رویکردهای هوشمند در شناسایی وبگاه‌های دام‌چینی

یگانه ستاری [۱]، غلامعلی منتظر [۲]

[۱] دانشجوی کارشناسی ارشد مهندسی فنّاوری اطلاعات، دانشگاه تربیت مدرس، تهران، ایران
y.sattari@modares.ac.ir
[۲] استاد مهندسی فنّاوری اطلاعات، دانشگاه تربیت مدرس، تهران، ایران
montazer@modares.ac.ir

چکیده

«دام‌چینی» نوعی حمله سایبری است که اغلب از طریق ایجاد وبگاهی جعلی با طراحی کاملاً مشابه وبگاهی قانونی و با هدف فریب کاربران اینترنتی به افشای اطلاعات شخصی خود مانند رمز عبور و مشخصات کارت‌های اعتباری صورت می‌گیرد. دام‌چینی، امنیت شبکه را نه تنها برای کاربران بلکه برای سازمان‌ها به خطر انداخته و امروزه با گسترش اینترنت و فناوری‌های ارتباطی، به یکی از جدّی‌ترین تهدیدهای امنیتی در فضای سایبری تبدیل شده است. تاکنون محققان روش‌های گوناگونی را برای شناسایی دام‌چینی ارائه کرده‌اند، با این حال چنین حمله‌هایی هنوز وجود دارند. در سال‌های اخیر روش‌های هوشمند با سازگاری بالا در مواجهه با حمله‌های جدید، مورد توجه محققان زیادی قرار گرفته است.بنابراین در این مقاله، به منظور بررسی شکاف‌های غالب و جهت‌گیری تحقیقات آینده، به تحلیل و بررسی رویکردهای هوشمند یادگیری ماشینی در شناسایی وبگاه‌های دام‌چینی پرداخته شده است. از جمله چالش‌های موجود، می‌توان به نحوه استخراج و انتخاب ویژگی و همچنین زمان پاسخگویی اشاره کرد.

کلمات کلیدی: امنیت سایبری، امنیت شبکه، دام‌چینی، هوشمند، یاد گیری ماشینی.

۱ مقدمه

«دام‌چینی [۱]» در لغت به معنای «طعمه‌گذاری مهاجمان برای به دام‌انداختن کاربران» است و به نوعی حملهٔ سایبری اشاره دارد که در آن مهاجمان با بهره‌گیری از مهندسی اجتماعی [۲] و ترفندهای فنی [۳]، کاربران اینترنتی را به افشای اطلاعات محرمانه و شخصی خود (مانند نام کاربری، رمز عبور و مشخصات کارت‌های اعتباری) ترغیب می‌کنند [۱، ۲]. دام‌چینی نزدیک به سه دههٔ اخیر تهدیدی امنیتی در فضای سایبری بوده و امروزه با گسترش فناوری و ارتباط صدها میلیون کاربر در سطوح شخصی یا تجاری تعداد این نوع حمله‌ها افزایش

[1] phishing
[2] social engineering
[3] technical tricks

یافته‌اند؛ به طوری که بیش از نیمی از جرایم سایبری را در بر می‌گیرد [۳، ۴، ۵]. در اغلب این حمله‌ها، مهاجمان با نام سازمان‌های معتبر ظاهر می‌شوند و با طراحی وبگاهی جعلی کاملاً مشابه وبگاهی قانونی و معتبر، از طریق کانال‌های ارتباطی مختلفی (مانند رایانامه[۴]) کاربران را به بازدید از این وبگاه ترغیب می‌کنند [۶]. در سال ۲۰۲۱ حدود سه میلیون وبگاه دام‌چینی منحصر به فرد در سراسر جهان کشف شده که نسبت به سال ۲۰۲۰ دو برابر افزایش داشته است. علاوه بر این در سه ماه اول سال ۲۰۲۲ تعداد این وبگاه‌ها به بیش از یک میلیون رسیده است [۷]. اکثر کاربران درک درستی از تهدید دام‌چینی ندارند و نمی‌دانند چنین حمله‌هایی چگونه اجرا می‌شوند [۸]. بنابراین نیاز به ساز و کاری مؤثر برای محافظت از کاربران در برابر حمله‌های دام‌چینی است. روش‌های شناسایی وبگاه‌های دام‌چینی را می‌توان به دو دسته‌ی اصلی تقسیم‌بندی کرد [۴]: دستهٔ اول راه حل‌های ترویجی – آموزشی با هدف افزایش سطح آگاهی کاربران نهایی و دستهٔ دوم راه‌حل‌های فنی با هدف تشخیص خودکار وبگاه‌های دام‌چینی از طریق توسعهٔ روش‌های مبتنی بر نرم‌افزار.

