فصلنامهٔ پژوهش‌های اطلاعاتی و جنایی، سال یازدهم
شمارهٔ اول، بهار ۱۳۹۵

ارزیابی روش‌های شناسایی وب‌سایت فیشینگ

تاریخ دریافت: ۱۳۹۴/۰۴/۱۵ تاریخ پذیرش: ۱۳۹۴/۰۸/۲۲

از صفحهٔ ۹ تا ۳۸

سلمان کمالی زاده[۱]، غلامرضا شاه‌محمدی[۲]

چکیده

فیشینگ[۳] یکی از تکنیک‌های مهندسی اجتماعی برای فریب کاربران است که به معنای تلاش برای به دست آوردن اطلاعات محرمانه مانند نام کاربری، گذرواژه یا اطلاعات حساب بانکی است. امروزه از مهم‌ترین چالش‌های موجود در اینترنت، خطر حملات فیشینگ و کلاه‌برداری‌های اینترنتی است. این حملات تنها در آمریکا، سالیانه چندین میلیارد دلار خسارت به بار می‌آورد. از این‌رو، پژوهشگران تلاش‌های زیادی در جهت شناسایی و مقابله با این گونه حملات داشته‌اند. هدف این تحقیق، ارزیابی روش‌های شناسایی وب‌سایت فیشینگ است. این تحقیق از نظر هدف کاربردی و از نظر ماهیت از نوع توصیفی - تحلیلی است. در این مقاله، ضمن معرفی حمله فیشینگ و روش‌های موجود، شناسایی وب‌سایت فیشینگ، بر اساس مطالعات انجام شده و تجارب محققان با پیشنهاد معیارهایی، روش‌های شناسایی وب‌سایت فیشینگ مورد ارزیابی قرار می‌گیرند. نتایج به دست آماده حاکی از آن است روش‌هایی که از تکنیک‌های مختلف شناسایی در کنار هم استفاده می‌کنند و همچنین اکثر ویژگی‌های صفحات وب را بررسی می‌کنند، در شناسایی حمله از موفقیت بیشتری برخوردار می‌باشند.

کلیدواژه‌ها

فیشینگ، شناسایی حمله فیشینگ، ارزیابی روش‌های شناسایی فیشینگ.

---

۱. عضو هیئت علمی دانشگاه جامع علمی کاربردی هرمزگان: skamalizadeh@gmail.com

۲. استادیار دانشگاه علوم انتظامی امین، نویسندهٔ مسئول: shah_mohammadi@yahoo.co.uk

3.Phishing

مقدمه

فیشینگ یک نوع حملهٔ رایانه ای است که مهاجم از طریق کانال‌های ارتباط الکترونیکی[۱]، با ایجاد ارتباط با انسان‌ها و با استفاده از پیام‌های مهندسی اجتماعی[۲]، به ترغیب آن‌ها، برای انجام کارهایی که به نفع مهاجم است، اقدام می‌کند(خونجی، عراقی و جونز، ۲۰۱۳، ص ۲).

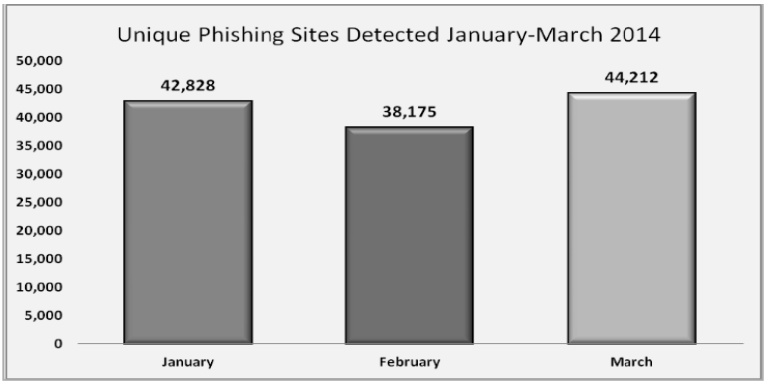
امروزه از مهم‌ترین چالش‌های موجود در بانکداری اینترنتی، خطر حملات فیشینگ و کلاه‌برداری‌های اینترنتی است و خسارت‌های فراوانی به مشتریان و به سازمان‌ها وارد می‌کند. علاوه بر پول و زمان از دست رفته؛ اعتماد افراد به برنامه‌ها و خدمات برخط[۳] از دست می‌رود و درنتیجه، اعتبار سازمان‌ها کم‌رنگ می‌شود. بر اساس گزارش گرتنر[۴](۲۰۰۶)، این حملات تنها در آمریکا سالیانه چندین میلیارد دلار خسارت به بار می‌آورد(دانم، ۲۰۰۹، ص ۱۳۱).

از این‌رو، متخصصان 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 آنجایی که فیشرها، همواره روش کار خود را با هزینهٔ اندک تغییر می‌دهند، این ابزارها نیازمند به‌روز شدن روش‌های شناسایی خود هستند.

به گزارش سیمانتک[۵](۲۰۱۴)، نرخ حملات فیشینگ در سال ۲۰۱۳ افزایش یافته است، به گونه‌ایی که این نرخ در سال ۲۰۱۲ از هر ۴۱۴ ایمیل، یک ایمیل فیشینگ بوده است، درحالی‌که در سال ۲۰۱۳ به ازای هر ۳۹۲ ایمیل یک ایمیل فیشینگ گزارش شده است. همچنین نرخ حملات فیشینگ در ماه مه میلادی سال ۲۰۱۴ به ازای هر ۳۹۵ ایمیل یک ایمیل فیشینگ گزارش شده است.

