد. منظور از سعی و خطا، انتخاب تصادفی یک نمونه کوچک از مسئله تحقیق (در این مورد، انتخاب تصادفی ۱۰۰ وب‌گاه با ۸ ویژگی) می‌باشد، که با مقادیر مختلف پارامترهای الگوریتم اجرا می‌شود. بهترین نتایج با توجه به کیفیت جواب و زمان محاسبه، به‌منظور پیاده‌سازی بروری مسئله اصلی انتخاب می‌شوند. در این تحقیق، به‌منظور تنظیم مقادیر پارامترهای مسئله و الگوریتم مانند: تعداد ابرصفحه‌ها، تعداد جمعیت اولیه و شرط خاتمه الگوریتم، مقادیر متعددی در پیاده‌سازی مسئله نمونه و الگوریتم مورد نظر قرار گرفتند. در نهایت، نتایج مناسب مربوط به ۴ طبقه‌بند، برای طبقه‌بندی مسئله شناسایی وب‌گاه فیشینگ، ۵۰ گویی برای اندازه جمعیت الگوریتم و ۱۰۰۰ تکرار به‌عنوان شرط خاتمه الگوریتم انتخاب شدند. همان‌طور که در بخش‌های قبلی

1- Classifier

اشاره شد، وظیفه الگوریتم بهینه‌سازی صفحات شیب‌دار در طبقه‌بندی وب‌سایت‌ها تخمین مناسب بردار ضرایب معادله (۱) به‌عنوان ضرایب ابرصفحات طبقه‌بندی می‌باشد، به‌طوری که میزان خطا وب‌گاه های 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 می باشد.



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 به‌طول ۶۴ دارد و با مقادیری که به‌طور تصادفی بین ۱- و ۱ انتخاب شده‌اند، تکمیل می‌شوند.

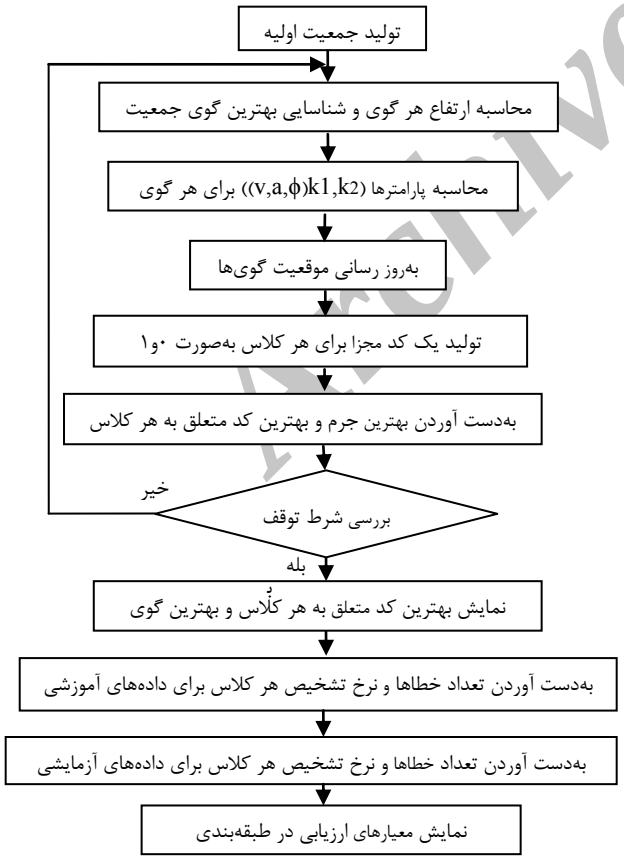
گام۲ (به‌دست آوردن بهترین گوی): هر گوی در جمعیت اولیه، نمایان‌گر ضرایب چهار ابرصفحه طبقه‌بندی‌کننده وب‌گاه‌ها می‌باشد. لذا ارتفاع (کیفیت) هر گوی، میزان خطای حاصل از طبقه‌بندی آن گوی می‌باشد که توسط معادله (۴) محاسبه می‌گردد. گویی که کمترین خطا (ارتفاع) را دارد، بهترین گوی جمعیت اولیه انتخاب می‌شود.