در این مقاله، با تمرکز بر راه‌حل‌های مبتنی بر نرم‌افزار، به دنبال بررسی روش‌های جدید و چالش‌های آنها هستیم. در سال‌های اخیر با توسعهٔ روش‌های یادگیری ماشینی، محققان زیادی به ارائهٔ رویکردهای هوشمند برای شناسایی وبگاه‌های دام‌چینی روی آوردند. بنابراین به تحلیل و مقایسهٔ برخی از آنها بر اساس نوع الگوریتم یادگیری ماشینی مورد استفاده در سه دسته شامل الگوریتم‌های معمولی[۵]، یادگیری ژرف[۶] و منطق فازی[۷] می‌پردازیم. در ادامه ساختار مقاله بدین شرح است: در بخش ۲ مروری مختصر بر روش‌های دام‌چینی و در بخش ۳ بررسی انواع روش‌های شناسایی وبگاه‌های دام‌چینی و تحلیل رویکردهای هوشمند یادگیری ماشینی ارائه شده است. در نهایت در بخش ۴ نتیجه‌گیری شده است.

۲ روش‌های دام‌چینی

مهاجمان پس از طراحی و ایجاد یک وبگاه جعلی کاملاً مشابه وبگاهی قانونی، به روش‌های مستقیم و یا غیر مستقیم کاربر/کاربران مورد هدف را به سمت وبگاه هدایت می‌کنند. در روش‌های مستقیم، مهاجم از طریق کانال‌های ارتباطی مختلفی مانند رایانامه و شبکه‌های اجتماعی با کاربر ارتباط برقرار کرده و با بهره‌گیری از مهندسی اجتماعی، سعی دارد خواننده را متقاعد کند تا هرچه سریع‌تر از پیوندی که برای او ارسال کرده بازدید و اطلاعات خود را به‌روزرسانی یا تأیید کند. اغلب حمله‌های دام‌چینی با ارسال انبوه رایانامه آغاز می‌شوند [۹]. طبق گزارشی که وبگاه آوانان[۸] در سال ۲۰۲۱ در مورد حمله‌های سایبری منتشر کرد، ۵ درصد از تمام رایانامه‌ها، رایانامه‌های دام‌چینی هستند [۱۰]. علاوه بر این مهاجمان رویکردهای فنی مختلفی را برای هدایت غیر مستقیم کاربران به سمت وبگاه به کار می‌گیرند؛ برای مثال ترفندهایی برای افزایش رتبهٔ پیوند

[4] email
[5] conventional
[6] deep learning
[7] fuzzy logic
[8] Avanan

در نتایج موتور جستجو اعمال می‌کنند زیرا کاربرانی که برای یافتن وبگاه ارائه‌دهندهٔ خدمات یا محصول مورد نظر از موتورهای جستجو استفاده می‌کنند، ممکن است با بی‌دقتی روی پیوند دام‌چینی در نتایج جستجو کلیک کنند [۱۱]. علاوه‌بر این گاهی مهاجمان یوآرال[9] وبگاه معتبر را به گونه‌ای جعل می‌کنند تا اگر کاربر خطای نوشتاری مرتکب شود و یا اسم وبگاهی را که شنیده اشتباه وارد کند، به وبگاه دام‌چینی هدایت شود [۱۱، ۱۲].

۳ روش‌های شناسایی دام‌چینی

در سال‌های اخیر مقاله‌های مختلفی در حوزهٔ دام‌چینی و بررسی روش‌های شناسایی آن‌ها توسط محققان منتشر شده است [۱۳، ۵، ۱۴، ۲]. به عنوان نمونه، بَسیت و همکاران، مروری بر حمله‌های دام‌چینی و روش‌های هوشمند در شناسایی آنها ارائه کردند. آنها انواع حمله‌های دام‌چینی و همچنین راه‌حل‌های متقابل را در چهار دستهٔ «یادگیری ماشینی»، «یادگیری ژرف»، «مبتنی بر سناریو» و «ترکیبی» به همراه رویکردهای محققان بررسی کرده و در نهایت چالش‌های موجود را ارائه کردند [۱۳]. با این حال همواره نیاز به مطالعاتی است که رویکردهای ارائه‌شده در شناسایی دام‌چینی را به منظور بهبود آنها و ارائهٔ چالش‌های موجود بررسی کنند. در ادامه، روش‌های شناسایی وبگاه‌های دام‌چینی را در دو دستهٔ روش‌های «مبتنی بر مقایسه[10]» و روش‌های «هوشمند» بررسی می‌کنیم.