شکل (۱) آمار اعلام‌شده توسط گروه کاری ضدفیشینگ[۶](۲۰۱۴) از وب‌سایت‌های فیشینگ شناسایی شده در اولین فصل سال ۲۰۱۴ را ارائه می‌دهد.

1.Electronic Communication Channels
2. Social engineering
3.Online Services
4.Gartner
5.Symantec
6.APWG



شکل ۱ – گزارش گروه کاری ضد فیشینگ از سایت‌های فیشینگ شناسایی‌شده از ژانویه تا مارس ۲۰۱۴

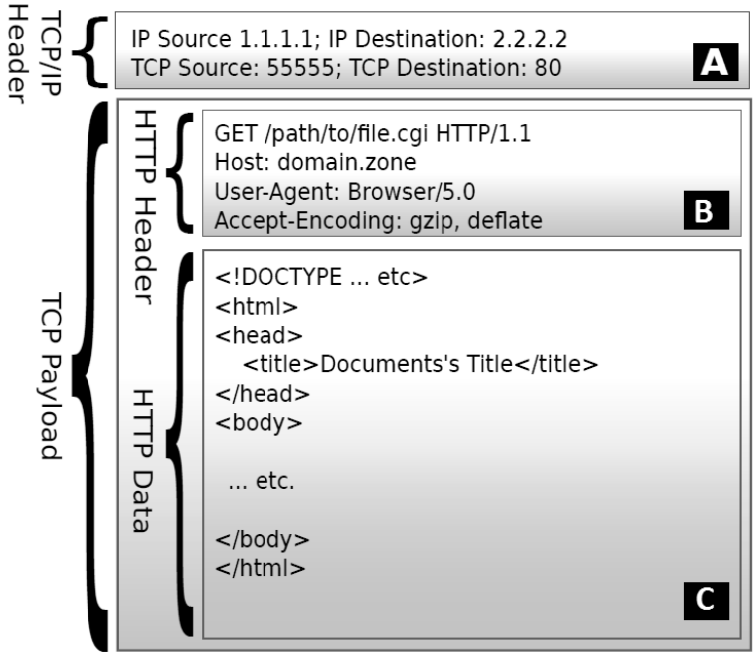
سرعت رشد این حملات به‌گونه‌ای است که نیاز به شناسایی و مقابله با آن‌ها روزبه‌روز بیشتر احساس می‌شود. امروزه باوجود طراحی ابزارهای فراوان برای کشف و مقابله با این نوع حملات و همچنین با گسترش آگاهی‌های عمومی دربارهٔ فیشینگ، این حملات هنوز هم خطری جدی در اینترنت محسوب می‌شوند و روزبه‌روز تعداد آن‌ها رو به افزایش است.

این پژوهش به دنبال ارزیابی روش‌هایی است که برای شناسایی حمله فیشینگ از طریق وب‌سایت جعلی ارائه شده‌اند. روش‌های شناسایی وب‌سایت فیشینگ، روی یک یا چند قسمت دادهٔ صفحهٔ وب، کار تجزیه و تحلیل را انجام می‌دهند که شامل موارد زیر هستند:

- **قسمت A :** سرآیند[1] TCP/IP[2] که حداقل شامل IP مبدأ و مقصد، منبع TCP و همچنین مقصد TCP است.
- **قسمت B :** سرآیند HTTP که حداقل شامل آدرس دقیق صفحهٔ وب، میزبان[3](دامنه[4] و منطقه[5]) است.
- **قسمت C :** HTTP Data شامل داده‌های صفحه وب است که در قالب HTML[6] ترجمه شده است.

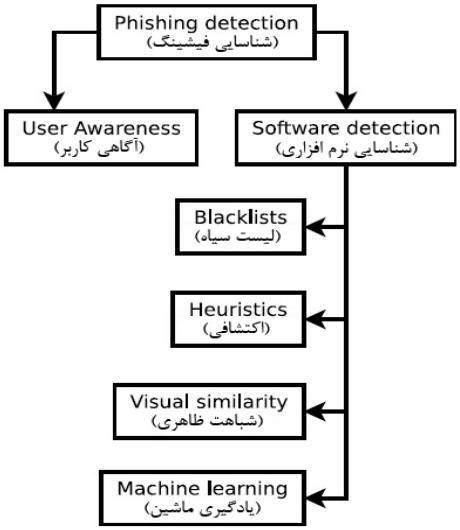
1.Header
2.Transmission-Control Protocol/Internet Protocol
3.Host
4.Domain
5.Zone
6.HyperText Markup Language

در شکل (۲)، بخش‌های دادهٔ وب‌سایت برای تحلیل، توسط روش‌های شناسایی حملهٔ فیشینگ ارائه شده است.



شکل ۲ – بخش‌های داده وب‌سایت برای تحلیل توسط روش‌های شناسایی
(خونجی، عراقی و جونز، ۲۰۱۳، ص ۲۳)

روش‌های شناسایی وب‌سایت فیشینگ: دسته‌بندی ارائه شده در شکل(۳)، روش‌های شناسایی فیشینگ را به دو دستهٔ آگاهی کاربر و شناسایی نرم‌افزاری تقسیم می‌کند.



