مجله علمی - پژوهشی "پدافند الکترونیکی و سایبری"
سال چهارم، شماره ۳، پاییز ۱۳۹۵، ص ۲۶-۱۱

ارائه روشی برای شناسایی وب سایت فیشینگ سرویس پرداخت اینترنتی

غلامرضا شاه‌محمدی[1*]، سلمان کمالی‌زاده[2]

۱- دانشیار، گروه فناوری اطلاعات، دانشگاه علوم انتظامی امین

۲- دانشجوی کارشناسی‌ارشد، گروه مهندسی کامپیوتر، دانشگاه آزاد اسلامی، واحد بندرعباس

(دریافت: ۹۴/۰۹/۲۳، پذیرش: ۹۵/۰۵/۱۱)

چکیده

فیشینگ یکی از فنون مهندسی اجتماعی برای فریب کاربران است که برای کسب اطلاعات محرمانه مانند نام کاربری، گذرواژه یا اطلاعات حساب بانکی کاربران استفاده می‌شود. از مهم‌ترین چالش‌های موجود در اینترنت، خطر حملات فیشینگ و کلاهبرداری‌های اینترنتی است. ازاین‌رو پژوهشگران، تلاش‌های زیادی برای شناسایی و مقابله با این‌گونه حملات داشته‌اند. هدف این تحقیق، ارائه روش جدیدی برای شناسایی وب‌سایت فیشینگ در بانکداری اینترنتی است. روش پیشنهادی نقاط ضعف روش‌های شناسایی و مقابله با فیشینگ مانند عدم بررسی وضعیت رایگان بودن میزبانی وب‌سایت‌ها و عدم تمرکز روی بانکداری اینترنتی را ندارد. این روش از مزایای فنون مختلف شناسایی وب‌سایت فیشینگ استفاده می‌کند و امکان شناسایی لحظه صفر وب‌سایت‌های فیشینگ فارسی‌زبان که سرویس‌های بانکداری اینترنتی در ایران را موردحمله قرار می‌دهند، رادار است. روش پیشنهادی با استفاده از مجموعه وب‌سایت‌های قانونی شامل ۵۰۰ صفحه نمایه‌شده در سایت پیوندها و مجموعه وب‌سایت‌های فیشینگ شامل ۱۰۰ سایت فیشینگ شناسایی‌شده در PhishTank و ۲۰ سایت فیشینگ طراحی‌شده و ۱۱ سایت فیشینگ شناسایی‌شده در طول نگارش مقاله، ارزیابی‌شده است. نتایج ارزیابی نشان می‌دهد، روش پیشنهادی، وب‌سایت‌های قانونی را فیشینگ تلقی نمی‌کند (خطای مثبت صفر درصد) و تنها ۳٪ از وب‌سایت‌های فیشینگ را تشخیص نمی‌دهد (خطای منفی ۳٪).

واژه‌های کلیدی: فیشینگ، وب‌سایت فیشینگ، مقابله با فیشینگ، حملات اینترنتی، Phishing، Anti-Phishing

۱- مقدمه

فیشینگ یک نوع حمله کامپیوتری است که مهاجم از طریق کانال‌های ارتباط الکترونیکی[1]، با ایجاد ارتباط با انسان‌ها و با استفاده از پیام‌های مهندسی اجتماعی[2]، به ترغیب آن‌ها، برای انجام کارهایی که به نفع مهاجم است، اقدام می‌کند [۲-۱]. به‌عبارت‌دیگر، یک کلاهبرداری است که معمولاً از طریق ایمیل انجام می‌شود تا اطلاعات شخصی افراد سرقت شود [۳]. در خصوص نام‌گذاری این حمله، اکثر متخصصین معتقدند کلمه فیشینگ نمایانگر به دام انداختن قربانیان حمله، به‌مثابه ماهیگیری[3]، است [۴-۱]. کلمه فیشینگ ابتدا در سال ۱۹۹۶ استفاده گردید که فیشرها[4] با استفاده از مهندسی اجتماعی، مجوزهای کاربری را مورد حمله قرار دادند [۵-۱]. کلاهبرداران اینترنتی، از ایمیل برای صید[5] اطلاعات مالی و گذرواژه‌های کاربران اینترنتی استفاده کردند [۵-۱]. کلاهبرداری فیشینگ معمولاً از طریق ارسال ایمیل‌های فراوان که به‌ظاهر از طرف شرکت یا بانک معتبر است، شروع می‌شود و قربانیان از طریق URL[6]های موجود در ایمیل، به وب‌سایت‌هایی غیرقانونی (کاملاً شبیه وب‌سایت‌های قانونی) هدایت می‌شوند. این وب‌سایت‌های جعلی از کاربر می‌خواهد که اطلاعاتی مانند نام کاربری و رمز عبور خود را وارد کرده و ارسال نمایند.

۱-۱- ضرورت مقابله با فیشینگ

امروزه از مهم ترین چالش‌های موجود در بانکداری اینترنتی، خطر حملات فیشینگ و کلاهبرداری‌های اینترنتی است و خسارت‌های فراوانی به مشتریان و به سازمان‌ها وارد می‌کند. علاوه بر پول و زمان از دست‌رفته؛ اعتماد افراد به برنامه‌ها و خدمات برخط[7] از دست می‌رود و در نتیجه، اعتبار سازمان‌ها کم‌رنگ می‌شود. این حملات تنها در آمریکا سالیانه چندین

* رایانامه نویسنده مسئول: shah_mohammadi@yahoo.co.uk

1 -Electronic Communication Channels
2 -Social engineering
3 -Fishing
4 -Phishers
5 -Fish
6 -Uniform Resource Locator
7 -Online Services

میلیارد دلار خسارت به بار می‌آورد [۶-۵]. از این‌رو متخصصین و پژوهشگران، تلاش‌های زیادی برای شناسایی و مقابله با این‌گونه حملات داشته‌اند. ابزارهای زیادی برای شناسایی و مقابله با آن‌ها ساخته شده است، ولی از آنجایی‌که فیشرها، همواره روش کار خود را با هزینه اندک تغییر می‌دهند، نیاز به به‌روز شدن روش‌ها و ابزارهای شناسایی و مقابله با فیشینگ است. امروزه جعبه‌ابزارهای[1] فیشینگ نیز در اینترنت یافت می‌شود که کار حمله را بسیار راحت‌تر کرده است. جعبه‌ابزارهای فیشینگ، مجموعه‌ای از ابزارهاست که به یک فرد نه‌چندان ماهر، این قابلیت را می‌دهند تا به‌سادگی سایت جعلی برای حمله فیشینگ طراحی کند و جزئیات ورود[2] را در سایت خود قرار دهد. این ابزارها حاوی نرم‌افزار طراحی وب‌سایت، گرافیک، کدهای وب و متونی است که برای ساخت صفحات وب، موردنیاز است. همچنین نرم‌افزارهای ارسال هرزنامه[3]، به مهاجم امکان ارسال صدها یا هزاران نامه‌ی الکترونیکی به قربانیان را می‌دهد [۲]. به گزارش [۷]، نرخ حملات فیشینگ در سال ۲۰۱۳ افزایش‌یافته است، به‌گونه‌ای که این نرخ در سال ۲۰۱۲ از هر ۴۱۴ ایمیل، یک ایمیل فیشینگ بوده است، درحالی که در سال ۲۰۱۳ به‌ازای هر ۳۹۲ ایمیل، یک ایمیل فیشینگ گزارش‌شده است. همچنین نرخ حملات فیشینگ در ماه مه میلادی سال ۲۰۱۴ به ازای هر ۳۹۵ ایمیل، یک ایمیل فیشینگ گزارش شده است.

این تحقیق به دنبال ارائه روشی جدید برای شناسایی وب‌سایت فیشینگ سرویس پرداخت اینترنتی است. در این پژوهش پس از بررسی پژوهش‌های موجود در حوزه شناسایی و مقابله با حمله فیشینگ، ضمن ارزیابی پژوهش‌های انجام‌شده بر اساس معیارهای پیشنهادی، روش جدیدی برای شناسایی وب‌سایت فیشینگ ارائه و پیاده‌سازی می‌شود.

۲- پژوهش‌های مرتبط

برخی از روش‌های شناسایی وب‌سایت فیشینگ در جدول (۱) ارائه شده است.

۳- روش پیشنهادی

در این بخش به ارائه روش پیشنهادی پرداخته می‌شود. در ابتدا کلیات روش شرح داده می‌شود و سپس به‌طور کامل، روند انجام کار آن ارائه می‌شود. در پایان، نحوه پیاده‌سازی روش پیشنهادی تشریح می‌گردد.

۳-۱- ارائه روش پیشنهادی

وب‌سایت‌ها را می‌توان به دو قسمت زیر تقسیم کرد:

- **وب‌سایت قانونی:** وب‌سایتی که مطمئناً معتبر است و روش باید تمامی آن‌ها را به‌عنوان وب‌سایت قانونی شناسایی کند.
- **وب‌سایت فیشینگ یا کلاه‌بردار[4]:** وب‌سایتی که شباهت زیادی به وب‌سایت قانونی داشته اما جعلی است که روش باید بتواند تمامی آن‌ها را به‌عنوان فیشینگ شناسایی کند.

