مقابله با حملات فیشینگ به کمک تکنیک‌های هوشمند تشخیص الگو در سیستم های پرداخت الکترونیک

نگارش:
سمیه سروش

چکیده

در دنیای امروز که دنیای ارتباطات و تجارت های الکترونیکی می باشد بیش از پیش باید مراقبت امنیت و تهدیدهای موجود درجامعه باشیم، اینترنت و شبکه های الکترونیک توانسته خواسته کاربران را در زمینه های شغلی ، دانشگاهی و همچنین انجام کلیه امور مالی سهل نماید. انجام تراکنش های مالی توسط سیستم های پرداخت الکترونیک مهمترین نسل ارتباطی مطرح در ادوار بانکداری بوده است. امروزه به دلیل ملاحظات بسیای که برای کاربران صورت گرفته جهت راحت شدن نحوه عملیات دریافت و پرداخت و همچنین عدم اتلاف وقت، جلوگیری از مراجعه حضوری به بانک ها، مراکز خرید و ... سیستم های پرداخت الکترونیک مطرح شده است.با توجه به اینکه سیستم های پرداخت الکترونیک توانسته جایگاه مناسبی بین کاربران برای خود ایجاد نماید اما به دلیل برخی مشکلات امنیتی موجود در شبکه های اینترنتی که عاملین آن هکرها هستند، کاربرانی هم وجود دارد که این نسل ارتباطی را تهدید دانسته و از ارتباط با آن خودداری می نمایند. درسیستم های پرداخت الکترونیکی که مدیریت آن توسط بانک عامل صورت می پذیرد باید امنیت طوری بررسی و طرح ریزی شده باشد که هکرها نتوانند به راحتی اعتماد کاربران را خدشه دار کنند .یکی از جرایم و حملات سایبری که تاکنون توانسته آسیب بسیاری به کاربران وارد نماید حملات فیشینگ بوده، در این روش فیشرها توانسته اند علی رغم وجود امنیت بالا، از طریق ایمیل و وب سایت های جعلی اطلاعات شخصی اشخاص را ربوده و سرمایه آنان را به تاراج ببرند.در این پایان نامه تکنیک های هوشمند تشخیص الگو با استفاده از الگوریتم شبکه عصبی تابع پایه شعاعی در ساختارهای مختلف برای عملکرد بهترین ساختار در شناسایی URL های فیشی، قانونی و مشکوک مورد بررسی قرار می گیرد که تا کنون با سیستم های یادگیری برخط در شناسایی الگو و تکنیک های وب کاوی آشنا شده و از دانسته های گذشته جهت بروز رسانی استفاده می کنیم.

واژگان کلیدی: سیستم های پرداخت الکترونیک، امنیت اینترنت،حملات فیشینگ، شبکه های عصبی پایه شعاعی، یادگیری برخط

مقدمه

فیشینگ را می‌توان نوعی تهدید آنلاین تعریف کرد که به عنوان بخشی از مداخله در یک وب سایت معتبر و با هدف بدست آوردن اطلاعات خصوصی کاربران مانند نام کاربری، رمز عبور، شماره های امنیتی اجتماعی مطرح می‌شود. وب سایت های فیشینگ، به وسیله ی افرادی سود جو ایجاد شده تا این هدف که وب سایت های معتبر را مورد تقلید قرار دهند. این وب سایت ها دارای سطوح بالایی از تشابه با وب سایت های اصلی بوده با این هدف که بتواند کاربران را فریب دهند. گزارشی که توسط شرکت گارتنر منتشر شده است نشان داده است که حملات فیشینگ به سرعت در حال رشد هستند. همچنین این شرکت تخمین زده است که سرقت‌هایی که به وسیله ی حملات فیشینگ صورت میگیرد، سالیانه حدود ۲٫۸ میلیارد دلار به بانک ها و شرکت های کارت های اعتباری در ایالات متحده خسارت وارد می‌کند. در سال ۲۰۱۱ میلادی، ندیر واحد تجاری– تکنولوژیک– امنیتی سیسکو، نگرانی های خود را در این مورد منتشر کرد که حملات اصلی امروزی، بر روی دسترسی به حساب های مالی شرکت هایی که دارای اطلاعات مالی حیاتی هستند متمرکز است گسترش فزآینده فن‌آوری اطلاعات و ارتباطات منجر به تحول و دگرگونی جوامع امروزی در ابعاد مختلف سیاسی، امنیتی، اجتماعی و اقتصادی شده است. زندگی جوامع امروزی به سیستم‌های رایانه‌ای، شبکه‌های رایانه‌ای و خدمات اینترنتی وابسته می‌باشد. اگر چه استفاده از فن‌آوری اطلاعات و ارتباطات دارای مزایای بسیار زیادی است ولیکن تهدیداتی را نیز در پی دارد. از جمله مهم‌ترین تهدیدات موجود در فضای سایبری می‌توان به تهدیدات درون‌سازمانی، تهدیدات برون‌سازمانی، نرم‌افزارهای مخرب[۱] و حملات فیشینگ[۲] اشاره نمود[۱].

---

۱. Malware

۲. Phishing Attacks

حملات فیشینگ:

حملات فیشینگ، صفحات وب جعلی[۳] هستند که توسط افراد با سوءنیت برای تقلید از وب‌سایت‌های واقعی و قانونی ایجاد می‌گردند. بسیاری از این نوع صفحات وب دارای شباهت‌های بصری بالایی برای کلاهبرداری از قربانیان می‌باشند. قربانیان وب‌سایت‌های فیشینگ ممکن است که حساب بانکی، رمز عبور، شماره کارت اعتباری و دیگر اطلاعات مهم خود را در اختیار صاحبان و طراحان وب‌سایت‌های فیشینگ قرار دهند[۲].