شکل ۳ – دسته‌بندی روش‌های شناسایی فیشینگ (خونجی، عراقی و جونز، ۲۰۱۳، ص ۵)

مطابق شکل (۳)، روش‌های نرم افزاری برای شناسایی وب‌سایت فیشینگ به چهار دستهٔ زیر تقسیم می‌شوند:

دستهٔ اول: روش‌های مبتنی بر لیست سیاه

لیست سیاه شامل فهرستی از URLها، IPها و کلیدواژه‌های فیشینگ است که اخیراً پیدا شده‌اند و غیرقانونی و کلاهبردارانه هستند. برخلاف لیست سیاه، لیست سفید شامل فهرستی از موارد قانونی و مطمئن است که نرخ خطای شناسایی را کاهش می‌دهد. طبق بررسی‌های انجام‌شده توسط شنگ و همکارانش[1](۲۰۰۹)، لیست‌های سیاه در کشف حملات فیشینگ لحظهٔ صفر بی‌فایده یا کم فایده هستند، به گونه‌ایی که تنها ۲۰ درصد از آن‌ها را کشف می‌کنند. این بررسی‌ها نشان می‌دهد ۴۷ درصد تا ۸۳ درصد از URLهایی که در لیست‌های سیاه هستند، بعد از ۱۲ ساعت قرارگرفته‌اند. این در حالی است که ۶۳ درصد از حملات فیشینگ در حدود ۲ ساعت به پایان می‌رسند.

۱- Google Safe Browsing API: شرکت گوگل[2] لیست سیاهی شامل goog-malware-shavar برای بدافزارها و goog-phish-shavar برای حملات فیشینگ به‌صورت مستمر به‌روز می‌کند. Google Safe Browsing API این قابلیت را در اختیار

1.Sheng
2.Google

نرم‌افزارهای ایستگاه‌های کاری قرار می‌دهد که URL به آن بدهند و بررسی کنند که در لیست سیاه گوگل وجود دارد یا خیر (توسعه دهندگان گوگل، ۲۰۱۶). این پروتکل هنوز هم مورد آزمایش و استفاده قرار می‌گیرد و بخشی از مرورگرهای وب فایرفاکس[۱] و کروم[۲] است(مرور امن گوگل، ۲۰۱۶).

۲- فیش نت[۳]: هر تغییری در URL فیشینگ، نتیجه عدم برابری را به دنبال دارد. کار پراکاش[۴] و همکاران(۲۰۱۰) با عنوان فیش نت، محدودیت آدرس‌های کاملاً برابر را شناسایی می‌کند. برای این کار، URL های لیست سیاه با عنوان والد[۵] را پردازش می‌کند، سپس شکل‌های مختلف URLهای شبیه به آن را با عنوان فرزندان[۶] ایجاد می‌کند. پنج روش اکتشافی جهت انجام این کار در زیر لیست شده است:

- جایگزینی TLD[۷]: هر URL می‌تواند با تغییر TLD به ۳۲۱۰ شکل مختلف انشعاب پیدا کند.

- شباهت ساختار فهرست[۸]: URLهای فیشینگ دارای ساختار فهرست شبیه به هم با تفاوت‌های جزئی هستند که می‌توان URLهای فرزندان شبیه به ساختار فهرست را ایجاد کرد. به‌عنوان‌مثال:

–http://www.abc.com/online/ebay.html.
–and http://www.xyz.com/online.paypal.com

می‌توان به URL های فرزندان زیر انشعاب داد:

–http://www.xyz.com/online/ebay.html.

–and http://www.abc.com/online.paypal.com

- هم ارزی[۹] آدرس IP: URLهایی با ساختار فهرست مشابه اما با نام‌های دامنهٔ متفاوت، در صورتی که آدرس IP یکسان داشته باشند، می‌توانند برابر در نظر گرفته شوند.

- جانشینی رشته پرسش[۱۰]: مانند «شباهت ساختار فهرست»، می‌توان شکل‌های مختلف

1.FireFox
2.Chrome
3.PhishNet
4.Prakash
5.Parent
6.Childeren
7.Replace Top Level Domains
8.Directory
9.Equivalence
10.Query

یک URL را با رشته‌های پرسش متفاوت انشعاب داد. به‌عنوان مثال:

–http://www.abc.com/online/ebay.php?ABC.

–and http://www.abc.com/online.paypal.com?XYZ.

می‌توان به URLهای فرزندان زیر انشعاب داد:

–http://www.abc.com/online/ebay.php?XYZ.

–and http://www.abc.com/online.paypal.com?ABC.

- هم‌ارزی نام برند[1]: تولید URLهای فرزندان با URL مشابه اما با نام‌های برند متفاوت. مانند تبدیل http://www.abc.com/online/paypal.html به http://www.abc.com/online/ebay.html به عنوان یک URL فرزند. این روش، بخش‌های داده **A** و **B** صفحه را مورد تجزیه و تحلیل قرار می‌دهد. همچنین، برخی از نوار ابزارهایی که مبتنی بر لیست سیاه ساخته شده‌اند، شامل موارد زیر هستند:

- کلودمارک[2](۲۰۱۶)
- تراست واچ[3](۲۰۱۶)
- راهنمای سایت[4](۲۰۱۶)
- نت‌کرفت[5](۲۰۱۶)

دستهٔ دوم: روش‌های اکتشافی

الگوریتم‌ها با هر مکانیسمی برای شناسایی یا جلوگیری از حملات فیشینگ مورد استفاده قرار می‌گیرند. این الگوریتم‌ها می‌توانند به صورت اکتشافی عمل کنند. اکتشاف فیشینگ دارای مشخصاتی است که در حملات فیشینگ واقعی یافت شده است. گرچه ضمانتی نیست که این مشخصات در همهٔ حملات وجود داشته باشد، اما این دسته از روش‌های شناسایی برخلاف لیست سیاه(که در اغلب موارد نیازمند این است که حمله عیناً برابر با موارد ثبت شده در لیست خود باشد)، برای شناسایی حملات لحظهٔ صفر(که اخیراً دیده نشده‌اند) بسیار مفید و مؤثر خواهند بود.