همچنین می‌توان فنون موجود در شناسایی وب‌سایت فیشینگ را به‌صورت زیر طبقه‌بندی کرد:

- **طبقه اول:** فنونی که جوابی برابر ۱ یا ۰ دارند (False/True). یعنی به‌طور قطع فیشینگ بودن یا قانونی بودن یک وب‌سایت را درمی‌یابند. در این روش، این فنون شامل فن لیست سفید، لیست سیاه و بررسی اطلاعات وارد شده است.
- **طبقه دوم:** فنونی که جواب آن‌ها بین ۰ تا ۱ است. یعنی درصدی را به‌عنوان احتمال برای یک وب‌سایت که میزان فیشینگ بودن آن را نشان می‌دهد در نظر می‌گیرند. در این روش، این فنون شامل بررسی ویژگی‌های صفحه [۸]، رتبه صفحه و وضعیت میزبانی وب هستند.

این دو قسمت، به‌عنوان دوطبقه درروش پیشنهادی مورداستفاده قرار می‌گیرند. همچنین، جهت بررسی نتایج هرکدام از فنون، طبقه سوم روش پیشنهادی با عنوان جمع‌بندی نهایی قرار داده‌شده است. با توجه به توضیحات داده‌شده، می‌توان فنون به‌کاررفته در این روش را به‌صورت جدول (۲) نمایش داد.

نکته قابل‌توجه درروش پیشنهادی این است که می‌توان به طبقه‌های اول و دوم، فنون متفاوت به همراه ضریب اهمیتشان اضافه کرد.

در ادامه، به تشریح فنون به کار رفته درروش پیشنهادی پرداخته می‌شود.

۳-۱-۱- جستجو در لیست سفید

اطلاعاتی که در لیست سفید قرار می‌گیرند باید بسیار دقیق و کاملاً قانونی باشند. زیرا روش پیشنهادی این لیست را ملاکی برای قانونی دانستن یک صفحه وب‌سایت در نظر می‌گیرد و همچنین برای شناسایی یک وب‌سایت فیشینگ شبیه به یکی از آن‌ها استفاده می‌کند. جمع‌آوری داده برای لیست سفید، با کسب اطلاعات مربوط به آدرس قانونی صفحات بانکداری اینترنتی

1 -Toolkit
2 -Log-in
3 -Spam
4 -Fraudulent

بانک‌های ایران صورت گرفته است.

صفحات مربوط به بانکداری اینترنتی بانک‌های فعال در ایران که اطلاعات آدرس وب آن‌ها استخراج‌شده و در لیست سفید قرار داده‌شده‌اند، مطابق جدول (۳)، شامل سه دسته زیر است:

۱. صفحات بانکداری اینترنتی
۲. صفحات سرویس پرداخت اینترنتی
۳. سایر صفحات خاص (مربوط به امور مالی و اطلاع‌رسانی)

جدول (۱). روش‌های شناسایی وب‌سایت فیشینگ

ردیف	مبتنی بر	عنوان روش	توضیحات
۱	لیست سیاه	Phish Net [۹]	این روش، محدودیت آدرس‌های دقیقاً برابر را شناسایی می‌کند. برای این کار، URLهای لیست سیاه با عنوان والد[۱] را پردازش می‌کند، سپس شکل‌های مختلف URLهای شبیه به آن را با عنوان فرزندان[۲] ایجاد می‌کند. ۵ روش اکتشافی برای انجام این کار شامل جایگزینی TLD[۳]، شباهت ساختار فهرست[۴]، هم ارزی[۵] آدرس IP، جانشینی رشته پرسش[۶] و هم ارزی نام برند[۷] می‌باشند. ارزیابی این روش نشان می‌دهد، نرخ خطای منفی برابر با ۳٪ و نرخ خطای مثبت برابر با ۵٪ است.
۲	اکتشاف	Spoof Guard [۱۰]	این روش، یک افزونه[۸] است که فعالیت‌های فیشینگ مبتنی‌بر HTTP(S) را با عنوان یک نوارابزار مرورگر وب شناسایی می‌کند. این کار با وزن‌دهی به موارد نامتعارف شناسایی‌شده در محتوی HTML صورت می‌گیرد. ارزیابی این روش نشان می‌دهد، نرخ خطای منفی برابر با ۹٪ و نرخ خطای مثبت برابر با ۳۸٪ است.
۳		CID [۱۱]	بسیاری از روش‌های شناسایی و مکانیسم‌های جلوگیری، مبتنی بر شناسایی آدرس IP منبع مهاجم است. در طرف دیگر، مهاجم برای تغییر پی‌درپی آدرس IP منبع، از شبکه‌های متغیر پی‌درپی استفاده می‌کند. این روش باوجوداینکه هنوز به شکل قابل‌استفاده پیاده‌سازی نشده، نشان می‌دهد که می‌توان با تجزیه‌وتحلیل تمامی بخش‌های داده صفحه وب و همچنین استفاده ازاین‌روش در سامانه‌های تشخیص نفوذ[۹] و زیر نظر گرفتن تغییرات پی‌درپی آدرس IP، صفحات فیشینگ را شناسایی کرد.
۴		Phish Guard [۱۲]	این روش، در قالب یک افزونه مرورگر وب فایرفاکس[۱۰]، گام‌هایی ازجمله اصالت‌سنجی صفحه بازشده (با ارسال یک مجوز کاربری نادرست به صفحه موردنظر) را برای شناسایی صفحه فیشینگ، انجام می‌دهد.
۵		CANTINA [۱۳]	این روش شامل یک نوارابزار مرورگر وب است که برای تصمیم‌گیری در مورد فیشینگ بودن صفحه مشاهده شده است و محتوای آن را تجزیه‌وتحلیل می‌کند. این روش از الگوریتم TF-IDF[۱۱] برای بازیابی و بررسی اطلاعات و همچنین از موتورهای جستجو و برخی اکتشافات برای کاهش خطای مثبت، استفاده می‌کند. از جمله این اکتشافات، بررسی سن دامنه و تعداد نقطه‌ها در URL نام برد. ارزیابی این روش نشان می‌دهد، نرخ خطای منفی برابر با ۱۱٪ و نرخ خطای مثبت برابر با ۳٪ است.
۶		مولد لیست سیاه [۱۴]	این روش، مکانیسمی برای تولید لیست سیاه با استفاده از موتورهای جستجو مانند گوگل[۱۲] ارائه شده است. این روش پس از به دست آوردن نام شرکت در صفحه، آن را در گوگل جستجو می‌کند و اگر بین ده نتیجه اول جستجوی گوگل نباشد، آن را فیشینگ در نظر می‌گیرد. هدف کار این روش، شناسایی وب‌سایت‌های فیشینگ و سپس ثبت آن‌ها در پایگاه داده است. مجموعه داده ارزیابی شامل ۵۰۰ وب‌سایت قانونی تصادفی است که از نتایج جستجو کلیدواژه‌های مختلف تصادفی در گوگل تشکیل‌شده است و همچنین ۳۰ وب‌سایت فیشینگ که توسط PIRT[۱۳] طبقه‌بندی شده‌اند.

1- Parent
2- Childeren
3- Replace Top Level Domains
4- Directory
5- Equivalence
6- Query
7- Brand
8- Plug-in
9- IDS (Intrusion Detection System)
10- FireFox
11- Term Frequency/Inverse Document Frequency
12- Google
13- Phishing Incident Reporting and Termination