۱.۳ روش‌های مبتنی بر مقایسه

به طور کلی در روش‌های مبتنی بر مقایسه، اجزای اصلی وبگاه مشکوک (مورد بررسی) با وبگاه‌های معتبر مقایسه و سپس نوع آن تعیین می‌شود. این روش‌ها را می‌توان به سه دسته تقسیم کرد: رویکردهای «مبتنی بر فهرست[11]»، «اکتشافی[12]»، «مبتنی بر شباهت بصری[13]»

الف. مبتنی بر فهرست: این رویکردها با مقایسهٔ وبگاه‌ها با موارد موجود در فهرست‌های سیاه/سفید که حاوی اطلاعات وبگاه‌های جعلی/قانونی ازپیش‌شناسایی‌شده مانند یوآرال، نشانی آی‌پی، نام دامنه و غیره هستند، به تعیین وضعیت وبگاه‌ها می‌پردازند [۱۵]. این رویکردها سرعت عمل بالایی دارند و معمولاً به سادگی طراحی و پیاده‌سازی می‌شوند. با این حال اکثر این نوع رویکردها وبگاه‌های ساعت-صفر[14] (موارد جدید یا ازپیش‌دیده‌نشده) را شناسایی نمی‌کنند و همچنین فهرست‌ها باید به طور مکرر به‌روز شوند که این امر نیاز به تأیید و مداخلهٔ انسانی دارد [۱۶، ۱۷].

[9] URL
[10] comparison-based
[11] list-based
[12] heuristics
[13] visual-based
[14] zero-hour

ب. اکتشافی: در رویکردهای اکتشافی ویژگی‌های مختلفی مبتنی بر نشانی و یا محتوا از وبگاه استخراج شده و در صورتی که با وبگاه‌های معتبر و قانونی مطابقت بالایی داشته باشند، به عنوان دام‌چینی شناسایی می‌شوند. این رویکردها می‌توانند حمله‌های ساعت – صفر را شناسایی کنند. با این حال اکثر آنها «میزان مثبت نادرست[۱۵]» بالایی دارند که ممکن است منجر به از دست رفتن اعتماد کاربران به سیستم شود [۱۶].

ج. مبتنی بر شباهت بصری: این رویکردها وبگاه‌های دام‌چینی را از میزان شباهت ظاهری با وبگاه‌های قانونی شناسایی کرده و قادر به شناسایی حمله‌های جدید هستند. با این حال برای نگهداری تصاویر وبگاه‌ها و پردازش آنها به فضای ذخیره‌سازی بزرگ و زمان قابل‌توجهی نیاز دارند [۱۶].

۲.۳ روش‌های هوشمند

شناسایی وبگاه‌های دام‌چینی را می‌توان به عنوان یک مسئلهٔ دسته‌بندی[۱۶] در نظرگرفت که شامل دو دستهٔ «دام‌چینی (فریب)» و «قانونی» است. در روش‌های هوشمند با اعمال الگوریتم‌های یادگیری ماشینی از طریق استدلال الگوها و ویژگی‌های وبگاه‌های برچسب‌گذاری‌شده (که پیشتر به عنوان دام‌چینی یا قانونی شناسایی شده‌اند)، مدل‌هایی ساخته می‌شود که وبگاه‌های جدید و فاقد برچسب را دسته‌بندی می‌کنند. روش‌های هوشمند توانمندی بیشتری در شناسایی انواع جدید حمله‌ها دارند [۱۸]. از مهم‌ترین معیارهای ارزیابی در این رویکردها می‌توان به صحت[۱۷] (مجموع نمونه‌های دام‌چینی و قانونی که درست شناسایی شدند، نسبت به تمام نمونه‌های موجود) و زمان‌پاسخگویی[۱۸] (زمان بین ورود یوآرال به سامانه تا مشخص شدن نوع وبگاه) اشاره کرد. تاکنون محققان رویکردهای مختلفی را برای شناسایی وبگاه‌های دام‌چینی ارائه کرده‌اند که در ادامه برخی از آنها را به تفکیک نوع الگوریتم یادگیری ماشینی مورد استفاده، در سه دسته بررسی می‌کنیم: الگوریتم‌های معمولی، یادگیری ژرف و منطق فازی.