1.Brand
2.Cloudmark
3.TrustWatch
4.SiteAdvisor
5.NetCraft

باوجود اینکه این روش‌ها، ریسک دسته‌بندی وب‌سایت‌های قانونی(به عنوان وب‌سایت فیشینگ) را به همراه دارند، در حال حاضر بسیاری از مرورگرها مانند فایرفاکس و مرورگر اینترنت[1] با مکانیسم‌های محافظت از فیشینگ ساخته شده‌اند که اغلب از آزمون‌های اکتشافی جهت شناسایی کمک می‌گیرند (خونجی، عراقی و جونز، ۲۰۱۳، ص ۱۳).

۱- اسپوف گارد[2]: روش‌چو[3] و همکاران(۲۰۰۴) با عنوان اسپوف گارد، یک افزونه[4] ساخته شده توسط دانشگاه استانفورد[5] که فعالیت‌های فیشینگ مبتنی بر HTTP(S) را با عنوان یک نوار ابزار مرورگر وب شناسایی می‌کند. این کار با وزن‌دهی به موارد نامتعارف شناسایی شده در محتوی HTML صورت می‌گیرد. سه مثال زیر توسط اسپوف گارد قابل شناسایی است:

- صفت href مربوط به تگ Anchor، شبیه به یک URL در لیست سفید باشد. مانند www.gmail.com در مثال زیر:

Click Here

- صفت href مربوط به تگ Anchor، شامل یک URL پنهان شده باشد. به عنوان مثال:

Click Here

- متن مربوط به تگ Anchor، یک URL متفاوت با URL موجود در تگ href باشد. مانند:

www.gmail.com

این روش، بخش‌های داده **B** و **C** صفحه را مورد تجزیه و تحلیل قرار می‌دهد.

۲- تشخیص نفوذ مشترک یا CID[6]: بسیاری از روش‌های شناسایی و مکانیسم‌های جلوگیری، مبتنی بر شناسایی آدرس IP منبع مهاجم هستند. در طرف دیگر، مهاجم برای تغییر پی‌درپی آدرس IP منبع، از شبکه‌های متغیر پی‌درپی استفاده می‌کند. روش کار ژو[7] و همکاران(۲۰۰۸) با وجود اینکه هنوز به شکل قابل استفاده پیاده‌سازی نشده، نشان می‌دهد که می‌توان با تجزیه و تحلیل تمامی بخش‌های دادهٔ صفحهٔ وب (**A**، **B** و

1.Internet Explorer
2.SpoofGuard
3.Chou
4.Plug-in
5.Standford
6.Collaborative Intrusion Detection
7.Zhou

C) و همچنین استفاده از این روش در سیستم‌های تشخیص نفوذ[1] و زیر نظر گرفتن تغییرات پی‌درپی آدرس IP، صفحات فیشینگ را شناسایی کرد.

۳- فیش گارد[2]: روش کار لیکاریش[3] و همکاران(۲۰۰۸)، با عنوان فیش گارد، در قالب یک افزونهٔ مرورگر وب فایرفاکس، گام‌های زیر را برای شناسایی دنبال می‌کند:

- کاربر یک صفحه را مشاهده می‌کند.
- اگر صفحهٔ مشاهده شده، یک درخواست اصالت سنجی[4] را ارسال کرد و همچنین کاربر فرم مربوطه را پر کرد، فیش گارد کار خود را شروع می‌کند.
- فیش گارد نام کاربری یکسان به همراه رمز عبور تصادفی (که با رمز عبور واقعی برابر نباشد) n بار (به صفحهٔ مشاهده شده) ارسال می‌کند.
- اگر صفحهٔ پاسخ داده شده همراه با پیام‌های HTTP 200 OK باشد، این بدان معناست که کاربر با یک سایت فیشینگ روبرو است که تصدیق‌های جعلی را با پیام‌های موفقیت‌آمیز پاسخ می‌دهد.
- اگر صفحهٔ پاسخ داده شده همراه با پیام‌های HTTP 401 Unauthorized باشد، آنگاه امکان دارد یکی از موارد زیر باشد:

الف) یک سایت فیشینگ که به شکل کورکورانه و همیشگی پیام‌های تصدیق عدم موفقیت را پاسخ می‌دهد.

ب) این یک سایت قانونی است.