۷		شناسایی صفحه فیشینگ و کشف هدف [۱۵]	این روش، پس از بررسی موارد مختلف ازجمله، سن دامنه، وجود جعبه متن در صفحه، اطلاعات Whois و طول URL اقدام به شناسایی هدف صفحه با استفاده از Google API می‌کند.
۸		اکتشاف مبتنی بر URL [۱۶]	این روش مبتنی بر ویژگی‌های URL بوده و به‌طور خاص روی ظاهر آن‌ها تمرکز دارد. فازهای این روش شامل انتخاب ویژگی‌های URL، محاسبه شش مقادیر اکتشاف (از جمله رتبه صفحه، Alexa Rank و Alexa Reputation) و نهایتاً دسته‌بندی وب‌سایت می‌باشند. ارزیابی این روش با مجموعه داده‌ایی شامل ۱۱۶۰۰ سایت فیشینگ و همچنین ۵۰۰۰ سایت قانونی انجام‌گرفته است. نتیجه نشان می‌دهد که این روش ۹۷,۱۶٪ از وب‌سایت‌های فیشینگ را شناسایی کرده است.
۹		جلوگیری از حملات فیشینگ با استخراج رتبه صفحه، اعتبار و منبع کد [۱۷]	این روش برای شناسایی صفحات فیشینگ، رتبه صفحه و اعتبار آن را استخراج کرده و همچنین منبع کد صفحه موردنظر را بررسی کرده و با ویژگی‌های یک صفحه فیشینگ (از جمله طول URL، تعداد نقاط در URL) مقایسه می‌کند.
۱۰	شباهت ظاهری	شناسایی مبتنی بر شباهت ظاهری بدون داشتن اطلاعات سایت قربانی [۱۸]	هدف این روش، شناسایی وب‌سایت‌های فیشینگ با استفاده از شباهت ظاهری بدون داشتن لیست سفیدی از تصاویر وب‌سایت‌های قانونی است. البته پردازش تصویر در زمان مشاهده وب‌سایت در سیستم کاربر نهایی، می‌تواند تأخیر زمانی قابل‌توجهی را به دنبال داشته باشد. نرخ خطای منفی ۱۷/۴٪ و خطای مثبت ۸/۳٪ است. این روش دارای خطای منفی بالایی است.
۱۱		مبارزه با فیشینگ با ویژگی‌های کلیدی متمایز [۱۹]	این روش نیازمند آن است که مرورگر وب یک عکس فوری از سایت مشکوک بگیرد. سپس، عکس گرفته‌شده با لیست سفیدی از وب‌سایت‌های محافظت‌شده که مورد هدف فیشرها هستند (مانند PayPal، Amazon، eBay و وب‌سایت‌های بانک‌ها) مقایسه می‌شود. همچنین این روش از الگوریتم Harris-Laplace برای شناسایی تصویر استفاده می‌کند تا کار شناسایی را انجام دهد. نرخ خطای منفی و مثبت این روش کمتر از ۰,۱٪ است.
۱۲	یادگیری ماشین	ضد فیشینگ مبتنی بر متن و ظاهر: یک رویکرد بیزی [۲۰]	این روش، از فنی برای ترکیب کردن مدل‌های طبقه‌بندی متن و ظاهر با استفاده از یک مدل بیزی برای شناسایی صفحه‌ای که ازنظر متن و ظاهر شبیه به یک صفحه محافظت‌شده است، بهره می‌برد. در این روش، فرض بر این است که صفحات فیشینگ از اهدافشان تقلید می‌کنند.
۱۳		دسته‌بندی خودکار مقیاس بزرگ صفحات [۲۱]	این روش از API مربوط به‌مرور امن گوگل[1] برای انتشار نتایج به‌عنوان یک لیست سیاه استفاده می‌کند؛ بنابراین، ازآنجایی‌که در این روش از لیست سیاه استفاده می‌شود، این روش در مقابل حملات لحظه صفر ضعیف عمل می‌کند. خطای منفی این روش ۳۰-۱۶ درصد است و هیچ خطای مثبتی ندارد.
۱۴		نوارابزار ضد فیشینگ بیزی (B-APT) [۲۲]	این روش، یک نوارابزار، مخصوص مرورگر فایرفاکس بوده و از یک طبقه‌بندی بیزی برای تصمیم درباره فیشینگ بودن صفحه استفاده می‌کند. این تصمیم با تحلیل آماری کلمات کلیدی (یا نشانه‌ها ازجمله نام میزبانی و IP) در صفحه وب به دست می‌آید. این روش از لیست سفید برای کاهش نرخ خطای مثبت استفاده می‌کند. نرخ خطای منفی این روش برابر با صفر و خطای مثبت آن ۳٪ است.
۱۵		شناسایی خودکار هدف فیشینگ از صفحه فیشینگ [۲۳]	در این روش اشاره‌شده که شناسایی وب‌سایت‌های فیشینگ و همچنین اهداف آن‌ها با یافتن وب‌سایت‌های شبیه به صفحات مشکوک امکان‌پذیر است. اگر یک وب‌سایت مشکوک شبیه وب‌سایتی بانام دامنه متفاوت باشد، آنگاه وب‌سایت مشکوک به‌عنوان یک وب‌سایت فیشینگ در نظر گرفته می‌شود. برای مثال اگر وب‌سایت بسیار شبیه به Paypal باشد، مطمئناً یک وب‌سایت فیشینگ باهدف حمله به Paypal است. این روش مبتنی بر داده‌کاوی است و از فن دسته‌بندی استفاده می‌کند (ازجمله DBSCAN). مجموعه داده ارزیابی آن شامل ۸۷۴۵ وب‌سایت فیشینگ (استخراج‌شده از PhishTank) و همچنین ۱۰۰۰ وب‌سایت قانونی تصادفی از لینک‌های Yahoo! است. نتایج نشان می‌دهد، نرخ خطای منفی این روش برابر با ۸,۵۶٪ و نرخ خطای مثبت برابر با ۳,۴٪ است.

1- Google Safe Browsing

جدول (۲). فنون به‌کار رفته درروش پیشنهادی

طبقه	ردیف	عنوان فن	شرح مختصر فن درروش	ضریب اهمیت
اول	۱	جستجو در لیست سفید	درصورتی‌که آدرس وب‌سایت در این لیست باشد، به‌طورقطع این وب‌سایت یک وب‌سایت قانونی و معتبر است. در این لیست، URL وب‌سایت‌های معتبر ازجمله بانک‌های ایرانی ذخیره‌شده‌اند.	۱
	۲	جستجو در لیست سیاه	درصورتی‌که آدرس وب‌سایت در این لیست باشد، به‌طورقطع این وب‌سایت یک وب‌سایت فیشینگ است. این لیست از صفحات شناخته‌شده تاکنون ساخته‌شده است و به شکل مداوم به‌روز می‌شود.	۱
	۳	بررسی اطلاعات واردشده	پس از بررسی اطلاعات واردشده، درصورتی‌که شماره کارت‌بانکی وارد شود، جستجو در لیست سفید صورت می‌گیرد. درصورتی‌که در لیست سفید نباشد، از کاربر سؤال می‌شود و اگر پاسخ داد که در حال پرداخت اینترنتی است، فیشینگ محسوب می‌شود.	۱
دوم	۴	بررسی وضعیت میزبانی وب	با توجه به اینکه در اکثر موارد، وب‌سایت‌های فیشینگ از میزبان‌های وب رایگان استفاده می‌کنند، درصورتی‌که وب‌سایت از یکی از آن‌ها استفاده می‌کند، احتمال فیشینگ بودن آن وجود دارد. جهت کسب اطلاعات در مورد میزبانی وب یک صفحه، وضعیت میزبانی وب بررسی می‌گردد.	۰/۲
	۵	بررسی رتبه صفحه	با توجه به اینکه اغلب وب‌سایت‌های فیشینگ بدون رتبه صفحه[1] هستند، این مورد با استفاده Google API بررسی می‌شود و در صورت نداشتن رتبه، احتمال فیشینگ بودن صفحه وجود دارد.	۰/۲
	۶	بررسی ویژگی‌های صفحه	ویژگی‌های صفحه فیشینگ مانند طول URL و استفاده از اعتبارنامه SSL موردبررسی قرار می‌گیرند.	۰/۶
سوم	۷	جمع‌بندی نهایی	پس از بررسی نتایج فنون هرکدام از طبقه‌های اول و دوم، نتیجه نهایی نمایش داده می‌شود.	-

جدول (۳). اطلاعات مربوط به برخی از صفحات بانکداری اینترنتی بانک‌های ایران

ردیف	نام بانک	آدرس صفحات		
		بانکداری اینترنتی	سرویس پرداخت	سایر صفحات خاص
۱	بانک مرکزی	-	-	http://www.cbi.ir http://www.shaparak.ir
۲	بانک ملت	https://ebanking.bankmellat.ir/ebanking	https://bpm.shaparak.ir	http://www.bankmellat.ir
۳	بانک ملی	https://saba.bankmelli-iran.com	https://sadad.shaparak.ir	http://bmi.ir https://echargecard.bmi.ir
۴	بانک رفاه کارگران	https://www.rb24.ir	-	http://www.refah-bank.ir https://card.rb24.ir
۵	بانک مسکن	https://ib.bank-maskan.ir	https://epayment.bank-maskan.ir	http://bank-maskan.ir
۶	بانک سپه	https://www.ebanksepah.ir	https://www.ebanksepah.ir	http://www.banksepah.ir
۷	بانک تجارت	https://online.tejaratbank.net https://onlinetb.tejaratbank.net	https://pg.tejaratbank.net	http://www.tejaratbank.ir

1 PageRank

اطلاعات مربوط به آدرس صفحات کمک می‌کند تا یک صفحه قانونی، به اشتباه، به‌عنوان یک صفحه فیشینگ شناخته نشود و همچنین نرخ خطای مثبت را به حداقل رساند.

با توجه به اهمیت بسیار زیاد موضوع، تأییدیه اطلاعات فوق با مراجعه حضوری یا تماس تلفنی با مدیریت‌های شعب بانک‌ها و یا با کسب اطلاعات از طریق وب‌سایت‌های رسمی معرفی‌شده در بانک مرکزی ایران، صورت پذیرفت.

پس از جستجو در لیست سفید، درصورتی‌که URL مورد نظر در لیست نباشد اما حاوی URL قانونی باشد، به‌عنوان فیشینگ شناخته می‌شود. برای مثال:

http://bankmellat.ir.phish.com

در اینجا URL حاوی URL قانونی bankmellat.ir است اما دامنه اصلی آن برابر است با pish.com.

۳-۱-۲- جستجو در لیست سیاه

درصورتی‌که آدرس وب‌سایت یا دامنه اصلی آن (با www و بدون www) و همچنین آدرس‌های IP آن در این لیست باشد، به‌طورقطع این وب‌سایت یک وب‌سایت فیشینگ است. این لیست از صفحات شناخته‌شده تاکنون ساخته شده است و به شکل مداوم به‌روز می‌شود.

در صورت شناسایی یک URL به‌عنوان فیشینگ، URL به همراه دامنه اصلی آن (با www و بدون www) و آدرس‌های IP آن به‌طور جداگانه در این لیست قرار داده می‌شود.

۳-۱-۳- بررسی اطلاعات واردشده

با توجه به اینکه در اکثر موارد، فیشرها به دنبال اطلاعاتی مانند اطلاعات حساب‌های بانکی هستند، بررسی اینکه کاربر چه نوع اطلاعاتی را در یک صفحه وارد می‌کند بسیار اهمیت دارد.