در نتیجه بروز چنین تهدیداتی است که امنیت در فضای سایبری به یک دغدغه عمومی برای جوامع بین المللی تبدیل شده است. بنابراین ارائه روش‌هایی برای مقابله با این نوع تهدیدات در فضای سایبری، بسیار حائز اهمیت می‌باشد. امنیت در فضای سایبری می‌تواند از جنبه‌های مختلف مورد بحث و بررسی قرار گیرد. در این پژوهش، تمرکز اصلی بر روی تهدیدات سایبری است که توسط اشخاص حقیقی و حقوقی به منظور سوءاستفاده و کلاه‌برداری مالی انجام می‌شود.

وب‌سایت‌های فیشینگ با راه‌اندازی یک حمله مهندسی اجتماعی[٤] تلاش می‌کنند که اطلاعات شخصی قربانیان از قبیل شماره کارت اعتباری، اطلاعات حساب بانکی و شماره امنیتی را به منظور سوءاستفاده در اختیار بگیرند. حملات فیشینگ تأثیرات منفی زیادی بر درآمد سازمان‌ها، روابط مشتریان و تلاش‌های بازاریابی دارد. حملات فیشینگ می‌تواند ده‌ها یا صدها هزار دلار در هر حمله به یک شرکت و سازمان خسارت وارد کند. در برخی موارد هزینه‌های مربوط به صدمه زدن به نام تجاری و کاهش اطمینان مشتریان به میلیون‌ها دلار می‌رسد[۳]. رشد روز افزون وب‌سایت‌های فیشینگ به یک چالش بسیار بزرگ در زمینه تجارت الکترونیک و به ویژه بانکداری الکترونیک تبدیل شده است.

بیان مسئله

فیشینگ تکنیک مهندسی اجتماعی به منظور گمراه کردن کاربران اینترنتی برای بدست آوردن اطلاعات محرمانه آنان است. در این تکنیک فیشرها (کسانی که عمل فیشینگ را انجام می دهند) با طراحی یک سایت که شبیه به سایت مورد نظر می باشد ، کار خود را آغاز می کنند.پس از طراحی سایت به دنبال راهی هستند تا قربانیان را مجبور به ورود به سایت خود نموده و اطلاعات محرمانه خود نظیر اطلاعات اعتباری جزئیات کارت، جزئیات حساب بانکی، رمزهای عبور و غیره را در اختیار می‌گیرند. ممکن است خود شما هم تا به حال به طور ناخواسته و بدون اینکه متوجه شده باشید، یکی از قربانیان فیشینگ شده باشید (به اصطلاح در قلاب ماهیگیر افتاده باشید)

---

۳. Forged Web pages

٤. Social Engineering Attack

فیشینگ در ادبیات اینترنت مطرح شده و هدف اصلی آن ایجاد نیرنگ برای افشای اطلاعات حساس و شخصی کاربران است. مهاجمان به جهت رسیدن به اهداف مخرب خود اولین درخواست را برای افراد زیادی ارسال و منتظر پاسخ می مانند آنها امیدوار هستند که بتوانند حتی تعداد کمی از کاربران را مجبور به افشای اطلاعات شخصی خود نمایند. البته ناگفته نماند که مهاجمان در کارخود موفق بوده اند و توانسته اند قربانیان زیادی را در این قلاب گرفتار و شانس موفقیت آنها به لحاظ آماری افزایش یافته است. رمز موفقیت فیشرها جلب رضایت و اعتماد کاربران بوده که توانسته ضریب موفقیت آنها را افزایش دهد.کاربران اینترنتی بدلیل اینکه اطلاعات و آگاهی کافی در مقابل چنین حملاتی را ندارند در نتیجه زمینه را برای بالابردن شانس موفقیت مهاجمان جهت سرقت هویت کاربران افزایش میدهند.

در کل حملات کلاهبرداری اینترنتی باروش های زیر انجام می شود :

- ابتدا کلاهبردار وب سایتی جعلی ایجاد می کند که مشابه وب سایت مجاز است
- آنگاه کلاهبردار لینک هایی را به وب سایت در قالب پیام های کلاهبرداری به کاربر هدف می فرستد. به دلیل اینکه کلاهبردار تلاش می کنند قربانیان را متقاعد کنند تا وب سایت های تقلبی آنها را ببینند.
- با کلیک روی لینک قربانیان وب سایت های جعلی را مشاهده کرده و اطلاعات محرمانه شان را آنجا وارد می کنند.
- آنگاه کلاهبردار اطلاعات محرمانه را می دزدد و در کلاهبرداری شان مثل انتقال پول از حساب قربانی از آن استفاده می کند.

پس از توسعه خدمات مالی آنلاین و تجارت الکترونیک، حملات وب‌سایت‌های فیشینگ به یکی از حملات خطرناک و شایع در بستر اینترنت تبدیل شده است. صفحات وب فیشینگ در سال‌های اخیر با سرعت زیادی در حال افزایش هستند. برای جلوگیری از وب‌سایت‌های فیشینگ هر یک از دو طرف درگیر در تجارت الکترونیک (سازمان مالی آنلاین و مصرف‌کننده) باید فیشینگ و فناوری‌های ضد فیشینگ را درک نمایند و اقدامات امنیتی لازم را اعمال نمایند. دامنه و پیچیدگی وب‌سایت‌های فیشینگ به سرعت در حال افزایش است به طوریکه در سالیان اخیر از یک فعالیت غیرحرفه‌ای کم هزینه به یک جرم سازمان‌یافته اینترنتی تبدیل شده است. حملات وب‌سایت‌های فیشینگ نه تنها تأثیرات منفی زیادی بر درآمد شرکت‌ها و سازمان‌های مالی دارد بلکه باعث سلب اعتماد کاربران از تجارت الکترونیک و بانکداری الکترونیک می‌گردد[۵].