- جهت تمایز بین دو احتمال بالا، فیش گارد اعتبارنامه واقعی را یک بار دیگر ارسال می‌کند.
- اگر صفحه پاسخ داده شده همراه با پیام‌های HTTP 200 OK باشد آنگاه به این نتیجه می‌رسد که سایت قانونی است و امتحان خود را پس داده است.
- اگر صفحه پاسخ داده شده همراه با پیام‌های HTTP 401 Unauthorized باشد، آنگاه دو احتمال زیر وجود دارد:

الف) یک سایت فیشینگ که به شکل کورکورانه و همیشگی پیام‌های تصدیق عدم

1.IDS (Intrusion Detection System)
2.PhishGuard
3.Likarish
4.Authentication

موفقیت را پاسخ می‌دهد. در این مورد، اعتبارنامهٔ ورود کاربر برای فیشر ارسال‌شده است. این بدان معناست که این تکنیک تنها قادر به شناسایی زیرمجموعه‌ایی از سایت‌های فیشینگ است.

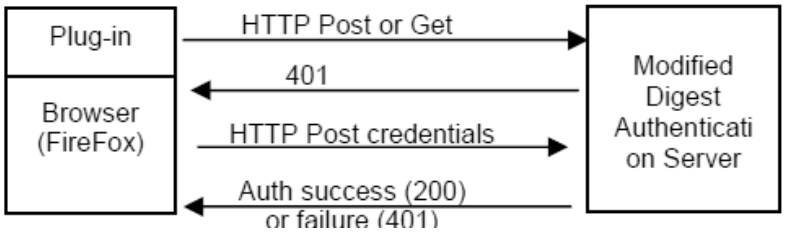
ب) کاربر، رمزعبور را اشتباه وارد کرده است.

- برای اطمینان از اینکه رمز عبور توسط کاربر اشتباه وارد نشده، فیش گارد رمزهای عبور درهم سازی شده[1] را در یک فایل ذخیره می‌کند و برای درخواست‌های ورود در آینده به کار می‌گیرد:

الف) اگر رمز عبور واردشده با یکی از رمزهای عبور درون فایل برابر باشد، فیش گارد نتیجه می‌گیرد که رمز عبور درست است و این یک سایت فیشینگ است.

ب) اگر مورد برابر پیدا نشد، آنگاه فیش گارد نتیجه می‌گیرد که رمز عبور اشتباه بوده و به کاربر پیام خطا را نمایش می‌دهد.

این روش با استفاده از آزمون‌های اکتشافی، قادر به شناسایی حملات لحظهٔ صفر است و همچنین بخش‌های دادهٔ **B** و **C** صفحه را مورد تجزیه و تحلیل قرار می‌دهد.



شکل ۴ – روش کار فیش گارد (لیکاریش، دان‌بار و هان‌سن، ۲۰۰۸، ص ۲)

۴- کانتینا[2]: کانتینا (ژانگ، هنگ و کرانور، ۲۰۰۷) یک نوار ابزار مرورگر وب اینترنت[3] است که برای تصمیم‌گیری در مورد اینکه صفحهٔ مشاهده شده یک صفحهٔ فیشینگ است، محتوای آن را تجزیه و تحلیل می‌کند. کانتینا از الگوریتم TF-IDF[4] برای بازیابی و بررسی اطلاعات توسط سالتن[5] و همکاران (۱۹۸۶) طراحی شده است و همچنین از موتورهای جستجو و برخی اکتشافات برای کاهش خطای مثبت استفاده می‌کند. گام‌های

1.Hashed
2.Cantina
3.Internet Explorer
4.Term Frequency/Inverse Document Frequency
5.Salton

زیر برای شناسایی وب‌سایت‌های فیشینگ توسط کانتینا انجام می‌گیرد:

- TF-IDF هر اصطلاح (عبارت یا واژه) در یک صفحهٔ وب مشکوک محاسبه می‌شود.
- پنج اصطلاح با بالاترین مقدار TF-IDF در سند گرفته می‌شوند. فلپس[1] و همکاران(۲۰۰۰)، آن را امضای لغوی[2] نامیده‌اند.
- پنج اصطلاح انتخاب شده به موتور جستجو ارسال می‌شوند. برای مثال پرسش جستجو در گوگل برابر است با http://www.google.ae/search?q=t1,t2,t3,t4,t5

سپس نام‌های دامنهٔ ابتدای نتیجهٔ پرسش(n دامنه) بازگشت داده می‌شود.

- اگر دامنهٔ مشکوک در بین نتایج بازگشت داده باشد، این سایت قانونی است.

همچنین برای کاهش خطای مثبت، برخی اکتشفات زیر انجام می‌شود:

- سن دامنه: اگر بیشتر از ۱۲ ماه باشد، احتمالاً یک صفحهٔ قانونی است.
- URL صفحهٔ مشکوک: اگر حاوی کاراکترهای @ و – باشد، صفحهٔ فیشینگ است.
- پیوندهای مشکوک در محتوای صفحه: اگر حاوی کاراکترهای @ و – باشد، صفحهٔ فیشینگ است.
- تعداد نقطه‌ها در URL: اگر بیش از ۵ نقطه باشد، صفحهٔ فیشینگ است.
- فرم‌ها: اگر حاوی فرم‌های HTML باشد، احتمالاً صفحهٔ فیشینگ است.
- TF-IDF: گرچه شناسایی صفحه فیشینگ مبتنی بر دسته‌بندی خود کانتینا است، اما آن‌هم مبتنی بر TF-IDF است.

این روش، بخش‌های دادهٔ **B** و **C** صفحه را مورد تجزیه و تحلیل قرار می‌دهد.