الف. الگوریتم‌های معمولی: در ایجاد مدل‌های مبتنی بر الگوریتم‌های معمولی یادگیری ماشینی، فرایند استخراج و انتخاب ویژگی نیازمند مهارت انسانی است و مجزا از عمل دسته‌بندی انجام می‌شود؛ به طوری که نمی‌توان آنها را برای بهبود عملکرد مدل در یک فاز ترکیب کرد [۱۷]. برخی از الگوریتم‌های معمولی پرکاربرد در شناسایی وبگاه‌های دام‌چینی عبارتند از [۲، ۱۴، ۱۷]: بیز ساده[۱۹]، نزدیک‌ترین k همسایه[۲۰]، ماشین بردار پشتیبان[۲۱] و جنگل تصادفی[۲۲]. به عنوان نمونه، گوپتا و همکاران ۹ ویژگی لغوی را مبتنی بر یوآرال، استخراج و عملکرد چهار دسته‌بند را روی مجموعه داده‌ای شامل ۱۱.۹۶۴

[15] False Positive Rate (FPR)
[16] classification
[17] accuracy
[18] response time
[19] Naïve Bayes
[20] K-Nearest Neighbor (KNN)
[21] Support Vector Machine (SVM)
[22] random forest

یوآرال دام‌چینی و قانونی، ارزیابی کردند که در نهایت صحت این مدل با الگوریتم جنگل تصادفی ۵۷/۹۹ درصد و زمان پاسخگویی ۵۱ میلی‌ثانیه برآورد شده است [۶].

ب. یادگیری ژرف: یادگیری ژرف نشان‌دهندهٔ نوعی از یادگیری ماشینی است که در آن از شبکه‌های عصبی ژرف برای آموزش و شناسایی وبگاه‌ها استفاده می‌شود. از جمله ویژگی‌هایی که این روش را از الگوریتم‌های معمولی مجزا می‌کند، توانایی یادگیری و انطباق با داده، استخراج خودکار ویژگی‌ها از داده‌های خام (مانند یوآرال) و کشف همبستگی پنهان بین آنها است. با این حال این روش‌ها نسبت به الگوریتم‌های معمولی تفسیرپذیری کمتری دارند و اغلب توضیح منطق پشت مفروضات، تصمیم‌گیری‌ها و نتیجه‌گیری‌هایی که یک شبکه عصبی انجام می‌دهد، ممکن نیست. علاوه بر این نسبت به الگوریتم‌های معمولی برای آموزش مدل به زمان و همچنین نمونه‌های آموزشی بیشتری نیاز دارند که جمع‌آوری آنها ممکن است بسیار پرهزینه و زمان‌بر باشد [۱۷]. الگوریتم‌های مختلفی در یادگیری ژرف وجود دارد که از پرکاربردترین آنها در شناسایی دام‌چینی می‌توان به شبکه عصبی پیچشی[۲۳]، شبکه عصبی بازگشتی[۲۴]، واحد بازگشتی دروازه‌ای[۲۵] و حافظهٔ کوتاه-مدت طولانی[۲۶] اشاره کرد. به عنوان نمونه، در تحقیقی مدل مبتنی بر حافظهٔ کوتاه-مدت طولانی برای شناسایی یوآرال‌های دام‌چینی پیشنهاد شده است که ابتدا رشته یوآرال را با استفاده از روش وان-هات[۲۷] رمزگذاری و سپس هر بردار رمزگذاری‌شده را برای آموزش و آزمایش در نرون‌های شبکه وارد می‌کند. این مدل بر روی مجموعه داده‌ای شامل یک میلیون یوآرال دام‌چینی و یک میلیون یوآرال قانونی ارزیابی شده که میزان صحت ۷/۹۸ درصد و زمان پاسخگویی ۲۸۱ ثانیه به ازای هر یوآرال برآورد شده است [۱۹].