دراین تحقیق سوال مهمی که مطرح می شود این است که چگونه بتوانیم با تکنیک های هوشمند تشخیص الگو امنیت بالایی برای کاربران استفاده از سیستم های پرداخت الکترونیک در بستر اینترنت فراهم کنیم؟

پیشینه تحقیق

فن آوری های جدید ارتباطات فرصت های جدید را برای تجارت الکترونیک و همچنین فرصت های جدید را برای انجام عملیات جرایم سایبری فراهم نموده است. کارشناسان خبره معتقدند که بزرگترین خطرات امنیت سایبری در چند سال آینده شامل تهدیدات مداوم ، محاسبات موبایل و استفاده از رسانه های اجتماعی است. دو روش مهندسی اجتماعی عبارتند از :

۱) پیش متن[۵]: استفاده از یک داستان که شخصی را متقاعد به افشای اطلاعات محرمانه می نماید. به عنوان مثال، یک هکر براحتی با ا ستفاده از شماره تلفن و نام در د سترس با تماس به عنوان یک کارمند ادعا می کند که مدیر سیستم که نیاز به تنظیم مجدد کلمه عبور برای محافظت از شرکت از هک است . در این حالت نام کاربری و پسورد را دریافت می کند.

۲) طعمه: ا ستفاده از، انگیزه ای برای یک کاربر برای انجام یک عمل نا امن ا ست. نمونه های از طعمه ارائه یک برنامه رایگان و یا ویدئو را برای کلیک کردن روی یک لینک در یک پیام متنی و رای گیری برای بهترین بازی های ویدئویی است. با کلیک بر روی نرم افزارهای مخرب دریافت لینک.

حملات امنیتی مختلفی در شبکه های کامپیوتری وجود دارد، یکی از مهمترین آنها حملات فیشینگ است که قصد سرقت اطلاعات مالی و بانکی کاربران را دارد. تمرکز این تحقیق روی شناسایی حملات فیشینگ به کمک آدرس های URL ایمیل های ارسالی می باشد.

عوامل موثر در تهدید و حمله در یک شبکه اینترنتی

۱- دسترسی بدون محدودیت به اینترنت
۲- ناشناس و گمنام بودن افراد
۳- سرعت بالای انتشار
۴- نداشتن ارتباط چهره به چهره
۵- دسترسی آزاد به خدمات و محتویات ارزشمند و همچنین نبود قوانین و توافقات مناسب

مجموعه تهدیدات و حملات فوق سبب ایجاد سختی پیگرد قانونی می شود. بانکداری الکترونیک خطراتی را برای مؤسسات اقتصادی بوجود آورده که این مؤسسات باید در برنامه های کاری خود کنترل و غربال گری و مدیریت ریسک جامع را ببینند. امنیت بانکداری الکترونیک مهمترین موضوع مطرح شده در تجارت الکترونیک است، هرچه خدمات و امکانات رفاهی الکترونیکی در معاملات افزایش پیدا کند، مبحث امنیت با دید بازتری باید مورد بررسی و طرح قرار گیرد[۸].

اهداف تحقیق

همانطور که عنوان شد، هدف اصلی از مطالعه این پژوهش، ارائه یک روش جدید ضد فیشینگ مبتنی بر تکنیکهای هوشمند ت شخیص الگو می‌با شد. برای ایجاد امنیت در ب ستر بانکداری الکترونیکی جهت انجام تراکنش‌ها با حداکثر سرعت و دقت لازم اهداف مورد نظر ارزیابی تکنیک هو شمند و در صد دقت قابل قبول آن سیستم واینکه در مقابل چه تعداد حمله مقاوم است برای ما دارای اهمیت است.

مقایسه ای در باره ی انواع تکنیک های ضد فیشینگ

روش مبتنی بر URL ن سبت به دیگر روش های ت شخیص فی شینگ سریعتر و مقیاس پذیرتر ه ستند زیرا آنها می توانند یک URL را با استفاده از ویژگی های ساختاری خود همان URL و برخی از ویژگی های باینری که نشان دهنده حضور

---

۵. Pretexting

کلمات خاص در آن اســت، بدون هرگونه جمع آوری اطلاعات بیشــتر طبقه بندی نمایند، رویکردی که تنها در ویژگی های مبتنی بر URL ارائه می شـــود. این روش نیاز به یک مقدار کافی از دانش دامنه برای انتخاب کلمات مربوط به ویژگی ها دارد.

اطلاعات DNS، خواص جغرافیایی، سرعت اتصال و عضویت در لیست سیاه، این ویژگی های با الگوریتم های طبقه بندی آنلاین مانند الگوریتمهای ماشین بردار پشتیبانی (SVM)، رگرسیون لجستیک و AROW و CW و بیز بکار گرفته می شود. AROW و CW دارای دقت بیشتری بوده و در عین حال حافظه و سربار محاسباتی کمتری در مقایسه با دیگر تکنیک ها را تحمیل می کنند. AROW نسبت به CW عملکرد بهتری دارد. روش های مبتنی بر محتوا بدلیل واکشی و تجزیه و تحلیل کامل محتوای صفحه وب زمانبرتر از سایر روش های تشخیص فیشینگ هستند.

در سـال های اخیر بسـیاری از الگوریتم های یادگیری ماشـین برای حل مشـکلات طبقه بندی توسـعه یافته اند و به طور گسترده ای برپا تشخیص وب سایت های فیشینگ بکار گرفته شده اند. الگوریتم ماشین بردار پشتیبان، به اختصار SVM، یک الگوریتم طبقه بندی سنتی است که در محاوره، به آن مدل طبقه بندی باینری می‌گویند که در شناسایی فیشینگ بسیار محبوب است این الگوریتم دقت بالایی در طبقه بندی از خود نشان داده است، اما الگوریتم SVM سرعت آموزش کندتری در زمان هایی که با مجموعه های آموزشی بزرگ مواجه می شویم ارائه می کند و این نقطه ضعف در کاربردهای عملی است.