در صفحات بانکداری اینترنتی، کاربر ملزم به ورود اطلاعاتی ازجمله نام کاربری و رمز عبور است که در واقع فرمتی برای آن‌ها تعیین نمی‌شود و تشخیص نوع آن‌ها بسیار دشوار است اما در صفحات پرداخت اینترنتی مانند شکل (۱)، اطلاعاتی مانند (شماره کارت بانکی، رمز عبور، CVV2، تاریخ انقضای کارت) وارد می‌شود که از این میان، شماره کارت بانکی و تاریخ انقضای کارت دارای فرمت مشخصی است.

در شکل (۱) صفحه سرویس پرداخت اینترنتی مربوط به بانک ملت ارائه شده است.

شماره کارت: یک عدد ۱۶ رقمی که شامل چهار قسمت است که هر قسمت چهار رقم را در خود جای می‌دهد (در برخی بانک‌ها ۱۹ رقمی و شامل پنج قسمت است). ۶ رقم ابتدایی شماره کارت بانکی در زمان نگارش این پایان‌نامه، مطابق جدول (۴) است.

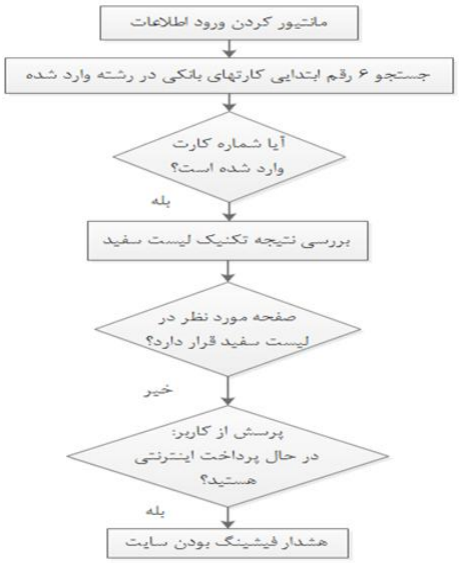
شکل (۱). صفحه سرویس پرداخت اینترنتی بانک ملت

جدول (۴). اطلاعات مربوط به شش رقم ابتدایی شماره کارت‌های بانکی برخی از بانک‌ها

ردیف	نام بانک	رقم ابتدایی شماره کارت
۱	بانک مرکزی	۹۳۶۴۵۰
۲	بانک ملت	۶۱۰۴۳۳
۳	بانک ملی	۶۰۳۷۹۹
۴	بانک رفاه کارگران	۵۸۹۴۶۳
۵	بانک مسکن	۶۲۸۰۲۳
۶	بانک سپه	۵۸۹۲۱۰
۷	بانک کشاورزی	۶۰۳۷۷۰
۸	بانک تجارت	۶۲۷۳۵۳
۹	بانک صادرات	۶۰۳۷۶۹

کاربر با هدف استفاده از سرویس پرداخت اینترنتی، شماره کارت بانکی خود را وارد می‌کند. در این صورت، اگر آدرس صفحه در لیست سفید نباشد، به‌احتمال‌زیاد این صفحه، یک صفحه فیشینگ است. جهت اطمینان از این موضوع، از کاربر سؤال می‌شود که آیا در حال پرداخت اینترنتی است؟ و در صورت جواب مثبت، مطمئناً این صفحه، یک صفحه جعلی پرداخت اینترنتی است.

در شکل (۲)، روند انجام کار در این فن نمایش داده شده است:



شکل (۲). روند انجام کار در فن بررسی اطلاعات واردشده

۳-۱-۴-بررسی وضعیت میزبانی وب

اطلاعات جمع‌آوری‌شده در خصوص استفاده فیشرها از میزبان‌های وب و سرویس‌های ثبت دامنه رایگان از گزارش‌های APWG در سال‌های مختلف، نشان می‌دهد که اکثر فیشرها از این نوع سرویس‌ها استفاده می‌کنند. جدول (۵)، برخی از این میزبان‌های وب و سرویس‌های ثبت دامنه رایگان را نشان می‌دهد.

جدول (۵). برخی از میزبان‌های وب و سرویس‌های ثبت دامنه رایگان

ردیف	نشانه	نام میزبان
۱	hol.es p.ht	Hostinger
۲	net.tf	UNONIC.COM
۳	usa.cc	freeavailabledomains.com
۴	cixx6.com	cixx6.com
۵	wink.ws	wink.ws
۶	fav.cc	fav.cc
۷	ias3.com	ias3.com
۸	co.vu	codotvu
۹	oicp.net	Oray

برای اطلاع از این موضوع، که صفحه موردنظر از کدام میزبان وب یا سرویس ثبت دامنه استفاده می‌کند، وضعیت میزبانی وب بررسی می‌گردد. درصورتی‌که صفحه موردنظر، یکی از این سرویس‌ها را استفاده می‌کند، احتمالاً یک وب‌سایت فیشینگ است و در غیر این صورت، خروجی این فن برابر با صفر است.

۳-۱-۵- بررسی رتبه صفحه[1]

با توجه به این‌که وب‌سایت‌های قانونی دارای رتبه صفحه بالایی در موتورهای جستجو هستند و در مقابل، وب‌سایت‌های فیشینگ، بدون رتبه صفحه هستند، این مورد می‌تواند در شناسایی یک وب‌سایت فیشینگ و همچنین برای کاهش نرخ خطای مثبت برای وب‌سایت‌های قانونی کمک کند.

با استفاده از API شرکت گوگل امکان به‌دست آوردن رتبه صفحه فراهم شده است.

۳-۱-۶- بررسی ویژگی‌های صفحه

بر اساس ویژگی‌های وب‌سایت‌های فیشینگ ارائه‌شده در [۸]، موارد زیر بررسی می‌شوند.

۱- استفاده از آدرس IP به‌عنوان URL
۲- استفاده از علامت @ در URL
۳- بررسی طول URL
۴- تعداد نقاط در URL
۵- استفاده از اعتبارنامه SSL
۶- استفاده از برگه HTML

برای جمع‌بندی نهایی نتایج هرکدام از فنون، گام‌های زیر انجام می‌شود:

۱. بررسی نتیجه فن لیست سفید و در صورت وجود آدرس صفحه در این لیست، مقدار قانونی نمایش داده می‌شود و سایر فنون بررسی نمی‌شود.

۲. بررسی نتیجه فن لیست سیاه و در صورت وجود آدرس صفحه در این لیست، مقدار فیشینگ نمایش داده می‌شود و سایر فنون بررسی نمی‌شود.

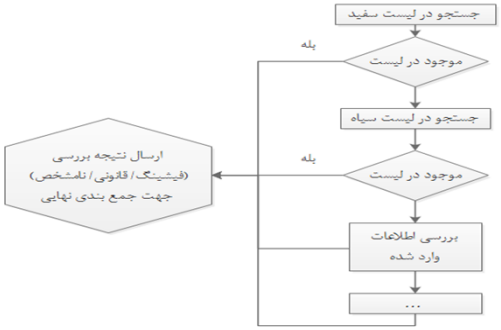
۳. بررسی نتیجه فن بررسی اطلاعات واردشده و در صورت پاسخ مثبت (فیشینگ) این فن، فیشینگ بودن صفحه به خروجی ارسال می‌شود و صفحه موردنظر در لیست سیاه قرار داده می‌شود.

۴. بررسی نتایج فنون وضعیت میزبانی وب، رتبه صفحه و ویژگی‌های صفحه انجام‌شده و سپس هرکدام در ضریب اهمیتشان ضرب می‌شوند و مجموع آن‌ها، خروجی طبقه دوم روش است. درصورتی‌که احتمال فیشینگ بودن صفحه موردنظر بیش از ۵۰ درصد باشد، پس از سؤال از کاربر، آدرس صفحه موردنظر در لیست سیاه قرار داده می‌شود. فرمول محاسبه طبقه دوم به‌صورت زیر است:

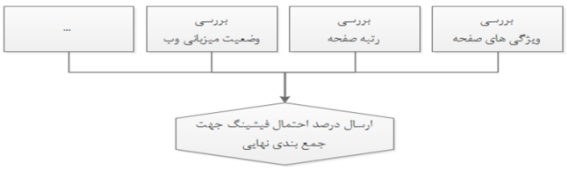
$$Percent = \sum Result_i * Factor_i$$

1 -PageRank

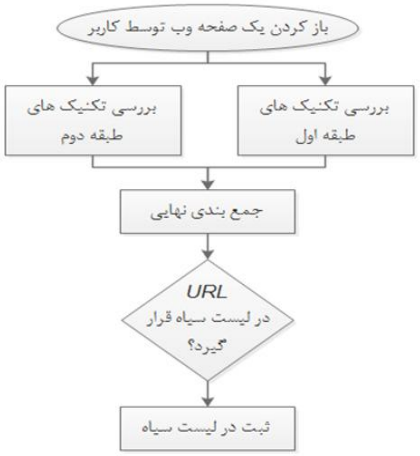
در شکل (۳)، روند انجام کار در طبقه اول، در شکل (۴) روند انجام کار در طبقه دوم و همچنین در شکل (۵) روند کلی انجام کار درروش پیشنهادی ارائه شده است:



شکل (۳). روند انجام کار در طبقه اول



شکل (۴). روند انجام کار در طبقه دوم



شکل (۵). روند کلی انجام کار در روش پیشنهادی

۳-۲- پیاده‌سازی روش پیشنهادی

روش پیشنهادی در قالب مرورگر وب با استفاده از ماکروسافت ویژوال استودیو[1] و زبان #C پیاده‌سازی شده است. داده‌ها در پایگاه داده سرویس‌گر SQL ماکروسافت[2] ذخیره می‌شوند.