گام۳ (محاسبه پارامترهای سرعت (v)، شتاب(a)، زاویه(ϕ) برای هر گوی): این پارامترها را می‌توان از طریق معادلات (۱۱، ۱۲ و ۱۴) که در بخش ۳-۲ تشریح شد، به‌دست آورد.

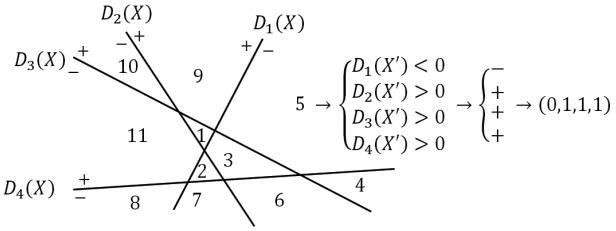
گام۴ (محاسبه موقعیت جدید هر گوی): براساس شتاب و سرعت هر گوی، موقعیت جدید آن محاسبه می‌شود. این امر را می‌توان از طریق معادله (۱۵) که‌در بخش ۳-۲ ارائه شد، به‌دست آورد.

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.

گام۵ (ارزیای 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 صفحه $D_1(X)$ $D_2(X)$ $D_3(X)$ و$D_4(X)$ می‌باشند که درصورتی‌که مقادیر ویژگی‌های یک وب‌گاه مانند X' در معادلات این ابرصفحه ها قرار گیرد، مقادیر، بزرگ‌تر یا کوچک‌تر از صفر خواهند شد. درصورتی‌که 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. مثلا 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. به‌عبارت دیگر، تعلق این ناحیه به کلاس سایت‌های قانونی، بدین معناست که درصورتی‌که 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 این‌صورت خطای طبقه‌بندی صورت گرفته است.



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



شکل (۵). نحوه کدگذاری کلاس‌ها

گام۶ (محاسبه بهترین گوی): گویی که کمترین خطا در نتیجه طبقه‌بندی ایجادشده از ضرایب آن را تولید کند، بهترین گوی (کم ارتفاع ترین گوی) و طبقه‌بندی متناظر آن، بهترین طبقه‌بندی انتخاب می‌شود.

گام۷ (بررسی شرط توقف): در گام ۷ شرط توقف الگوریتم که تعداد تکرارهای الگوریتم در نظر گرفته‌شده بررسی می‌شود.

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

گام ۹ و ۱۰: به‌دست‌آوردن توابع هدف برای داده‌های آموزشی و آزمایشی

گام ۱۱: نمایش معیارهای ارزیابی در طبقه‌بندی وب‌گاه‌ها

۴- نتایج شبیه‌سازی روش پیشنهادی

پیاده‌سازی این ایده به کمک نرم افزار متلب[1] در محیط سیستم عامل ویندوز ۷ و پردازشگر Intel Core i3 و RAM 4GB انجام گرفته است. نتایج حاصله جهت ارزیابی وب‌گاه فیشینگ در جدول (۳) ارایه شده است. این نتایج، با نتایج روش‌های فوق ابتکاری دیگر که توسط[۲۱،۲۲،۲۳] به‌دست‌آمده، در جدول (۴)، مقایسه شده است. پایگاه داده مورد استفاده در [۲۱،۲۲،۲۳] برای الگوریتم‌های PSO، ACO و MBAT شامل ۱۶۵۲وب‌گاه بوده است، که ۱۰۸۰ وب‌گاه برای آموزش و ۵۷۲ وب‌گاه برای آزمون به‌کار گرفته شده‌اند. این داده‌های محک برای BFOA برابر ۱۰۰۰ وب‌گاه درنظر گرفته‌شده که ۱۰۰ وب‌گاه برای آموزش و ۹۰۰ وب‌گاه ی دیگر برای آزمون انتخاب می‌شوند. داده‌های محک این مقاله از مجموعه داده‌های http://www.phishtank.com که شامل لیست url وب‌گاه‌های فیشینگ بوده و به‌طور منظم به‌روز رسانی می‌شوند مستخرج گردیده است. تعداد وب‌گاه‌های فیشینگ، وب‌گاه‌های قانونی و وب‌گاه‌های مشکوک در داده‌های محک به‌ترتیب ۵۰۰ و ۳۰۰ و ۲۰۰ می‌باشد. این داده‌ها در تاریخ ۱۱ جولای ۲۰۱۴ تا ۱۸ جولای ۲۰۱۴ و ۲۱ و ۲۲ آگوست ۲۰۱۴ جمع‌آوری شده است.