SVM قابلیت های خوبی در تشخیص وب سایت های فیشینگ دارا ست. در مقابل ، الگوریتم BVM نه تنها می تواند پیچیدگی مدل طبقه بندی را کاهش دهد، در عین حال دقت تشخیص وب سایت فیشینگ را بهبود می بخشد. الگوریتم طبقه بندی BVM مبتنی بر الگئریتم SVM است. با افزایش تعداد صفحات فیشینگ در نمونه آزمایشی، دقت الگوریتم BVM به شکل محسوسی از SVM بهتر است. سرعت آموزش BVM از SVM سریعتر است. بنابراین تشخیص وب سایت های فیشینگ با بکارگیری الگوریتم BVM از لحاظ زمانی بهینه تر از SVM می باشد. تکنیک Cantina با بکارگیری TF-IDF تاحدودی توانست بر برخی از مشکلات یافت نشده صفحه وب غلبه کند. TF-IDF یک الگوریتم بازیابی اطلاعات شناخته شده است که میتواند برای مقایسه و طبقه بندی اسناد، همچنین استخراج متن و بازیابی اسناد از یک مجموعه بزرگ استفاده نماید. در الگوریتم TF-IDF به لغات، بر اساس فراوانی آن در سند یک وزن داده می شود. در واقع این سیستم وزن دهی نشان می دهد چقدر یک کلمه برای یک سند مهم است. نسبت اهمیت کلمه با فراوانی آن در متن افزایش می یابد و بوسیله فرکانس کلمه در مجموع متعادل می شود. با این حال بدلیل مثبت کاذب بالا و تعداد محدود ویژگی های مؤثر برای تشـــخیص فیشـــیگ، دارای نقاط ضـــعف اســـت. در مقابل + Cantina که توســـعه ای از روش Cantina است توانست با افزایش ویژگی های مورد بررسی، میزان مثبت کاذب را تا سطح قابل قبولی کاهش داده و بر نقاط ضعف تکنیک های مبتنی بر لیست سیاه و مبتنی بر ویژگی غلبه و حملات تکراری را بلافاصله شناسایی نماید.

راه حل ها مختلفی جهت مقابله با فیشـــینگ توســـعه داده شـــده اند. اما با این حال هنوز هم فقدان دقت و راه حل های بلادرنگ مشکل اساسی است چون ترکیبی از ویژگی های کمی مانند تعداد اتصال یا عدم اطمینان می‌باشد. بیشتر الگوریتم

های یادگیری ما شین پارامترهای تعیین شده ای ذلرند، اما وفق دادن پارامترها با یک خروجی مطلوب د شوار ا ست. منطق فازی دقت بیشتری در مرحله تصمیم گیری فراهم می کند.

تکنیک های عصبی- فازی ترکیبی از منطق فازی و شبکه های عصبی هستند. این شبکه ساختاری متشکل از گره ها و ارتباطات است که از طریق آن گره ها به یکدیگر مرتبط می شوند. تنظیم پارامترهای مناسب با مجموعه ویژگی های جامع با بکارگیری سیستم عصبی- فازی میتواند عملکرد سیستم را بهبود ببخشد. سیستم عصبی- فازی توانایی تولید قوانین فازی با ویژگی های داده شده و همچنین قابلیت یادگیری ویژگی های جدید را دارد. تنظیم پارامتر بر ا ساس سی ستم ع صبی- فازی با ویژگی هاتی جامع می تواند عملکرد سیستم را در تشخیص بلادرنگ ارتقا دهد.

تکنیک های قطعی به اندازه کافی در تصمیم گیری انعطاف پذیر نیستند. در عین حال تکنیک های تکاملی همانند الگوریتم ژنتیک (GA) قادر است بر محدودیت های بسیاری از سیستم های تشخیص نفوذ جاری غلبه کن و منحصرا برای سیستم های ت شخیص نفوذ منا سب می با شد. خا صیت بهینه سازی تکاملی مزایای متعددی ن سبت به روش های سنتی دارد که عبارتند از ا ستحکام، یادگیری بدون نظارت، توانایی پیدا کردن یک راه حل تقریبا بهینه در ف ضاهای بزرگ و عملیات موازی ذاتی، خواص الگوریتم ژنتیک شامل ا ستحکام در برابر اختلال، قابلیت خود یادگیری و توانایی ساخت قوانین اولیه بدون نیاز به دانش پیشین می باشد.

آنها ذاتا موازی اند بدلیل اینکه آنها چندین فرزد تولید می کنند که به طور همزمان فضای راه حل را در چندین جهت کشف می کنند. هنگامیکه ف ضای راه حل بسیار بزرگ ا ست موازی سازی باعث می شود که این الگوریتم ها به خوبی با شرایط سازگار شوند که این امر به سیستم اجازه می دهد تا برای تکامل قوانین جدید به راحتی دوباره تعلیم ببینند.

در اکثر تکنیک های مبتنی بر تحلیل شمای بصری ابتدا ت صویری از صفحه وب گرفته می شود. پس از گرفتن ت صویر از صفحه وب با کمک الگوریتم EMD شباهت بصری صفحه وب مورد سنجش قرار می گیرد. در این تکنیک ها اگر مهاجم صفحه وبی بسیار مشابه به معادل واقعی آن بسازد آنگاه ممکن است با شکست مواجه شود. همچنین پس از گرفتن تصویر می توان از سیستم ت شخیص نوری (OCR) ا سفده نمود که برخلاف سایر تکنیک های مبتنی بر تحلیل شمای بصری قادر به تشخیص حمله روز صفر می باشد اما نسبت به حمله های ناشی از سیستم جستجوی گوگل آسیب پذیر است.

درمقابل سیستم وزنی قدرت منا سبی در شنا سایی وب سایت های فیشینگ از طریق نام برند در محتوای HTML دارد، تکنیک ا ستاندارد CSS و مبتنی بر تصویر لوگو از دیگر روش های مبتنی بر تحلیل شمای بصری هستند که راه حل های ارایه شده بر اساس CSS دارای دقت بسیار بالایی در تشخیص فیشینگ نسبت به سایر روش های تحلیل شمای بصری می با شند و برخلاف روش های مبتنی بر لوگو در ت شخیص هویت برا ساس اطلاعات متنی دارای ضعف نیستند. اما مزیت روش های مبتنی بر لوگو در ت شخیص هویت برا ساس اطلاعات متنی دارای ضعف نی ستند. اما مزیت روش های مبتنی بر لوگو عدم محدودیت در تشخیص وب سایت های فیشینگ غیر انگلیسی است.