تمام پیاده‌سازی‌ها روی کامپیوتر شخصی با قدرت پردازش ۲ گیگاهرتز با حافظه رم ۲ گیگابایت انجام شده است.

این مرورگر وب مطابق شکل (۶)، شامل یک جعبه متن[3] جهت ورود آدرس صفحه، یک شی مرور وب[4] و یک دکمه[5] که با کلیک کردن روی آن، صفحه وب آدرس واردشده در شی مرور وب نمایش داده می‌شود. همچنین این مرورگر شامل یک نوار وضعیت که حاوی منویی با گزینه‌های لیست سفید، لیست سیاه و لیست میزبان‌های رایگان است.

شکل (۶). شمای کلی از مرورگر وب پیاده‌سازی شده

۱. روال‌های مهم پیاده‌سازی شده، شامل پیاده‌سازی فن لیست سفید، پیاده‌سازی فن لیست سیاه، پیاده‌سازی فن بررسی اطلاعات واردشده، پیاده‌سازی فن بررسی وضعیت میزبانی وب، پیاده‌سازی فن بررسی رتبه صفحه و پیاده‌سازی فن بررسی ویژگی‌های صفحه است.

۴- معیارهای ارزیابی روش‌های شناسایی حمله فیشینگ

بر اساس ارزیابی‌های ارائه‌شده در پژوهش‌های مختلف در حوزه شناسایی حمله فیشینگ مانند [۱]، معیارهای زیر برای ارزیابی روش‌های شناسایی حمله فیشینگ ارائه و پیشنهاد می‌گردد. این معیارها در سه دسته زیر قرار می‌گیرند:

۴-۱-۱- دسته اول: ویژگی‌ها

این دسته شامل ویژگی‌های روش‌های شناسایی حمله فیشینگ است که قابلیت‌های آن‌ها را نشان می‌دهد.

۱. **شناسایی لحظه صفر**[6]: روش موردنظر توانایی شناسایی حمله فیشینگ در لحظه صفر را دارد یا خیر. منظور شناسایی آنی وب‌سایت‌ها یا ایمیل‌های فیشینگی است که همان لحظه، توسط فیشر، در دسترس قرار داده شده یا ارسال شده است.

1 -Microsoft Visual Studio
2 -Microsoft SQL Server
3 -Textbox
4 -WebBrowser
5 -Button
6 -Zero-Day

۲. **شناسایی فارمینگ**[1]: روش موردنظر، قابلیت شناسایی حمله فارمینگ به‌عنوان یکی از روش‌های فیشینگ را دارد یا خیر.

۳. **شناسایی هدف:** این معیار نشان می‌دهد که روش شناسایی به دنبال شناسایی هدف حمله فیشینگ هست یا خیر. منظور از هدف، می‌تواند همان وب‌سایت قانونی باشد که فیشر آن را جعل کرده است و یا این‌که وب‌سایت فیشینگ در پی به‌دست آوردن چه نوع اطلاعاتی (مانند اطلاعات حساب بانکی و یا نام کاربری و رمز عبور یک ایمیل) است.

۴. **شناسایی اشیاء جاسازی‌شده:** جعل برخی از وب‌سایت‌های قانونی توسط فیشرها با استفاده از اشیاء جاسازی‌شده مانند تصویر و اسکریپت انجام می‌شود. این معیار نشان می‌دهد که آیا روش موردنظر قابلیت شناسایی وب‌سایت‌های فیشینگی را دارد که از تصویر و اسکریپت و سایر اشیاء جاسازی سازی شده ساخته‌شده است یا خیر.

۵. **متمرکز بر سرویس‌دهنده:** روش شناسایی، در سرویس‌دهنده متمرکز است یا خیر؟

۶. **متمرکز بر ایستگاه کاری:** روش شناسایی، در ایستگاه کاری متمرکز است یا خیر؟

۷. **تبدیل به ابزار قابل‌استفاده:** روش شناسایی به ابزاری قابل‌استفاده توسط کاربر یا یک سرویس‌دهنده تبدیل شده است یا خیر؟

۸. **نیازمند ارتباطات شبکه:** برخی از روش‌ها برای شناسایی نیازمند استفاده از ارتباطات شبکه هستند. ارسال/دریافت اطلاعات به/از سرویس‌دهنده یا وب‌سایت مشکوک ازجمله این موارد است.

۹. **نیازمند ارتباط با کاربر:** روش موردنظر برای شناسایی حمله فیشینگ وابسته به تعامل با کاربر است یا خیر. البته بدیهی است نمایش نتیجه یک روش و یا اعلام هشدار فیشینگ بودن یک وب‌سایت به کاربر به‌عنوان وابستگی مطرح نمی‌شود.

۱۰. **وابستگی زبانی:** روش موردنظر محدودیت زبان دارد یا خیر. به‌عبارت‌دیگر آیا روش موردنظر وابسته به زبان خاصی (مانند انگلیسی یا فارسی) هست یا خیر؟

۴-۱-۲- دسته دوم: نتایج آزمون

برای آزمایش یک روش شناسایی حمله فیشینگ، می‌توان از مجموعه داده[2] شامل نمونه‌های قانونی و فیشینگ استفاده کرد که درنهایت چهار دسته احتمالی وجود خواهد داشت که مطابق جدول (۶)، این پراکندگی نمایش داده شده است:

جدول (۶). پراکندگی نتایج دسته‌بندی روش‌های شناسایی [۱]

	دسته‌بندی شده به‌عنوان فیشینگ	دسته‌بندی شده به‌عنوان قانونی
فیشینگ است	NP→P	NP→L
قانونی است	NL→P	NL→L

NP→P تعداد نمونه‌های **فیشینگ** که به‌درستی به عنوان **فیشینگ** دسته‌بندی شده‌اند.

NL→P تعداد نمونه‌های **قانونی** که به‌اشتباه به عنوان **فیشینگ** دسته‌بندی شده‌اند.

NP→L تعداد نمونه‌های **فیشینگ** که به اشتباه به عنوان **قانونی** دسته‌بندی شده‌اند.

۱۱. NL→L تعداد نمونه‌های **قانونی** که به درستی به عنوان **قانونی** دسته‌بندی شده‌اند.

بر این اساس، می‌توان معیارهای زیر را ارائه کرد:

۱. **تأخیر**[3]: روش شناسایی چقدر زمان برای انجام کار خود نیاز دارد.

۲. **خطای مثبت**[4] **یا FP**: با اجرای روش شناسایی فیشینگ، روی فهرستی از وب‌سایت‌های قانونی، درصد خطاهایی که در شناسایی این‌گونه وب‌سایت‌ها به‌عنوان وب‌سایت فیشینگ شناخته‌شده، خطای مثبت نامیده می‌شود (به‌عبارت دیگر سایت قانونی را به‌عنوان سایت فیشینگ تشخیص می‌دهد). فرمول زیر، نحوه محاسبه خطای مثبت را نشان می‌دهد.

$$\mathrm{FP} = \frac{NL \rightarrow P}{NL \rightarrow L + NL \rightarrow P} \qquad (۱)$$

۳. **خطای منفی**[5] **یا FN**: با اجرای روش شناسایی فیشینگ، روی فهرستی از وب‌سایت‌های فیشینگ، درصد خطاهایی که در شناسایی این‌گونه وب‌سایت‌ها به‌عنوان وب‌سایت قانونی شناخته‌شده، خطای منفی نامیده می‌شود (به‌عبارت‌دیگر سایت فیشینگ را به عنوان سایت قانونی تشخیص می‌دهد). فرمول زیر، نحوه محاسبه خطای منفی را نشان می‌دهد.

$$\mathrm{FN} = \frac{NP \rightarrow L}{NP \rightarrow P + NP \rightarrow L} \qquad (۲)$$

1- Pharming
2 -Dataset

3 -Delay
4 -False Positives
5 -False Negatives

۴-۱-۳- دسته سوم: فنون مورداستفاده

این دسته شامل فنونی است که روش‌های شناسایی حمله فیشینگ استفاده می‌کنند که شامل موارد زیر می‌شوند:

۱. **لیست سیاه**[1]: برخی از روش‌های شناسایی مبتنی بر لیست سیاه هستند که شامل موارد فیشینگ و غیرقانونی هستند. برخی نیز برای کاهش نرخ خطای منفی از آن استفاده می‌کنند.
۲. **لیست سفید**[2]: برخی از روش‌های شناسایی مبتنی بر لیست سفید هستند که شامل موارد قانونی و صحیح هستند. برخی نیز برای کاهش نرخ خطای مثبت از آن استفاده می‌کنند.
۳. **اکتشافی**[3]: این معیار نشان می‌دهد که آیا روش شناسایی در پی کشف مشخصاتی است که در حملات فیشینگ واقعی یافت شده است یا خیر. در این‌گونه موارد، روش‌ها در پی رسیدن به نزدیک‌ترین جواب صحیح هستند.
۴. **شباهت ظاهری**[4]: روش شناسایی شباهت ظاهری (معمولاً با گرفتن عکس از صفحه وب‌سایت) برای کار خود بهره می‌برد یا خیر؟
۵. **یادگیری ماشین**: روش شناسایی از یادگیری ماشین و روش‌های هوش مصنوعی مانند فنون داده کاوی برای کار خود بهره می‌برد یا خیر؟

۵- اعتبارسنجی روش پیشنهادی

در این بخش به اعتبارسنجی روش پیشنهادی و همچنین مقایسه با روش‌های شناسایی وب‌سایت فیشینگ ارائه‌شده در بخش سوم، پرداخته می‌شود. با توجه به ارزیابی صورت گرفته توسط [۱۴]، روش پیشنهادی با دو مجموعه آزموده شده است تا میزان خطای مثبت و منفی آن محاسبه گردد. یک مجموعه شامل تعدادی سایت درست و قانونی است و مجموعه دیگر شامل مجموعه‌ای از صفحات فیشینگ است. از نمونه‌های سایت درست برای محاسبه میزان خطای مثبت روش و از نمونه‌های فیشینگ برای محاسبه میزان خطای منفی این روش استفاده می‌شود.