1– Matlab

جدول (۳). نتایج به‌دست‌آمده از شناسایی وب‌گاه فیشینگ توسط الگوریتم بهینه‌سازی صفحات شیب‌دار

معیار ارزیابی	ارزیابی وب‌گاه مبتنی بر ۱۵ ویژگی
نرخ تشخیص وب‌گاه فیشینگ	٪۹۷/۰۸
نرخ تشخیص وب‌گاه قانونی	٪۹۸
نرخ تشخیص وب‌گاه مشکوک	٪۹۳/۰۸
تعداد کل داده‌ها	۱۰۰۰
تعداد داده‌هایی که اشتباه طبقه‌بندی شده‌اند	۴۴
تعداد داده‌هایی که درست طبقه‌بندی شده‌اند	۹۵۶
زمان تشخیص الگوریتم	۰/۷۹۰۷

جدول (۴). نتایج پیاده‌سازی الگوریتم IPO در تشخیص وب‌گاه‌های

معیار ارزیابی	IPO	PSO	ACO	BFOA	MBAT
	مبتنی بر ۱۵ ویژگی	مبتنی بر ۲۷ ویژگی			
نرخ دقت تشخیص	٪۹۶/۵۶	٪۹۲/۱۳	٪۸۸/۹۲	٪۹۱/۳۸	٪۹۷/۹۸
نرخ خطا	٪۳/۴۴	٪۷/۸۷	٪۱۱/۰۸	٪۸/۶۲	٪۲/۰۲
زمان تشخیص (برحسب ثانیه)	۰/۷۹۰۷	۱/۱۶۹۰۳	۱/۲۴۵۴۲	۰/۸۲۱۱	۰/۸۲۲۸۰
شرط خاتمه	۱۰۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰

به‌منظور پیاده‌سازی فرایند طبقه‌بندی داده‌های محک، ۱۰۰۰ وب‌گاه کاندید برای طبقه‌بندی به دو دسته داده‌های آموزشی، مشتمل بر ۶۰۰ وب‌گاه و داده‌های آزمون، مشتمل بر ۴۰۰ وب‌گاه تقسیم‌بندی شده‌اند. نتایج به‌دست‌آمده از داده‌های آموزشی و آزمایشی به‌طور جداگانه محاسبه شده است و متوسط این نتایج به‌عنوان خروجی الگوریتم در جدول (۳) قابل مشاهده است.

مقایسه ارزیابی نتایج به‌دست‌آمده اجرای الگوریتم IPO در شناسایی وب‌گاه های فیشینگ با سایر الگوریتم‌های فوق ابتکاری معتبر در جدول (۴) صورت پذیرفته است. در این مقایسه نتایج روش‌های فوق ابتکاری معتبری همچون الگوریتم بهینه‌سازی ازدحام ذرات (PSO)، الگوریتم بهینه‌سازی کلونی مورچگان (ACO)، الگوریتم بهینه‌سازی غذایابی باکتری (BFOA) و الگوریتم اصلاح‌شده خفاش (MBAT) در شناسایی وب‌گاه‌های فیشینگ مورد توجه قرار گرفته است. نتایج مقایسه‌شده الگوریتم IPOاز میانگین ارزیابی داده‌های آموزشی و داده‌های آزمایشی به‌دست آمده است. لازم به ذکر است که در طبقه‌بندی وب‌گاه ها توسط الگوریتم IPO، پارامتر