روش تحقیق

فیشینگ تلاشی از طرف یک فرد یا گروهی از اشخاص برای دزدیدن اطلاعات شخصی مانند گذرواژه[۶]، حساب بانکی و اطلاعات کارت اعتباری است. بیشتر این صفحات وب فیشینگ از نظر رابط وب‌سایت[۷] و آدرس مکان‌یاب منبع یکپارچه[۸] (URL) مشابه صفحات وب واقعی به نظر می‌آیند. تکنیک‌های بسیاری برای شناسایی وب سایت‌های فیشینگ مانند تکنیک مبتنی بر لیست سیاه[۹]، تکنیک مبتنی بر هیوریستیک[۱۰] و غیره پیشنهاد شده‌اند. اگرچه، این تکنیک‌ها هنوز ناکارآمد هستند. در حالی که تکنیک‌های مبتنی بر لیست سیاه نمی‌توانند سایت‌های فیشینگی را که در پایگاه داده‌ی[۱۱] لیست سیاه نیستند را شناسایی کنند، وزن‌های هیوریستیک[۱۲] در روش مبتنی بر هیوریستیک به صورت چشمگیری به آمار برگرفته از مجموعه داده‌ی آموزشی[۱۳] بستگی دارند. بنابراین، یک روش شناسایی فیشینگ جدید که در آن وب سایت‌های فیشینگ جدید می‌توانند شناسایی شوند یا وزن‌های هیوریستیک استخراج شوند، لازم است. در این مقاله، روش کارآمدی برای شناسایی وب سایت‌های فیشینگ مبتنی بر شبکه‌ی عصبی تک لایه تابع پایه شعاعی RBF پیشنهاد شده است. به طور مشخص، روش پیشنهادی مقدار هیوریستیک‌ها را به صورت بی طرفانه محاسبه می‌کند. سپس، وزن‌های هیوریستیک توسط یک شبکه‌ی عصبی تک لایه پایه شعاعی RBF ایجاد می شود. تکنیک پیشنهادی به همراه مجموعه داده‌ای از ۱۱۶۶۰ سایت فیشینگ و ۱۰۰۰۰ سایت قانونی[۱۴] ارزیابی می‌شود. نتایج نشان می‌دهند که این تکنیک می‌تواند بالای ۹۸٪ از سایت‌های فیشینگ را شناسایی کند.

بر اساس بسیاری از مطالعات، بیشتر سازمان های مالی و دولتی سرویس های آنلاین خود را برای مشتری‌هایشان گسترش داده‌اند. برای مثال، در سال ۲۰۱۱، ۸۳٪ آمریکایی‌ها و ۸۵٪ اروپایی‌ها به طور منظم به صورت آنلاین خرید می‌کردند (مجله‌ی Fortune، ۲۰۱۱). در نتیجه، نوع جدیدی از جرم الکترونیکی که به آن فیشینگ گفته می‌شود، ایجاد شده است. فیشینگ به عنوان ترفند جنایی آنلاین[۱۵] برای دزدیدن اطلاعات شخصی کاربران توسط فرستادن ایمیلی به کاربران و وادار کردن آن‌ها برای بازدید از یک سایت تقلبی که مشابه سایت اصلی به نظر می‌رسد، در نظر گرفته می شود. فیشینگ باعث تلفات اقتصادی در سر تا سر جهان شده است. در ایالات متحده، فیشینگ باعث ۳,۲ میلیون تلفات در سال ۲۰۰۷ شد. بنابراین، مسئله‌ی ضد فیشینگ[۱۶] یک مسئله‌ی حیاتی در جامعه‌ی مدرن است. مطالعات بسیاری برای شناسایی فیشینگ از جمله روش لیست سیاه، روش مبتنی بر هیوریستیک و غیره پیشنهاد شده است. اگرچه

---

۶. Password

۷. Interface website

۸. Uniform

۹. Blacklist-based

۱۰. Heuristic-based

۱۱. Database

۱۲. The weights of the heuristic

۱۳. Training dataset

۱۴. legitimate

۱۵. Online criminal trick

۱۶. Anti-phishing

هنوز یک روش مناسب وجود ندارد. در این مقاله، روشی جدید برای شناسایی سایت‌های فیشینگ که روی مشخصه‌های URL (دامنه‌ی اولیه۱۷، زیردامنه۱۸، دامنه‌ی مسیر۱۹) و رتبه‌بندی سایت (رتبه‌ی صفحه۲۰، رتبه‌ی Alexa۲۱، اعتبار Alexa۲۲) تمرکز دارد، پیشنهاد شده است. سپس، یک شبکه‌ی عصبی تک لایه پیشنهاد شده است که سیستمی است که خطا را کاهش داده و کارایی را افزایش می‌دهد.

امروزه، فیشرها۲۳ از بسته‌های ابزار نرم‌افزاری۲۴ برای ایجاد تعداد زیادی از URLهای سایت فیشینگ استفاده می‌کنند. بنابراین، روش مبتنی بر لیست سیاه نمی‌تواند به طور کارآمد وب سایت‌های فیشینگ را شناسایی کند. روش‌های شناسایی فیشینگ دیگر از جمله روش‌های هیوریستیک، روش‌های فراگیری ماشین۲۵ و روش‌های هیبرید۲۶ توجه بیشتری از تحقیقات را به خود معطوف کرده‌اند. Zhang و همکارانش ابزاری به نام Cantina در سال ۲۰۰۷ ارائه کردند. این روش از الگوریتم TF-IDF مبتنی بر ۲۷ مشخصه‌ی صفحه‌ی وب استفاده کرده است. این تکنیک می‌تواند ۹۷٪ سایت‌های فیشینگ را با ۶٪ مثبت‌های کاذب۲۷ شناسایی کند. اگرچه این تکنیک کارآمد است، اما ۲۷ مشخصه‌ی صفحه‌ی وب استخراج شده‌اند که در آن برخی مشخصه‌ها برای بهبود دقت شناسایی فیشینگ لازم نیستند. به علاوه، از آن جا که Cantina هیوریستیک‌های بسیاری را معرفی می‌کند، منجر به محاسبات پیچیده و در نتیجه زمان پاسخ طولانی می‌شود. در نتیجه، در سال ۲۰۱۱، Cantina به +Cantina ارتقا یافت. +Cantina از تکنیک‌های یادگیری ماشین مبتنی بر ۱۵ مشخصه‌ی صفحه‌ی وب استفاده کرد. اگرچه، تنها ۶ مشخصه از ۱۵ مشخصه برای شناسایی فیشینگ از جمله "فرم بد۲۸"، "حوزه‌های فعالیت بد۲۹"، "URLهای نامطابق۳۰"، "صفحه در نتایج جستجوی برتر۳۱"، "نشان حق کپی‌رایت جستجو به اضافه‌ی دامنه۳۲" و "نشان حق کپی‌رایت جستجو به اضافه‌ی نام میزبان۳۳" کارآمد هستند. در ، نویسنده از