۵- مولد لیست سیاه: شریفی و همکاران(۲۰۰۸)، مکانیسمی جهت تولید لیست سیاه با استفاده از موتورهای جستجو مانند گوگل ارائه داده‌اند. هدف کار آن‌ها، شناسایی وب‌سایت‌های فیشینگ و سپس ثبت آن‌ها در پایگاه داده است. مجموعهٔ دادهٔ ارزیابی شامل ۵۰۰ وب‌سایت قانونی تصادفی است که از نتایج جستجوی کلیدواژه‌های مختلف تصادفی در گوگل تشکیل شده است و همچنین ۳۰ وب‌سایت فیشینگ که توسط PIRT[3] طبقه‌بندی شده‌اند.

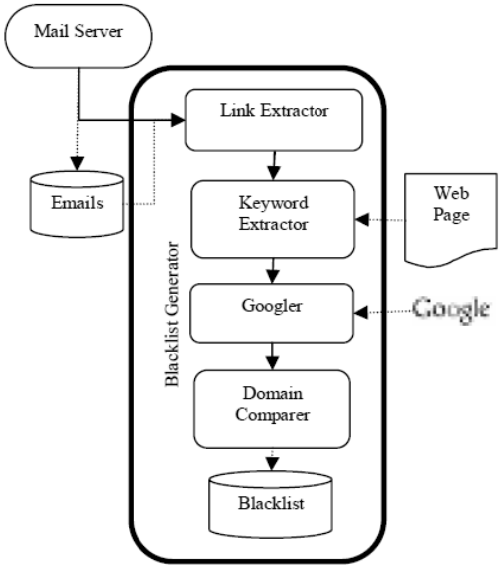
1.Phelps
2.Lexical Signature
3.Phishing Incident Reporting and Termination

الگوریتم اکتشافی آن‌ها، شامل مراحل زیر است:

- به دست آوردن نام شرکت از URL مشکوک؛
- جستجوی نام شرکت در گوگل و برگرداندن ۱۰ نتیجهٔ اول؛
- اگر URL مشکوک شامل ۱۰ نتیجهٔ اول گوگل باشد، این صفحه قانونی است.
- اگر URL مشکوک شامل ۱۰ نتیجهٔ اول گوگل نباشد، این صفحه مشکوک به فیشینگ است.
- اگر URL مشکوک به‌عنوان فیشینگ شناخته شد، در پایگاه داده ذخیره می‌شود.

زمانی که آزمایش اکتشافی بالا روی وب‌سایت‌های قانونی اجرا شد، تنها ۴۵ وب‌سایت از ۵۰۰ وب‌سایت به اشتباه، به‌عنوان وب‌سایت فیشینگ دسته‌بندی شدند. همچنین زمانی که روی وب‌سایت‌های فیشینگ اعمال شد، تمامی آن‌ها به عنوان وب‌سایت فیشینگ دسته‌بندی شدند.

این روش بخش‌های دادهٔ **B** و **C** صفحه را مورد تجزیه و تحلیل قرار می‌دهد.



شکل ۵ – معماری مولد لیست سیاه (شریفی و سیادتی، ۲۰۰۸، ص ۳)

۶- شناسایی صفحهٔ فیشینگ و کشف هدف: روش کار سینگ و همکاران(۲۰۱۴) شامل الگوریتم شناسایی صفحهٔ فیشینگ و الگوریتم کشف هدف است. الگوریتم شناسایی گام‌های زیر را دنبال می‌کند:

- جعبهٔ متن: اگر URL حاوی یک یا چند جعبهٔ متن باشد، الگوریتم کشف هدف اقدام

کند، در غیر این صورت گام ۲.

- سن دامنه: اگر سن دامنه از ۶ ماه کمتر باشد، الگوریتم کشف هدف اقدام کند، در غیر این صورت گام ۳.

- Google safe browsing: اگر جواب این گام منفی بود، الگوریتم کشف هدف اقدام کند، در غیر این صورت گام۴.

- اطلاعات Whois: اگر اطلاعات Who is مورد قبول بود، گام ۵. در غیر این صورت الگوریتم کشف هدف اقدام کند.

- طول URL: اگر طول URL<54 آنگاه گام ۶، در غیر این صورت الگوریتم کشف هدف اقدام کند.

- نقطه‌های چندگانه: اگر تعداد نقطه‌ها بیش از ۳ بود، الگوریتم کشف هدف اقدام کند.

- منبع URL صفحهٔ وب مشخص باشد: گام ۸، در غیر این صورت الگوریتم کشف هدف اقدام کند.

- URL معتبر است.

الگوریتم کشف هدف با استفاده از Google API شامل سه گام زیر است:

- تولید کلید و قرار دادن آن در جاوا اسکریپت[1]؛

- اضافه کردن محتوای HTML برای جستجوی وب؛

- فراخوانی Google Search API توسط جاوا اسکریپت؛

این روش با استفاده از آزمون‌های اکتشافی، قادر به شناسایی حملات لحظهٔ صفر است و همچنین بخش‌های دادهٔ **B** و **C** صفحه را مورد تجزیه و تحلیل قرار می‌دهد.

۷- اکتشاف مبتنی بر URL: روش ارائه‌شده توسط نویین[2] و همکاران(۲۰۱۴) مبتنی بر ویژگی‌های URL بوده و به طور خاص روی ظاهر آن‌ها تمرکز دارد. فازهای این روش شامل موارد زیر است:

– انتخاب ویژگی‌های URL: در این فاز، چهار ویژگی از URL استخراج می‌شود که شامل (دامنه، دامنهٔ اصلی، زیر دامنه و مسیر دامنه) است.