زمان تشخیص، بیان‌کننده زمان آموزش مدل طبقه‌بندی وب‌گاه‌ها بعد از ۱۰۰۰بار تکرار الگوریتم (شرط توقف الگوریتم) است، به‌طوری این پارامتر در الگوریتم‌های دیگر، بعد از ۱۰۰ بار تکرار الگوریتم (شرط توقف الگوریتم) حاصل شده است. باعنایت به این‌که 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 دراد. 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 گرفته اند. 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 وب‌گاه فیشینگ طبقه‌بندی‌شده	**به‌عنوان وب‌گاه قانونی طبقه‌بندی‌شده**
وب‌گاه فیشینگ	۱۸۹ TN = ۰/۹۴۵	۱۱ FN = ۰/۰۵۵
وب‌گاه قانونی	۵ FP = ۰/۰۴۱	۱۱۵ TP = ۰/۹۵۸

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://s3rvimail.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/bbtdotcom/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/2121signinebyeswseBayISAPIdllSignInUsingSSL1pUserIdco_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/2121signinebyeswseBayISAPIdllSignInUsingSSL1pUserIdco_partnerId2siteid186ru=http3A2F2Fmyebyes2Fws2FeBayISAPIdll3F.php
True	✓			http://www.ezonebanking.com/category/banking/

۵- نتیجه‌گیری

یکی از جدیدترین تهدیدات امنیتی در فضای مجازی، سرقت اطلاعات شخصی و مالی افراد توسط فیشر می‌باشد. روش‌های متنوعی در شناسایی وب‌گاه فیشینگ مورد بررسی و تحلیل قرار گرفته‌اند. در روش‌های موجود، به‌طور همزمان به طول عمر کوتاه وب‌گاه‌های فیشینگ، 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 کارامد برای مجموعه‌های نظارتی در حوزه بانک‌داری الکترونیک، رویکرد جدیدی ارائه و پیاده‌سازی شد.

در این رویکرد، ابتدا ۱۵ ویژگی اثرگذار و حساس به جعل، که خواسته یا ناخواسته توسط فیشر دستخوش تغییر می گردند، انتخاب و استخراج می‌گردد. گروه وب‌گاه‌های هدف مطالعه، براساس ویژگی‌های مستخرج‌شده توسط الگوریتم بهینه‌سازی صفحات شیب‌دار به سه طبقه وب‌گاه‌های قانونی، مشکوک و فیشینگ طبقه‌بندی شدند. همان‌طور که اشاره شد، الگوریتم بهینه‌سازی صفحات شیب‌دار برای اولین‌بار 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 سه‌طبقه قانونی، فیشینگ، مشکوک شد. این تحقیق راه را برای ورود به شناسایی خودکار وب‌گاه فیشینگ 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 فوق ابتکاری جدید، جهت افزایش نرخ دقت تشخیص در مسائل دنیای واقعی اشاره نمود.

۶- منابع

[1] Y. Zhang, S. Egelman, L. Cranor, and J. Hong, "Phinding 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.

[4] L. F. Cranor, S. Egelman, J. I. Hong, and Y. Zhang, "Phinding 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. Kirda, and C. Kruegel, "Visual-similarity-based phishing detection," in Proceedings of the 4th international conference on Security and privacy in communication netowrks, 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. Kirda, 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 malcode, 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] SS. 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 ecrime researchers summit, pp. 60-69, 2007.

[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. Aburrous, 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.

[26] M. Aburrous, 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. Aburrous, 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. Wenyin, 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.

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.

* Corresponding Author Email: rebrahimi@guilan.ac.ir