---

۱۷ . PrimaryDomain

۱۸ . SubDomain

۱۹ . PathDomain

۲۰ . PageRank

۲۱ . AlexaRank

۲۲ . AlexReputation

۲۳. Phishers

۲۴ . Software tool-kits

۲۵ . Machine learning

۲۶ . Hybrid

۲۷ . False positives

۲۸ . Bad form

۲۹ . Bad action fields

۳۰ . Non-matching URLs

۳۱ . Page in top search results

۳۲ . Search copyright brand plus domain

۳۳ . Search copyright brand plus hostname

URL برای شناسایی خودکار سایت‌های فیشینگ توسط استخراج و بررسی کلمات مختلف URL از طریق موتور جستجو[۳۴] استفاده کرده است. اگرچه این مقاله تکنیک جدید جالبی ارائه داده است، اما نرخ شناسایی نسبتا پایین است (۵۴,۳٪). روشی مبتنی بر محتوا[۳۵] برای شناسایی فیشینگ ایجاد کرده است. به خصوص، نویسنده رتبه‌ی صفحه‌ی گوگل[۳۶] مربوط به یک صفحه را برای ارزیابی سطح فیشینگ صفحه‌ی وب در نظر می‌گیرد. اگرچه، رتبه‌ی صفحه‌ی گوگل ممکن است منجر به چندین شناسایی نادرست در مواردی که وب سایت قانونی به تازگی ایجاد شده باشد و رتبه‌ی صفحه‌ی گوگل پایینی داشته باشد، شود. روش دیگر برای شناسایی وب سایت‌های فیشینگ استفاده از کد مرجع است . اگرچه این روش ممکن است کارآمد باشد، اما زمان شناسایی طولانی خواهد بود. برای غلبه بر مشکلات موجود در روش‌های شناسایی فیشینگ پیشنهاد شده ارائه شده است. اگرچه، مقدار هیوریستیک‌ها و وزن‌ها به آمارهای مجموعه داده‌ی آموزشی بستگی دارد.

کار ما با بقیه در ۳ جنبه تفاوت دارد:

۱. هیوریستیک‌های جدید برای شناسایی کارآمدتر و سریعتر وب‌سایت فیشینگ ارائه شده‌اند.
۲. مقدار هیوریستیک‌ها به صورت بی طرفانه محاسبه شده و به مجموعه داده‌ی آموزشی بستگی ندارند.
۳. وزن هیوریستیک‌ها بهینه‌تر هستند زیرا وزن‌ها توسط شبکه‌ی عصبی آموزش یافته‌اندطراحی سیستم

شبکه های عصبی مصنوعی[۳۷]

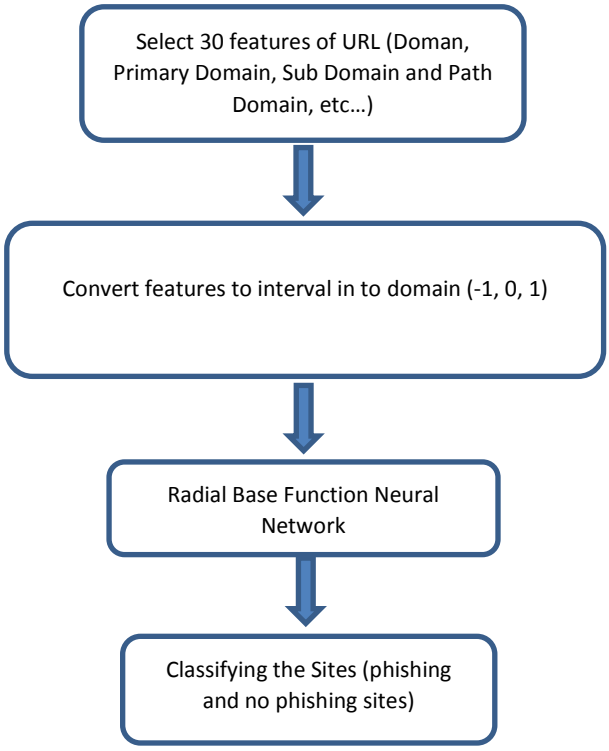
شبکه‌های عصبی مصنوعی (ANNs) به دلیل عملکرد موفق در مسائل مختلف طبقه‌بندی جهان واقعی به خصوص در صنعت، تجارت و علم، شناخته شده هستند .. ANNها مدل پردازش اطلاعات[۳۸] هستند که از روش پردازش اطلاعات انسان الهام گرفته‌اند. ANNها به فراگیری دانش از طریق یک فرایند یادگیری نیاز دارند، در حالی که از قدرت اتصال بین نورونی[۳۹] که به عنوان وزن‌های سیناپتیک[۴۰] شناخته می‌شود، برای ذخیره‌ی دانش استفاده می‌شود . بنابراین با این قابلیت‌ها، شبکه‌های عصبی راه حل مناسبی برای مسائل طبقه‌بندی یا شناخت الگو فراهم می‌کنند. آموزش شبکه‌ی عصبی مسئله‌ی پیچیده‌ای است که برای یادگیری نظارتی[۴۱] حائز اهمیت است. جذاب‌ترین مشخصه‌ی شبکه‌ی عصبی امکان یادگیری است. شبکه زمانی که مثال‌ها با نتایج مشخص شده به آن ارائه می شود، یاد می‌گیرد. عوامل وزن‌دهی[۴۲] توسط الگوریتم