ج. منطق فازی: منطق فازی، مدل‌سازی دقیق شیوه‌های استدلال تقریبی است که استدلال و تصمیم‌گیری را در محیطی غیر قطعی و غیر دقیق با اطلاعات ناقص (مانند گفتگوی انسان‌ها با زبان طبیعی) ممکن می‌سازد [۲۰]. سامانه‌های مبتنی بر منطق فازی قدرت یادگیری ندارند و شامل قواعد «اگر-آنگاه» هستند که توسط خبرگان هر حوزه استخراج می‌شود که عینی نیستند [۲۱].

به‌عنوان نمونه، رویکردی مبتنی بر منطق فازی پیشنهاد شده است که ابتدا ۶ ویژگی عددی را از مجموعه داده استخراج و سپس آنها را به مقادیر زبانی که میزان تأثیر هر ویژگی را با واژه‌های «زیاد»، «متوسط» و «کم» نشان می‌دهد، تبدیل می‌کند. در نهایت با اعمال مجموعه قواعد اگر-آنگاه که توسط الگوریتم تکامل تفاضلی[۲۸] بهینه شده‌اند، وبگاه‌ها دسته‌بندی می‌شوند. صحت روی مجموعه داده‌ای شامل ۲۰.۰۰۰ وبگاه، ۶/۹۷ درصد برآورد شده است [۲۲].

[23] Convolutional Neural Networks (CNN)
[24] Recurrent Neural Networks (RNN)
[25] Gated Recurrent Unit (GRU)
[26] Long short-term memory (LSTM)
[27] one-hot
[28] Differential Evaluation (DE)

جدول ۱: مقایسهٔ رویکردهای هوشمند یادگیری ماشینی در شناسایی وبگاه‌های دام‌چینی

مدل یا الگوریتم	نوع	مجموعه داده	مزیت‌ها	محدودیت‌ها	صحت (%)
جنگل تصادفی [6]	معمولی	۱۹.۹۶۴ نمونه: ۹.۹۶۴ یوآرال دام‌چینی و ۱۰.۰۰۰ یوآرال قانونی با ۹ ویژگی متنی مبتنی بر یوآرال	– صحت بالا – زمان پاسخگویی کوتاه (۵۱ میلی ثانیه) – استخراج ویژگی مستقل از خدمات شخص ثالث	– عدم اعتبارسنجی و ارزیابی استحکام مدل با مجموعه داده‌های مختلف – محدود بودن ویژگی‌ها	۹۹/۵۷
جنگل تصادفی [23]	معمولی	۴.۰۵۹ وبگاه: ۲.۱۴۱ نشانی دام‌چینی و ۱.۹۱۸ نشانی قانونی ۱۹ ویژگی مبتنی بر نشانی و محتوای صفحه	–مقایسهٔ پنج الگوریتم مختلف –زمان پاسخگویی قابل قبول (۵.۸۰۰ میلی ثانیه)	–مجموعه داده کوچک – برخی از ویژگی‌ها مبتنی بر مقایسه با وبگاه‌های معتبر	۹۹/۰۹
جنگل تصادفی [4]	معمولی	۷۳.۵۷۵ نمونه: ۳۷.۱۷۵ یوآرال دام‌چینی و ۳۶.۴۰۰ یوآرال قانونی با ۴۰ ویژگی	– استخراج ۲۷ ویژگی به روش پردازش زبان طبیعی – مستقل از زبان وبگاه	– استخراج یک ویژگی مبتنی بر رتبهٔ صفحه از پایگاه اَلکسا (از دسترس خارج شده)	۹۷/۹۸
ماشین بردار پشتیبان [24]	معمولی	۲.۰۰۰ نمونه: ۱.۰۰۰ یوآرال دام‌چینی و ۱.۰۰۰یوآرال قانونی با ۶ ویژگی	– استخراج ویژگی مبتنی بر یوآرال	– مجموعه‌داده کوچک – نیاز به یوآرال وبگاه قانونی در استخراج یک ویژگی – محدود بودن ویژگی‌ها	۹۵/۸۰
ماشین بردار پشتیبان [25]	معمولی	۲.۱۳۴ نمونه: ۱.۴۴۸ یوآرال دام‌چینی و ۶۸۶ یوآرال قانونی با ۱۷ ویژگی	– صحت بالا – زمان پاسخگویی قابل‌قبول (حداکثر ۶.۳۰۰ میلی‌ثانیه)	– مجموعه داده کوچک – برخی از ویژگی‌ها مبتنی بر کد صفحه (HTML)	۹۸/۶۵
حافظهٔ کوتاه- مدت طولانی [19]	یادگیری ژرف	۲ میلیون نمونه: ۱ میلیون یوآرال دام‌چینی و ۱ میلیون یوآرال قانونی	– مجموعه‌داده بزرگ – صحت بالا – بدون فرایند استخراج ویژگی و فقط با دریافت یوآرال – مستقل از خدمات شخص ثالث	– زمان پاسخگویی زیاد (۲۸۱ ثانیه) – زمان آموزش زیاد (۲۳۸ دقیقه)	۹۸/۷
حافظهٔ کوتاه- مدت طولانی + شبکه عصبی پیچشی [26]	یادگیری ژرف	۱ میلیون نمونه: ۱ میلیون یوآرال دام‌چینی و ۱۰.۰۰۰ تصویر با ۳۵ ویژگی	– مجموعه‌داده بزرگ – ترکیب دو شبکه عصبی – ویژگی‌های ترکیبی مبتنی بر متن، قالب و تصویر وبگاه	– صحت پایین و زمان پاسخگویی زیاد (۲۵ ثانیه) – برخی از ویژگی‌ها وابسته به خدمات شخص ثالث و یا مبتنی بر کد صفحه (جاوااسکریپت)	۹۳/۲۸