مجموعه وب‌سایت‌های قانونی شامل ۵۰۰ صفحه که در سایت پیوندها به آدرس www.peyvandha.ir و www.iran.ir نمایه شده‌اند، است. همچنین مجموعه وب‌سایت‌های فیشینگ شامل ۱۰۰ سایت فیشینگ شناسایی‌شده در Phish Tank و ۲۰ سایت فیشینگ طراحی‌شده و ۱۱ سایت فیشینگ شناسایی‌شده در طول زمان انجام تحقیق بوده است. روش انجام آزمون به این صورت بوده است که، هر یک از آدرس های اینترنتی از هر مجموعه را در نوار آدرس مرورگر وارد کرده تا صفحه مربوط به آن در مرورگر نشان داده شود. مرورگر بر اساس الگوریتم پیاده‌سازی شده نتیجه را بازمی‌گرداند. حاصل انجام آزمون روی نمونه‌های مذکور نشان می‌دهد که میزان خطای مثبت روش پیشنهادی صفر درصد است و این بدان معنی است که این روش، هیچ سایت قانونی را به‌عنوان فیشینگ تشخیص نمی‌دهد. دلیل این امر این است که، لیست سفید این روش شامل آدرس‌های صفحات وب بانکداری اینترنتی بانک‌های ایران است و همچنین در اکثر موارد، ویژگی‌های وب‌سایت‌های فیشینگ مورد بررسی در این روش، در وب‌سایت‌های قانونی یافت نمی‌شود و یا احتمال وجود آن‌ها زیر ۵۰٪ درصد است که این روش، آن‌ها را به‌عنوان فیشینگ در نظر نمی‌گیرد.

میزان خطای منفی این روش ۳٪ است و این بدان معنا است که در ۳٪ از موارد، وب‌سایت‌های فیشینگ به‌عنوان غیر فیشینگ یا به‌عبارت‌دیگر قانونی (در برخی موارد مشکوک) در نظر گرفته شده‌اند.

۵-۱- مقایسه روش پیشنهادی با سایر روش‌ها

با بهره‌گیری از مقایسه [۱] و بررسی هرکدام از روش‌های شناسایی، به‌صورت جداگانه، ارزیابی انجام‌شده بر اساس معیارهای معرفی‌شده در بخش چهارم، در جدول (۷) نمایش داده شده است.

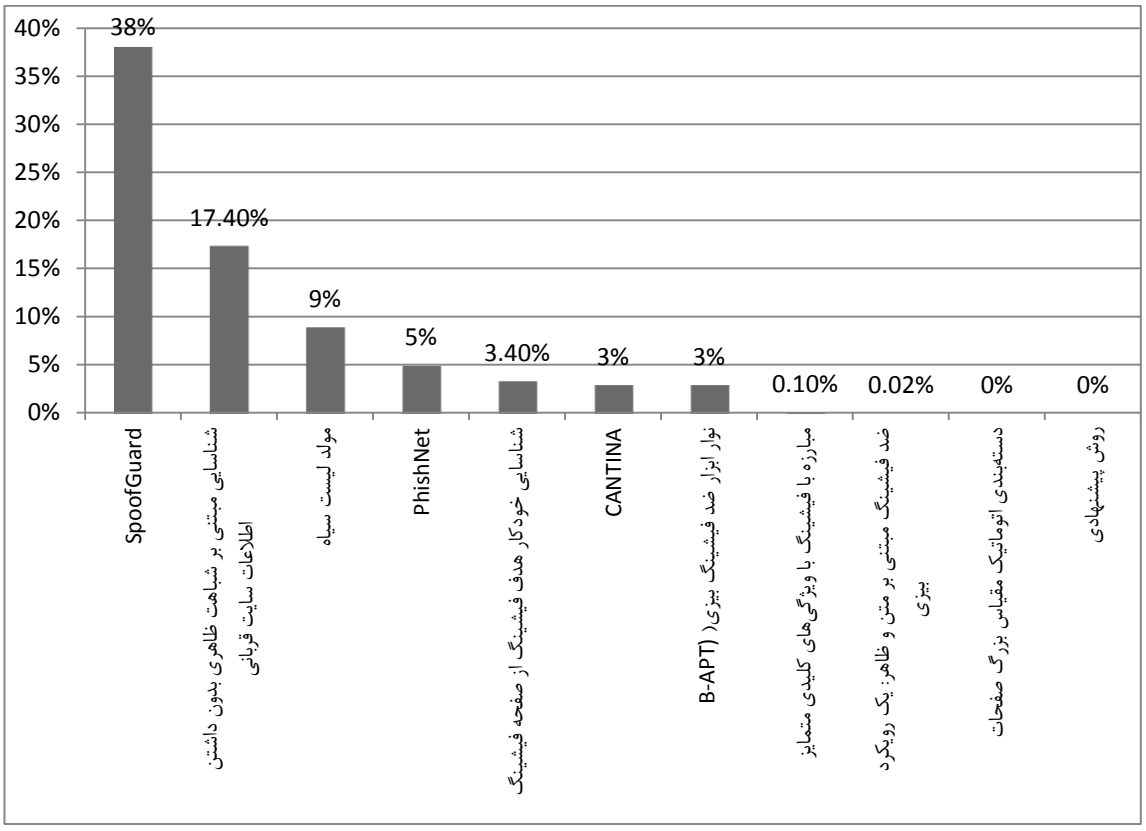
در این جدول، در مقابل نام روش، معیارهای ارزیابی جهت مقایسه ارائه شده است. درصورتی‌که روش مورد نظر از فنون لیست سیاه، لیست سفید، اکتشافی، شباهت ظاهری و یادگیری ماشین استفاده نماید و همچنین ویژگی‌های مختلف از جمله شناسایی لحظه صفر، شناسایی فارمینگ، شناسایی هدف، شناسایی اشیاء جاسازی‌شده، متمرکز بر سرویس‌دهنده، متمرکز بر ایستگاه کاری، تبدیل به ابزار جهت استفاده، عدم نیاز به ارتباطات شبکه، عدم نیاز به ارتباط با کاربر و عدم وابستگی زبانی را داشته باشد، با علامت ✓ مشخص‌شده است. در ستون‌های خطای مثبت و خطای منفی نیز نتیجه آزمون هرکدام ارائه‌شده است.