---

[۳۴] . Search engine
[۳۵] . Content-based
[۳۶] . Google page rank
[۳۷] . Artificial Neural Networks
[۳۸] . Information processing model
[۳۹] . Interneuron connection strength
[۴۰] . Synaptic
[۴۱] . Supervised learning
[۴۲] . Weighting factors

برای نزدیک‌تر کردن خروجی نهایی به نتایج مشخص شده تنظیم می‌شوند. یک URL (مکان‌یاب منبع یکپارچه) برای مکان‌یابی منابع استفاده شده است . فیشرها معمولا تلاش می‌کنند تا سایت‌های فیشینگ را برای کاربران آنلاین احمق به سایت‌های قانونی شبیه سازند. آن‌ها نمی‌توانند از URL دقیق مربوط به سایت قانونی استفاده کنند، آن‌ها بیشتر از اشتباه املایی[۴۳] برای مشخصه‌های URL از جمله دامنه‌ی اولیه، زیردامنه و دامنه‌ی مسیر استفاده می‌کنند. برای مثال، URLای مانند www.applle.com که مشابه وب‌سایت شناخته شده‌ی www.apple.com یا http://www.apple.attack.com می‌باشد، اگر کاربران هوشیار نباشند تصور خواهند کرد که در سایت "apple" هستند. واضح است که وب‌سایت‌های فیش شده[۴۴] نه توسط کاربران دسترسی پیدا کرده و نه توسط دیگر وب‌سایت–ها پیوند یافته‌اند[۴۵]. بنابراین، رتبه‌بندی سایت مانند رتبه‌ی صفحه، رتبه‌ی Alexa، اعتبار Alexa نیز می‌تواند برای شناسایی سایت‌های فیشینگ کمک کند. فیشرها معمولا از سایت‌های معروف سایت تقلبی ایجاد می‌کنند، اما رتبه‌بندی سایت تقلبی بالا نیست. همچنین، می‌توانیم از رتبه‌بندی‌ها برای طبقه‌بندی یک سایت برای فیشینگ بودن یا نبودن سایت استفاده کنیم.

طراحی مدل سیستم



شکل ۳–۱: مدل سیستم پیشنهادی

[۴۳] . Spelling mistakes

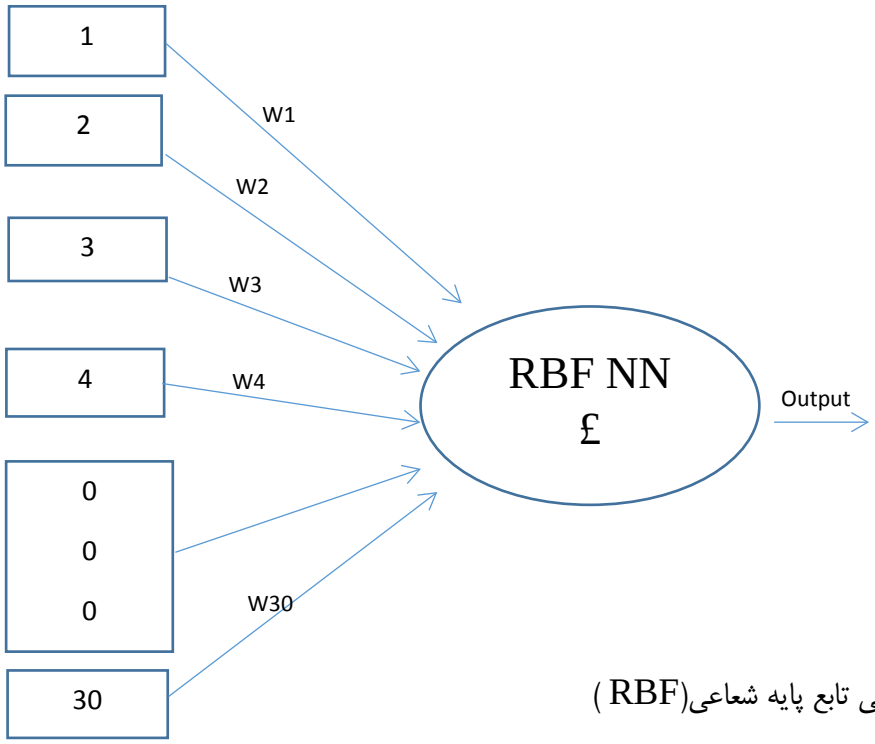
[۴۴] . Phished

[۴۵] . linked

۱) فاز ۱- انتخاب ۳۰ م شخ صه‌ی URL: ۴ م شخ صه از URL مانند دامنه، دامنه‌ی اولیه، زیردامنه و دامنه‌ی م سیر و غیره استخراج می‌شوند.
۲) فاز ۲- تبدیل ویژگی ها به مقادیر (۱-و۰و۱).
۳) فاز ۳- شبکه‌ی عصبی: شبکه‌ی عصبی تابع پایه شعاعی (RBF) برای محاسبه‌ی مقدار گره‌ی خروجی اجرا می‌شود.
۴) فاز ۴- طبقه‌بندی وب‌سایت‌ها: بر اساس مقدار گره‌ی خروجی تصمیم می‌گیریم که یک وب‌سایت، وب سایت فیشینگ است یا نه.

۳-۳- مدل شبکه عصبی

مدل شبکه‌ی عصبی تابع پایه شعاعی به صورت شکل ۲-۲ طراحی شده است.