شبکه عصبی بازگشتی + شبکه عصبی پیچشی [15]	یادگیری ژرف	۴۹۰.۴۰۸ نمونه: ۲۴۵.۳۸۵ یوآرال دام‌چینی و ۲۴۵.۰۲۳ یوآرال قانونی	– مجموعه داده بزرگ – ارزیابی استحکام مدل – بدون فرایند استخراج ویژگی و فقط با دریافت یوآرال – ترکیب دو شبکه عصبی	– صحت پایین – زمان پاسخگویی زیاد (۴۰ ثانیه) – حداکثر طول یوآرال ۲۵۵ کاراکتر	۹۵/۷۹
شبکه عصبی پیچشی [27]	یادگیری ژرف	۳۱۸.۶۴۲ نمونه: ۱۵۷.۶۲۶ یوآرال دام‌چینی و ۱۶۱.۰۱۶ یوآرال قانونی با ویژگی‌های سطح کاراکتر مبتنی بر یوآرال	–استفاده از چهار مجموعه داده مختلف برای ارزیابی –استخراج و مقایسهٔ چهار گروه مختلف از ویژگی‌ها	–حداکثر طول نشانی وبگاه ۲۰۰ کاراکتر –زمان آموزش طولانی (۸۹ دقیقه)	۹۵/۰۲
شبکه عصبی بازگشتی + واحدبازگشتی دروازه‌ای [28]	یادگیری ژرف	۱۲۰.۰۰۰ نمونه: ۶۰.۰۰۰ یوآرال دام‌چینی و ۶۰.۰۰۰ یوآرال قانونی	– میزان صحت بالا – استخراج ویژگی با پردازش زبان طبیعی – پیاده‌سازی به عنوان یک افزونهٔ مرورگر	– ذخیره یوآرال تا ۲۰۰ کاراکتر – عدم پشتیبانی از یوآرال‌های کوتاه	۹۹/۱۸
سامانهٔ فازی [22]	منطق فازی	۲۰.۰۰۰ نمونه: ۱۰.۰۰۰ یوآرال دام‌چینی و ۱۰.۰۰۰ یوآرال قانونی با ۶ ویژگی	– بهینه‌سازی مجموعه قواعد با الگوریتم تکامل تفاضلی	– استخراج یک ویژگی مبتنی بر رتبهٔ صفحه از پایگاه اَلکسا (از دسترس خارج شده)	۹۷/۶
انفیس [29]	منطق فازی	۱۳.۰۰۰ نمونه: ۴.۸۹۸ یوآرال دام‌چینی، ۱.۹۴۵ یوآرال مشکوک و ۶.۱۵۷ یوآرال قانونی با ۳۵ ویژگی	– صحت بالا – ویژگی‌های ترکیبی مبتنی بر متن، قالب و تصویر وبگاه	– برخی از ویژگی‌ها وابسته به خدمات شخص ثالث، نیازمند بارگیری کامل صفحه و یا مبتنی بر کد صفحه فقط به زبان جاوااسکریپت	۹۸/۵۵