1 -Blacklists
2 -Whitelists
3 -Heuristics
4 -Visual Similarity

جدول(۷). خلاصه مقایسه روش پیشنهادی با سایر روش‌ها

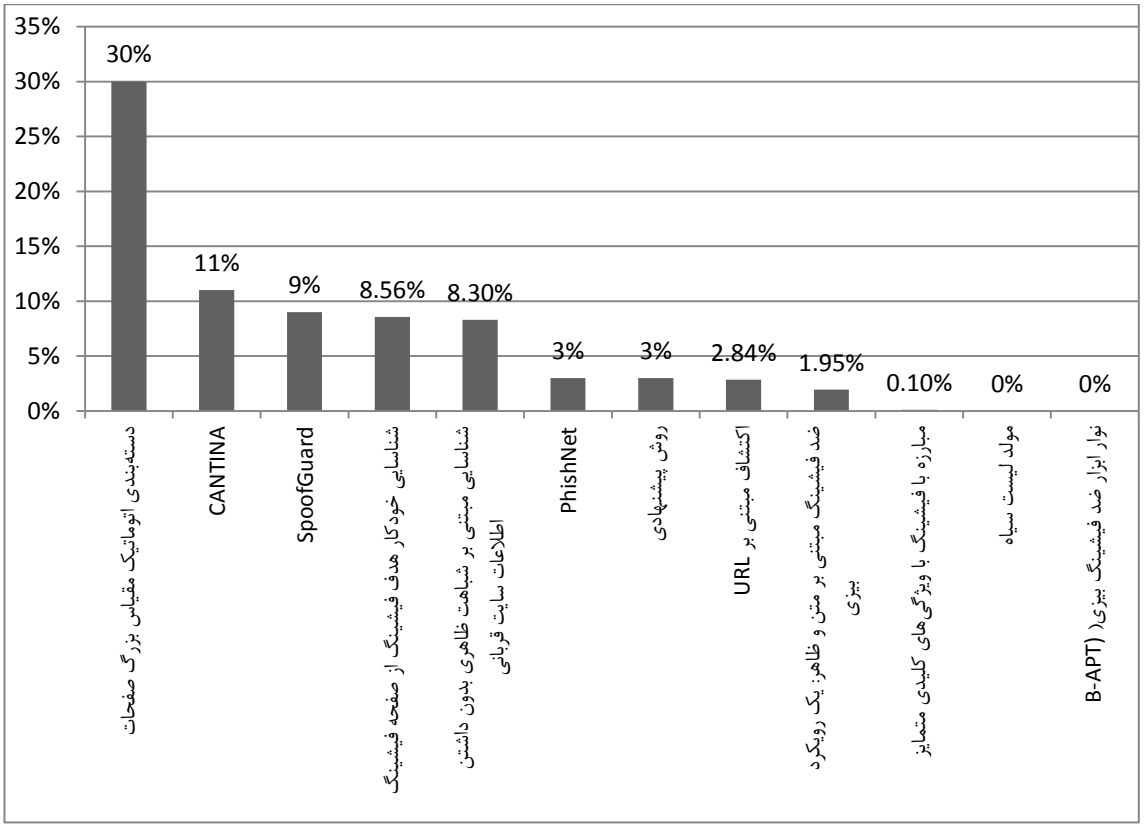
ردیف	روش شناسایی	فن مورداستفاده					ویژگی‌ها										نتیجه آزمون	
		لیست سیاه	لیست سفید	اکتشافی	شباهت ظاهری	یادگیری ماشین	شناسایی لحظه صفر	شناسایی فارمینگ	شناسایی هدف	شناسایی اشیاء جاسازی‌شده	متمرکز بر سرویس‌دهنده	متمرکز بر ایستگاه کاری	تبدیل به ابزار جهت استفاده	عدم نیاز به ارتباطات شبکه	عدم نیاز به ارتباط با کاربر	عدم وابستگی زبانی	خطای مثبت	خطای منفی
۱	PhishNet	✓		✓								✓	✓	✓	✓	✓	٪۵	٪۳
۲	SpoofGuard			✓			✓					✓	✓	✓	✓		٪۳۸	٪۹
۳	CID			✓				✓			✓				✓	✓	–	–
۴	PhishGuard			✓			✓					✓	✓	✓	✓		–	–
۵	CANTINA			✓			✓					✓	✓		✓		٪۳	٪۱۱
۶	مولد لیست سیاه			✓			✓					✓	✓		✓	✓	٪۹	٪۰
۷	شناسایی صفحه فیشینگ و کشف هدف			✓			✓		✓			✓			✓		–	–
۸	اکتشاف مبتنی بر URL			✓			✓				✓	✓			✓	✓	–	٪۲,۸۴
۹	جلوگیری از حملات فیشینگ با استخراج رتبه صفحه، اعتبار و منبع کد			✓			✓					✓			✓		–	–
۱۰	شناسایی مبتنی بر شباهت ظاهری بدون داشتن اطلاعات سایت قربانی	✓	✓		✓		✓					✓			✓	✓	٪۱۷,۴	٪۸,۳
۱۱	مبارزه با فیشینگ با ویژگی‌های کلیدی متمایز	✓			✓		✓					✓			✓	✓	<٪۰,۱	<٪۰,۱
۱۲	ضد فیشینگ مبتنی بر متن و ظاهر: یک رویکرد بیزی	✓	✓	✓		✓	✓					✓			✓	✓	٪۰-٪۰,۰۲	۰- ٪۱,۹۵
۱۳	دسته‌بندی خودکار مقیاس بزرگ صفحات	✓	✓	✓		✓					✓	✓	✓		✓		٪۰	٪۳۰-۱۶
۱۴	نوارابزار ضد فیشینگ بیزی (B-APT)		✓			✓	✓						✓		✓		٪۳	٪۰
۱۵	شناسایی خودکار هدف فیشینگ از صفحه فیشینگ				✓	✓	✓		✓			✓			✓		٪۳,۴	٪۸,۵۶
۱۶	روش پیشنهادی	✓	✓	✓			✓					✓	✓	✓	✓	✓	٪۰	٪۳

بر اساس مقایسه انجام‌شده در جدول (۷)، مقایسه روش‌ها بر اساس نرخ خطای مثبت آن‌ها در نمودار شکل (۷) ارائه شده است:



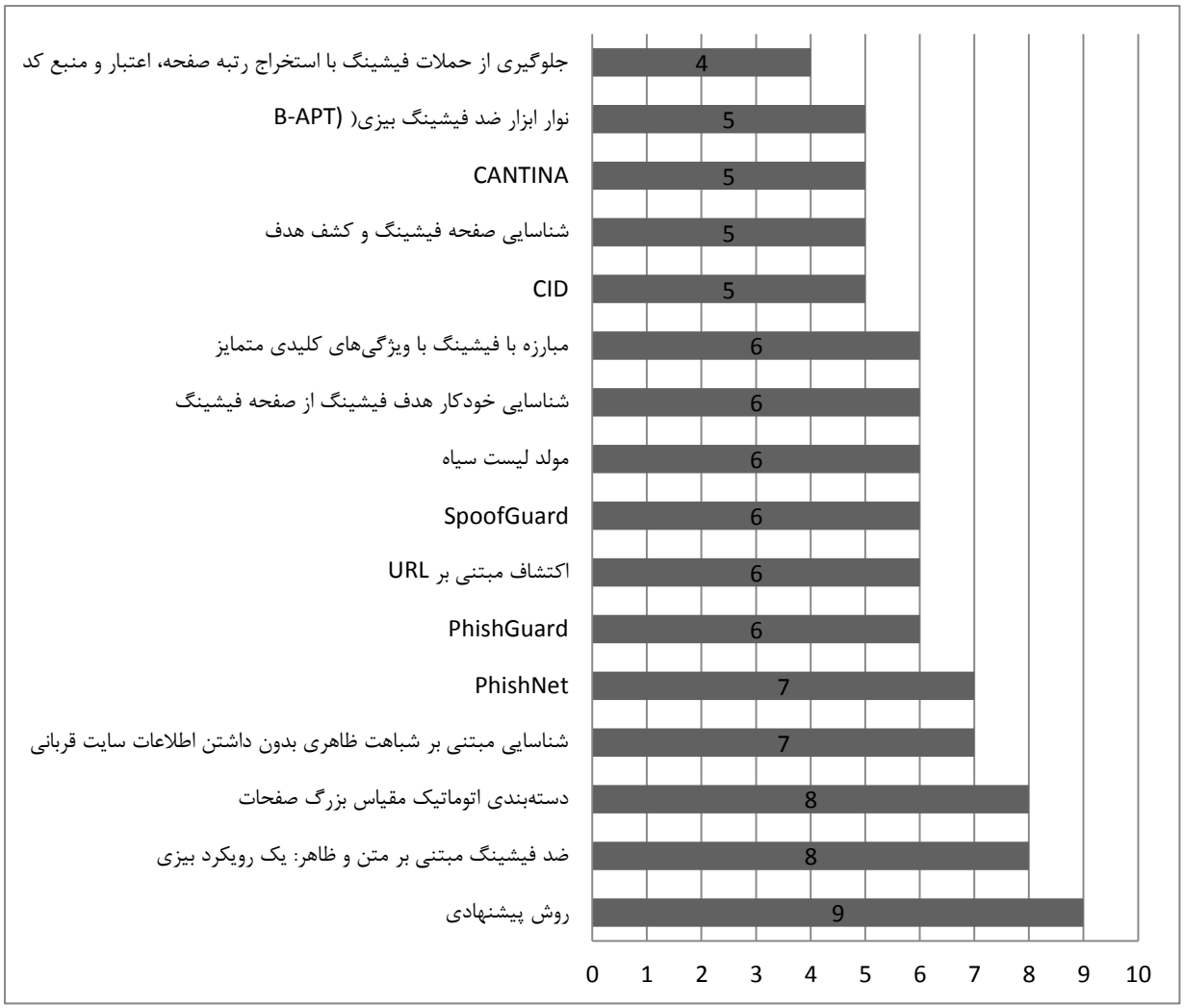
شکل (۷). مقایسه روش‌ها بر اساس نرخ خطای مثبت

بر اساس مقایسه انجام‌شده در جدول (۷)، مقایسه روش‌ها بر اساس نرخ خطای منفی آن‌ها در نمودار شکل (۸) ارائه شده است:



شکل (۸). مقایسه روش‌ها بر اساس نرخ خطای منفی

بر اساس مقایسه انجام‌شده در جدول (۷)، مقایسه روش‌ها بر اساس تعداد ویژگی‌ها و فنون مورداستفاده آن‌ها در نمودار شکل (۹) ارائه شده است:



شکل (۹). مقایسه روش‌ها بر اساس تعداد ویژگی‌ها و فنون مورد استفاده آن‌ها

۶- نتیجه‌گیری

در این مقاله کلاهبرداری فیشینگ و شیوه‌های انجام آن به طور اجمالی بررسی شد. بررسی نشان داد از مهم‌ترین چالش های موجود در بانکداری اینترنتی، خطر حملات فیشینگ و کلاهبرداری‌های اینترنتی محسوب می‌شود و خسارت‌های فراوانی به مشتریان و سازمان‌ها وارد می‌کند. ازاین‌رو، تلاش‌های زیادی برای شناسایی و مقابله با این‌گونه حملات شده است؛ اما ازآنجایی‌که فیشرها، همواره روش خود را با هزینه اندک تغییر می‌دهند، روش‌های تشخیص فیشینگ باید بهبود یابد تا قادر به شناسایی و مقابله با روش‌های جدید فیشینگ باشد. در این مقاله، پس از بررسی پژوهش‌های موجود در حوزه شناسایی و مقابله با فیشینگ و ارزیابی تحقیقات مرتبط بر اساس معیارهای پیشنهادی، روش جدیدی برای تشخیص وب‌سایت فیشینگ ارائه گردید.