شکل ۳-۲ : مدل شبکه‌ی عصبی تابع پایه شعاعی(RBF)

مدل دارای دو لایه شامل یک لایه‌ی ورودی و یک لایه‌ی خروجی است. لایه‌ی داخلی شامل ۳۰ گره که ۳۰ هیوریستیک مانند دامنه‌ی اولیه، زیردامنه، دامنه‌ی مسیر، رتبه‌ی صفحه، رتبه‌ی Alexa، اعتبار Alexa و غیره هستند، است. لایه‌ی خارجی یک گره‌ی خروجی دارد. شبکه‌ی ع صبی پی شنهادی از تابع فعال سازی سیگموئید[۴۶] ا ستفاده می‌کند، مقدار گره‌ی خروجی در محدوده‌ی ۰ تا ۱ قرار دارد. مدل پیشنهادی به دو کلاس طبقه‌بندی می‌شود، بنابراین اگر مقدار گره‌ی خروجی کمتر از ۰,۵ باشد سایت از نوع فیشینگ و اگر مقدار گره‌ی خروجی بزرگتر یا برابر ۰,۵ باشد، سایت قانونی است.

[۴۶] . Sigmoid activation function

نتیجه گیری

در این تحقیق ما تعداد ۱۱۰۸۶ داده را با ۳۰ ویژگی ترکیب کرده ایم و یک خروجی از آن حا صل شده ا ست که نتیجه هر خروجی با ساختار مورد نظر در بالا توصیف شده است. در اینجا برخی از ویژگیهای مؤثر در شناسایی وب سایت را ارائه نمودیم.

جدول(۴-۱) نمونه ای از ویژگیهای مؤثر را به همراه ارزش آن معرفی می کند.

جدول ۴-۱: ویژگیهای مؤثر در شناسایی وب سایت

Feature	Value
Using IP address	1
Long URL	0
URL havine @ symbol	0
Adding Perfix and Sufiix	1
Sub Domain(S)	1
Misuse of HTTPS	0
Request URL	1
URL of Anchor	-1
Server Form Handler	1
Abnormal URL	1
Redirect Page	-1
Using Pop- Up Window	-1
Hiding Suspicious Link	0
DNS Record	1
Websites Traffic	1
Age Of Domain	0
Disabling Right click	1

برای هر ساختار تعداد داده های تحت آموزش مورد بررسی قرار گرفته اند که ۳۲ داده اول همان مشخصات هدر می باشند در این تحقیق ما برای هر ساختار تعدادی داده را آموزش داده ایم و نتایجی از آن حاصل شده است که در جدول (۴-۲) مینیمم جمع مربعات خطا[۴۷] را برای آموزش و آزمایش نشان داده شده است .

جدول۴-۲: نتیجه ارزیابی

[۴۷]. Minimum Sum Square Error(MSE)

ساختار	آموزش مینیمم جمع مربعات خطا	آزمایش مینیمم جمع مربعات خطا
Phishing Exact RBFN	1.15×10^{-28}	۰,۱۱۱۵
Phishing GRNN	0	۰,۰۸۸۰
Phishing MLP[58]	1.57×10^{-10}	۱,۸۹۷۳
Phishing RBFN	۰,۳۹۶۴	۰,۳۹۰۸
Phishing RBFN Bayse	۰,۳۹۸۴	۰,۳۹۲۳
Phishing SVM[59]	۰,۱۲	۰,۸۹

در این مقاله، تکنیکی جدید برای شناسایی کارآمد سایت‌های فیشینگ ارائه کرده‌ایم. در تکنیک پیشنهادی، مدل سیستم برای شناسایی سایت‌های فیشینگ توسط شبکه‌ی عصبی دو لایه و 30 هیوریستیک (دامنه‌ی اولیه، زیردامنه، دامنه‌ی مسیر، رتبه‌ی صفحه، رتبه‌ی Alexa، اعتبار ... , Alexa) طراحی شده است. این تکنیک به همراه مجموعه داده‌ی آموزشی از پایگاه داده سایت معتبر PhishTank می باشد که ۱۱۰۸۶ داده در مدت ۳۰ روز جمع آوری و در ساختارهای مختلف از الگوریتم شبکه عصبی آموزش و نتیجه آزمایش نشان داده شده است، پیش از این ساختار Phishing MLP مورد بررسی واقع شده بود، کار جدید ما روی ساختارهای RBF، شبکه عصبی تابع پایه شعاعی می باشد که در بهترین نتایج نشان می‌دهند که ساختار رگرسیون محلی با ٪۹۹,۲۲ برای سایت‌های فیشینگ توسط تکنیک پیشنهادی شناسایی شده‌اند. دریافتیم که از بقیه بهینه‌تر است. در آینده، سیستم می‌تواند با استفاده از مجموعه داده‌های بزرگتر و پارامترهای هیوریستیک بیشتر بهبود یابد.

روش پیشنهادی

در این قسمت الگوریتم های شناسایی برخط را روی یک مجموعه داده مرتبط با حملات فیشینگ ارزیابی می شوند. دیتا بیس تهیه شده مرتبط با ۳۰ روز می باشد که در هر روز ۴۰۰ آدرس URL فیشینگ و غیر فیشینگ تهیه شده است. نتایج آزمایش را روی ۱۱۰۸۶ نمونه آموزشی با ساختارهای شبکه عصبی تابع پایه شعاعی بدست آمده است. نسخه های مختلف الگوریتم شبکه عصبی با روش های مختلف انتخاب ویژگی بهترین عملکرد مورد بررسی قرار گرفت و بهترین عملکرد به شبکه عصبی تابع پایه شعاعی رگرسیون کلی اختصاص یافت.

در این پژوهش، یک روش جدید و موثر برای تشخیص وب سایت‌های فیشینگ و حفاظت از کاربران با عنوان تکنیک های هوشمند تشخیص الگو ارائه گردید.

پیشنهادات کارهای آتی

روش پیشنهادی تکنیک های هوشمند تشخیص الگو علاوه بر شناسایی URL برای شناسایی ایمیل های هرزنامه استفاده کنیم از یادگیری بر خط وب کاوی و متن کاوی جهت کاربرد های بلادرنگ قابل استفاده است. اگرچه روش ضد فیشینگ پیشنهاد شده در این پژوهش از کارآیی بالایی برخوردار است.