مقایسه‌ای از صحت به‌دست‌آمده در این رویکردها به همراه مزیت‌ها و محدودیت‌های آنها در جدول ۱ گردآوری شده است که نشان می‌دهد عملکرد آن‌ها نه‌تنها به نوع الگوریتم یادگیری ماشینی، بلکه به نوع ویژگی‌ها و نحوهٔ استخراج آنها وابسته است. ویژگی‌های مبتنی بر متن و یا کد صفحه در صورتی که وابسته به زبان خاصی باشند، در مواجهه با وبگاه‌هایی که به زبان دیگر نوشته شده‌اند، کارایی ندارند. ویژگی‌هایی که استخراج آن‌ها وابسته به خدمات شخص ثالث است، ممکن است منجر به افزایش زمان پاسخگویی و در برخی موارد، ناپایداری سامانه شوند؛ به عنوان مثال رویکردهایی که ویژگی‌های مبتنی بر رتبهٔ صفحه را از پایگاهی مانند اَلکسا[۲۹] استخراج می‌کردند، با خاتمه‌یافتن فعالیت این پایگاه در تاریخ ۱ مه ۲۰۲۲، کارایی خود را از دست داده‌اند. علاوه بر این ویژگی‌های متنی مبتنی بر یوآرال به‌رغم اینکه زمان کمی برای استخراج

[29] alexa

نیاز دارند و مستقل از خدمات شخص ثالث هستند، ممکن است تمام مؤلفه‌های وبگاه‌های دام‌چینی را به خوبی پوشش ندهند و از طرفی اگر مهاجم از روش‌های کوتاه‌کننده‌ی پیوند (مانند بیتلی [۳۰]) استفاده کند، این ویژگی‌ها کارایی ندارند.

۴ نتیجه‌گیری

در این مقاله به طور مختصر انواع روش‌های هوشمند در شناسایی وبگاه‌های دام‌چینی، مورد بررسی قرار گرفت. این رویکردها دارای مزایای زیادی از جمله سازگاری با حمله‌های جدید و همچنین صحت بیش از ۹۸ درصدی در برخی موارد بوده‌اند؛ با این حال هنوز مسئلهٔ شناسایی وبگاه‌های دام‌چینی با چالش‌هایی جدّی روبه‌روست که از مهم‌ترین آن‌ها می‌توان به انتخاب ویژگی‌های کارآمد و ارائهٔ مدلی قوی به منظور بررسی آنها و پاسخگویی سریع در مورد وبگاه‌های جدید اشاره کرد.

مراجع

[1] R. M. Mohammad, F. Thabtah, and L. McCluskey, “Tutorial and critical analysis of phishing websites methods,” Computer Science Review, vol. 17, 2015, pp. 1-24.

[2] L. Tang, and Q. H. Mahmoud, “A Survey of Machine Learning-Based Solutions for Phishing Website Detection,” Machine Learning and Knowledge Extraction, vol. 3, 2021, pp. 672-694.

[3] A. Shaikh, A. Shabut, and A. Hossain, “A literature review on phishing crime, prevention review and investigation of gaps,” In 10th International Conference on Software, Knowledge, Information Management & Applications (SKIMA), 2016, pp. 9-15.

[4] O. K. Sahingoz, E. Buber, O. Demir, and B. Diri, “Machine learning based phishing detection from URLs,” Expert Systems with Applications, vol. 117, 2019, pp. 345-357.

[5] A. K. Jain and B. B. Gupta, “A survey of phishing attack techniques, defence mechanisms and open research challenges,” Enterprise Information Systems, 2021, pp. 1-39.

[6] B. B. Gupta, K. Yadav, I. Razzak, K. Psannis, A. Castiglione, and X. Chang, “A novel approach for phishing URLs detection using lexical based machine learning in a real-time environment,” Computer Communications, vol. 175, 2021, pp. 47-57.

[7] APWG. Phishing Activity Trends Report. 1th quarter 2019: 1th quarter 2022. available at https://apwg.org/trendsreports/ Last acceced on 4 Mar, 2022.

[8] M. Wu, R. C. Miller, and S. L. Garfinkel, “Do security toolbars actually prevent phishing attacks?,” In Proceedings of the SIGCHI conference on Human Factors in computing systems, 2006, pp. 601-610.