1.JavaScript
2.Nguyen

– محاسبهٔ شش مقدار اکتشاف:

- محاسبه مقدار اکتشاف **دامنهٔ اصلی**: الگـوریتم مـوردنظر در شـکل (۶) آورده شـده است:

```
Input: PrimaryDomain
Output: Value of heuristic "PrimaryDomain"

If PrimaryDomain is IP then value= 0; // doubt phishing
If PrimaryDomain is not IP then
        result = Suggestion_Google(PrimaryDomain);
        If result is null then
            value = 1; // no doubt phishing
        End if
        If result is not null then
            value=min(1, Levenshtein(result, PrimaryDomain)/4);
        End if
End if
```

شکل ۶ – محاسبه مقدار اکتشاف دامنهٔ اصلی (نویین و همکاران، ۲۰۱۴، ص ۳)

- محاسبهٔ مقادیر اکتشاف **زیر دامنه** و **مسیر دامنه** مانند **دامنـــه اصـــلی**: الگـوریتم موردنظر مطابق شکل(۷) است:

```
Input: m     // m is SubDomain or PathDomain
Output: Value of heuristic "m"

If SubDomain is null then value=1; // no doubt phishing
If SubDomain is not null then
    result = Suggestion_Google(m);
    If result is null then
        value = 1; // no doubt phishing
    End if
    If result is not null then
        value=min(1, Levenshtein(result, PrimaryDomain)/4);
    End if
End if
```

شکل ۷ – محاسبهٔ مقادیر اکتشاف زیر دامنه و مسیر دامنه (نویین و همکاران، ۲۰۱۴، ص ۳)

- محاسبهٔ مقدار اکتشافی رتبهٔ صفحه[1]: الگوریتم موردنظر مطابق شکل(۸) است:

```
Input: Domain
Output: Value of heuristic "PageRank"

value = Google_PageRank(Domain);
If value < 0 then
    value = 0 ; // phishing site
else
    value = value/10;
End if
```

شکل ۸ – محاسبهٔ مقدار اکتشافی رتبهٔ صفحه (نویین و همکاران، ۲۰۱۴، ص ۳)

- محاسبهٔ مقدار اکتشافی رتبهٔ آلکس[2]: الگوریتم موردنظر مطابق شکل(۹) است:

```
Input: Domain
Output: Value of heuristic "AlexaRank"

value = 1- min(1,AlexaRank(Domain)/3000000);
```

شکل ۹ محاسبه مقدار اکتشافی رتبه آلکس (نویین و همکاران، ۲۰۱۴، ص ۳)

- محاسبهٔ مقدار اکتشافی **AlexaReputation** : الگوریتم موردنظر در شکل(۱۰) آورده شده است:

```
Input: Domain
Output: Value of heuristic "AlexaReputation"

value = min(1,AlexaReputation(Domain)/30);
```

شکل ۱۰ – محاسبهٔ مقدار اکتشافی (نویین و همکاران، ۲۰۱۴، ص ۳)

- محاسبهٔ مقدار سیستم(vs) : مقدار سیستم بر اساس فرمول زیر به دست می‌آید:

$$vs = \sum v_i * w_i$$

که در آن Vi برابر است با مقدار هر اکتشاف و Wi برابر است با وزن هر اکتشاف.

– دسته‌بندی وب‌سایت: در این فاز، جهت تصمیم‌گیری در خصوص فیشینگ بودن وب‌سایت، مقدار vs با مقدار آستانه[3] مقایسه می‌شود.

1.PageRank
2.AlexaRank
3.Threshold

ارزیابی این روش با مجموعهٔ داده هایی شامل ۱۱۶۰۰ سایت فیشینگ و همچنین ۵۰۰۰ سایت قانونی انجام گرفته است. نتیجه نشان می‌دهد که این روش ۹۷/۱۹ درصد از سایت‌های فیشینگ را شناسایی کرده است. این روش، بخش دادهٔ B صفحه را مورد تجزیه و تحلیل قرار می‌دهد.

۸- جلوگیری از حملات فیشینگ با استخراج رتبهٔ صفحه، اعتبار و منبع کد: روش کار پاندا[1] و همکاران(۲۰۱۴) 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 مقایسه می‌کند. ارزیابی کارایی این روش ارائه نشده است اما قابلیت شناسایی لحظه صفر وب‌سایت‌های فیشینگ را دارد و بخش‌های دادهٔ B و C صفحه را مورد تجزیه و تحلیل قرار می‌دهد.

دستهٔ سوم: شباهت ظاهری

این دسته از روش‌های شناسایی وب‌سایت فیشینگ، مبتنی بر شباهت ظاهری هستند و با تجزیه و تحلیل کد منبع صفحه و بررسی اطلاعات سطح شبکه مخالف‌اند. ازجمله این روش‌ها می‌توان به کار هارا[2] و همکاران(۲۰۰۹) با عنوان «شناسایی مبتنی بر شباهت ظاهری بدون داشتن اطلاعات سایت قربانی» اشاره کرد. همچنین روش زیر نیز مبتنی بر شباهت ظاهری است.