این روش شامل دوطبقه (طبقه اول شامل فنون که جواب قطعی را ارائه می‌دهند و طبقه دوم شامل فنونی که جواب احتمالی بین ۰ تا ۱ را به خروجی ارسال می‌کنند) و هفت فن شامل جستجو در لیست سفید، جستجو در لیست سیاه، بررسی اطلاعات واردشده، بررسی رتبه صفحه، بررسی اطلاعات Whois، بررسی ویژگی‌های صفحه، به همراه جمع‌بندی نهایی است. این روش قابلیت شناسایی لحظه صفر وب سایت های فیشینگ را دارد و بر ایستگاه کاری تمرکز دارد. همچنین این روش به ابزاری جهت استفاده (مرورگر وب) تبدیل‌شده است. این روش به ارتباطات شبکه و ارتباط با کاربر نیاز ندارد و همچنین وابستگی زبانی ندارد.

روش پیشنهادی با دو مجموعه آزموده شده است تا میزان

خطای مثبت و منفی آن محاسبه گردد. یک مجموعه شامل تعدادی سایت قانونی است و مجموعه دیگر شامل مجموعه‌ای از صفحات فیشینگ است. از نمونه‌های سایت‌های قانونی برای محاسبه میزان خطای مثبت روش و از نمونه‌های سایت‌های فیشینگ برای محاسبه میزان خطای منفی روش پیشنهادی استفاده شد. مجموعه وب‌سایت‌های قانونی شامل ۵۰۰ صفحه نمایه شده در سایت پیوندها به آدرس www.peyvandha.ir و www.iran.ir بود. همچنین مجموعه وب‌سایت‌های فیشینگ شامل ۱۰۰ سایت فیشینگ شناسایی‌شده در PhishTank و ۲۰ سایت فیشینگ طراحی‌شده و ۱۱ سایت فیشینگ شناسایی‌شده در طول زمان نگارش مقاله بوده است. نتیجه آزمون روش پیشنهادی حاکی از آن است که نرخ خطای مثبت برابر با صفر و خطای منفی آن ۳٪ است. بر اساس ارزیابی انجام‌شده روش پیشنهادی موارد زیر را پوشش می‌دهد:

۱. بررسی داده‌های واردشده توسط کاربر، برای بهبود نتیجه شناسایی.

۲. بررسی اینکه وب‌سایت مورد نظر از میزبان‌های وب رایگان استفاده می‌کند یا خیر (با توجه به اینکه طی بررسی‌های صورت گرفته، اغلب وب‌سایت‌های فیشینگ از میزبان‌های وب رایگان استفاده می‌کنند)

۳. تمرکز در حوزه بانکداری اینترنتی (با توجه به اینکه بیش از ۶۰٪ اهداف فیشرها، سرویس‌های پرداخت و امور مالی بوده است)

۴. ارائه طبقه‌بندی مناسب برای ایجاد چارچوب کاری برای بهره‌گیری از توان فنون مختلف شناسایی در کنار هم.

۵. امکان شناسایی وب‌سایت‌های فیشینگ فارسی‌زبان که سرویس‌های بانکداری اینترنتی در ایران را موردحمله قرار می‌دهند و همچنین تلاش در جهت شناسایی لحظه صفر این‌گونه موارد (برخلاف اغلب روش‌های شناسایی یادشده که امکان شناسایی وب‌سایت‌های فیشینگ فارسی‌زبان که سرویس‌های بانکداری اینترنتی در ایران را موردحمله قرار می‌دهند را ندارند و همچنین احتمال شناسایی لحظه صفر این‌گونه موارد توسط روش‌های یادشده، بسیار ضعیف است)

۷- مراجع

[1] S. S. Silva, R. M. Silva, and R. C. Pinto, "Botnets: A survey," Computer Networks, vol. 57, no. 2, pp. 372-403, 2013.

[2] A. Cole, M. Michael, and D. Noyes, "Botnets: The rise of the machines," In Proceedings on the 6th Annual Security Conference, 2007.

[3] M. Khonji, Y. Iraqi, and A. Jones, "Phishing Detection: A Literature Survey," IEEE Communications Surveys & Tutorials, vol. 15, 2013.

[4] H. Rouhani, "Detection of Phishing Websites Using Fast Flux Service Networks (Master's thesis)," Sharif University of Technology, 2010.

[5] PhishTank [Internet]. PhishTank; [cited 2016 May 27]. Available from: http://www.phishtank.com.

[6] A. Ramachandran, "Fishing for Phishing from the Network Stream," Georgia Tech. CSS Technical Report GT-CS-08-08, 2008.

[7] D. Ken, "Mobile Malware Attacks and Defense," Syngress Publishing, 2009.

[8] Gartner [Internet]. Gartner: [cited 2016 May 27]. Available from: http://www.gartner.com.

[9] Symantec, "Internet Security Threat Report," Symantec, 2014

[10] M. Aburrous, "Predicting Phishing Websites using Classification Mining Techniques with Experimental Case Studies," In Proceedings on the 7th International Conference on Information Technology, IEEE, 2010.

[11] P. Prakash, "Phishnet: predictive blacklisting to detect phishing attacks," In Proceedings of the 29th conference on Information communications, pp. 346–350, 2010.

[12] N. Chou, "Client-side defense against web-based identity theft," The Internet Society conference [NDSS], 2004.

[13] C. V. Zhou, "A self-healing self-protecting collaborative intrusion detection architecture to traceback fast-flux phishing domains," NOMS Workshops, IEEE, 2008.

[14] P. Likarish, D. Dunbar, and T. E. Hansen, "B-apt: Bayesian anti-phishing toolbar," IEEE International Conference on Communications, 2008.

[15] Y. Zhang, J. I. Hong, and L. F. Cranor, "Cantina: a content-based approach to detecting phishing web sites," Proceedings of the 16th international conference on World Wide Web, New York, USA, pp. 639–648, 2007.

[16] M. Sharifi and S. H. Siadati, "A Phishing Sites Blacklist Generator," IEEE, 2008.

[17] P. Singh and M. D. Patil, "Identification of Phishing Web Pages and Target Detection," International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), vol. 3, 2014.

[18] L. A. T. Nguyen, "A Novel Approach for Phishing Detection Using URL-Based Heuristic," IEEE, 2014.

[19] R. Panda and R. Tiwari, "Protection from Phishing Attacks by Exploiting Page Rank,Reputation and Source Code of the Webpage," India : International Journal of Advanced Research in Computer Science and Software Engineering, 2014.

[20] M. Hara, A. Yamada, and Y. Miyake, "Visual similarity-based phishing detection without victim site information," IEEE Symposium on Computational Intelligence in Cyber Security, pp. 30-36, 2009.

[21] K. Chen, "Fighting phishing with discriminative keypoint features," Internet Computing, IEEE, vol. 13, pp. 56-63, 2009.

[22] H. Zhang, "Textual and visual contentbased anti-phishing: A bayesian approach," IEEE Transactions on Neural Networks, vol. 22, pp. 1532 –1542, 2011.

[23] C. Whittaker, B. Ryner, and M. Nazif, "Large-scale automatic classification of phishing pages," NDSS, vol. 10, 2010.

[24] P. Likarish, D. Dunbar, and T. E. Hansen, "B-apt: Bayesian anti-phishing toolbar," IEEE International Conference on Communications, 2008.

[25] G. Liu, B. Qiu, and L. Wenyin, "Automatic detection of phishing target from phishing webpage," 20th International Conference on Pattern Recognition (ICPR), pp. 4153-4156, 2010.

Journal Of Electronical & Cyber Defence
Vol. 4, No. 3, 2016, Serial No.15

Providing a method for identifying phishing website of Internet Payment Service

G. Shahmohammadi*, S. Kamalizadeh

*Olum Entezami Amin University, Iran

(Received: 12/14/2015, Accepted: 01/08/2016)

ABSTRACT

Phishing is one of the social engineering techniques to deceive users to obtain confidential information such as user names, passwords or bank account information is used. The most important challenges in the Internet, is the risk of phishing attacks and internet fraudulent. So researchers have been attempted to identify and deal with such attacks. The aim of this study is to present a new method for identifying phishing website in Internet banking. The proposed method have not weaknesses of identifying and deal with phishing methods such as the lack of checking the status of free of charge website hosting and lack of focus on internet banking. This method uses the advantages of the different techniques to identify phishing website and identify possible moment zero Persian-language phishing websites that attack to Iran's Internet banking services. The proposed method is assessed by using a set of legal websites includes 500 pages indexed in links site and phishing websites set contains 100 phishing site detected in PhishTank and 20 phishing site designed and 11 phishing site identified in during writing the article. The results indicated that the proposed approach, does not consider phishing websites as legitimate (positive error of zero percent) and only 3% of phishing websites does not recognize (3% negative error).

Keywords: Phishing, Phishing Website, Deal with Phishing, Internet Attacks, Anti-Phishing

* Corresponding Author Email: shah_mohammadi@yahoo.co.uk