[9] B. B. Gupta, N. A. G. Arachchilage, and K. E. Psannis, “Defending against phishing attacks: taxonomy of methods, current issues and future directions,” Telecommunication System, vol. 67, 2018, pp. 247-267.

[30] bitly

[10] Avanan. 1H Cyber Attack Report. 2021. available at https://www.avanan.com/resources/white-papers/1h-cyber-attack-report Last acceced 26 Aug, 2022.

[11] R. Alabdan, "Phishing Attacks Survey:Types,Vectors, and Technical Approaches," Future Internet, vol. 12, 2020, pp. 1-37.

[12] K. L. Chiew, K. S. C. Yong, and C. L. Tan, "A survey of phishing attacks: Their types, vectors and technical approaches," Expert Systems with Applications, vol. 106, 2018, pp. 1-20.

[13] A. Basit, M. Zafar, X. Liu, A. R. Javed, Z. Jalil, and K. Kifayat, "A comprehensive survey of AI-enabled phishing attacks detection techniques," Telecommunication Systems, vol. 76, 2021, pp. 139-154.

[14] P. Kalaharsha and B.M. Mehtrea, "Detecting Phishing Sites - An Overview," 2021, arXiv, arXiv:2103.12739.

[15] W. Wang, F. Zhang, X. Luo, and S. Zhang, "Pdrcnn: precise phishing detection with recurrent convolutional neural networks," Security and Communication Networks, 2019.

[16] M. Khonji, Y. Iraqi, and A. Jones, "Phishing detection: a literature survey," IEEE Communications Surveys & Tutorials, vol. 15, 2013, pp. 2091-2121.

[17] N. Q. Do, A. Selamat, O. Krejcar, E. Herrera-Viedma, and H. Fujita, "Deep Learning for Phishing Detection: Taxonomy, Current Challenges and Future Directions," IEEE Access, 2022.

[18] E. Buber, Ö. Demir, and O. K. Sahingoz, "Feature selections for the machine learning based detection of phishing websites," International Artificial Intelligence and Data Processing Symposium (IDAP), 2017, pp. 1-5.

[19] A. C. Bahnsen, E. C. Bohorquez, S. Villegas, J. Vargas, and F. A. González, "Classifying phishing URLs using recurrent neural networks," In 2017 APWG symposium on electronic crime research (eCrime), IEEE, 2017, pp. 1-8.

[20] L. A. Zadeh, "Fuzzy logic," Computer, vol. 21, 1988, pp. 83-93.

[21] G. A. Montazer and S. ArabYarmohammadi, "Detection of phishing attacks in Iranian e-banking using a fuzzy–rough hybrid system," Applied Soft Computing, vol. 35, 2015, pp. 482-492.

[22] R. M. Abdul-Hussein, A. H. Mohammed, and A. A. Kadhim, "Detecting Phishing Cyber Attack Based on Fuzzy Rules and Differential Evaluation," TEM Journal, vol. 11, 2022, pp. 543-551.

[23] A. K. Jain and B. B. Gupta, "Towards detection of phishing websites on client-side using machine learning based approach," Telecommunication Systems, vol. 68, 2018, pp. 687-700.

[24] M. Zouina and B. Outtaj, "A novel lightweight URL phishing detection system using SVM and similarity index," Human-centric Computing and Information Sciences, vol. 7, 2017, pp. 1-13.

[25] M. Moghimi and A. Y. Varjani, "New rule-based phishing detection method," Expert systems with applications, vol. 53, 2016, pp. 231-242.

[26] M. A. Adebowale, K. T. Lwin, and M. A. Hossain, "Intelligent phishing detection scheme using deep learning algorithms," Journal of Enterprise Information Management, 2020.

[27] A. Aljofey, Q. Jiang, Q. Qu, M. Huang, and J. Niyigena, "An Effectiv Phishing Detection Model Based on Character Level Convolutional Neural Network from URL," Electronics, vol. 9, 2020.

[28] L. Tang, and Q. H. Mahmoud, "A Deep Learning-Based Framework for Phishing Website Detection," IEEE Access, vol. 10, 2021, pp. 1509-1521.

[29] M. A. Adebowale, K. T. Lwin, E. Sanchez, and M. A. Hossain, "Intelligent web-phishing detection and protection scheme using integrated features of Images, frames and text," Expert Systems with Applications, vol. 115, 2019, pp. 300-313.