مبارزه با فیشینگ با ویژگی‌های کلیدی متمایز: برخلاف سایر مکانیسم‌های ضد فیشینگ، راه‌حل ارائه‌شده توسط چن[3] و همکاران(۲۰۰۹)، جهت شناسایی فیشینگ مبتنی بر محتوای ارائه‌شده به جای کد محتوا است. به‌عبارت‌دیگر، این مکانیزهٔ شناسایی فیشینگ، مخالف زیر نظر قرار دادن کدها و فناوری‌هایی است که درنهایت به خروجی قابل نمایش برای کاربر ترجمه می‌شوند. برای مثال، یک سایت فیشینگ که شبیه یک سایت قانونی است با نمایش محتوای شبیه به آن با استفاده از تگ‌های img زبان HTML ممکن است مکانیسم‌های ضد فیشینگی که شناسایی آن‌ها بر اساس محتوای HTML است را کنار بگذارد. این‌گونه موارد توسط این روش قابل شناسایی است.

1.Panda
2.Hara
3.Chen

این روش نیازمند آن است که مرورگر وب یک عکس فوری[1] از سایت مشکوک بگیرد. سپس، عکس گرفته شده با لیست سفیدی از وب‌سایت‌های محافظت شده که مورد هدف فیشرها هستند(مانند PayPal، Amazon، eBay و سایت‌های بانک‌ها) مقایسه می‌شود. همچنین این روش از الگوریتم هریس - لاپلاس[2] برای شناسایی تصویر استفاده می‌کند تا کار شناسایی خود را انجام دهد. این روش تنها بخش دادهٔ C را بررسی می‌کند.

دستهٔ چهارم: یادگیری ماشین

تکنیک‌هایی شناسایی وب‌سایت فیشینگ که در این دسته فعالیت می‌کنند، به دنبال راه‌حل‌هایی برای دسته‌بندی سند یا مسئله خوشه‌بندی هستند. در آن‌ها، مدل‌هایی با استفاده از یادگیری ماشین و الگوریتم‌های خوشه‌بندی ساخته می‌شود. ازجملهٔ این الگوریتم‌ها، می‌توان به نزدیک‌ترین همسایه(k-NN)، C4.5، ماشین بردار پشتیبانی(SVM)، k-means و DBSCAN اشاره کرد.

از جمله روش‌های ارائه شده در این دسته عبارت‌اند از:

- ضد فیشینگ مبتنی بر متن و ظاهر: یک رویکرد بیزی(ژانگ، لیو و چو، ۲۰۱۱)؛
- دسته‌بندی خودکار مقیاس بزرگ صفحات(ویتکر، رینر و نزیف، ۲۰۱۰)؛
- نوار ابزار ضد فیشینگ بیزی(B-APT[3]) (لیکاریش، دان‌بار و هان‌سن، ۲۰۰۸)؛

همچنین شرح روش لیو[4] و همکاران(۲۰۱۰) که از داده‌کاوی بهره می‌برد در زیر ارائه شده است:

-شناسایی خودکار هدف فیشینگ از صفحه فیشینگ:ذلیو و همکاران(۲۰۱۰) به این مورد اشاره می‌کنند که شناسایی وب‌سایت‌های فیشینگ و همچنین اهداف آن‌ها با پیدا کردن وب‌سایت‌های شبیه به صفحات مشکوک امکان‌پذیر است. اگر یک وب‌سایت مشکوک شبیه وب‌سایتی با نام دامنه متفاوت باشد، آنگاه وب‌سایت مشکوک به عنوان یک وب‌سایت فیشینگ در نظر گرفته می‌شود. برای مثال، اگر وب‌سایت بسیار شبیه به Paypal باشد، مطمئناً یک وب‌سایت فیشینگ با هدف حمله به Paypal است. این روش مبتنی بر داده‌کاوی[5] بوده و از تکنیک دسته‌بندی استفاده می‌کند(از جمله DBSCAN).

1.Snapshot
2.Harris-Laplace
3.Bayesian Anti-Phishing Toolbar
4.Liu
5.Data Mining

مجموعهٔ دادهٔ ارزیابی آن شامل ۸۷۴۵ وب‌سایت فیشینگ (استخراج شده از فیش‌تانک[1]) و همچنین ۱۰۰۰ وب‌سایت قانونی تصادفی از لینک‌های Yahoo! است. این روش، بخش‌های دادهٔ B و C صفحه را مورد تجزیه و تحلیل قرار می‌دهد.

جدول (۱)، خلاصهٔ روش‌های شناسایی وب‌سایت فیشینگ را نشان می‌دهد.

جدول ۱ – خلاصهٔ روش‌های شناسایی حمله فیشینگ

ردیف	عنوان روش	مبتنی بر	بررسی بخش دادهٔ A	بررسی بخش دادهٔ B	بررسی بخش دادهٔ C
۱	فیش نت	لیست سیاه	✓	✓	
۲	اسپوف گارد	اکتشاف		✓	✓
۳	CID		✓	✓	✓
۴	فیش گارد			✓	✓
۵	کانتینا			✓	✓
۶	مولد لیست سیاه			✓	✓
۷	شناسایی صفحه فیشینگ و کشف هدف			✓	✓
۸	اکتشاف مبتنی بر URL			✓	
۹	جلوگیری از حملات فیشینگ با استخراج رتبهٔ صفحه، اعتبار و منبع کد			✓	✓
۱۰	شناسایی مبتنی بر شباهت ظاهری بدون داشتن اطلاعات سایت قربانی	شباهت ظاهری		✓	✓
۱۱	مبارزه با فیشینگ با ویژگی‌های کلیدی متمایز			✓	✓
۱۲	ضد فیشینگ مبتنی بر متن و ظاهر: یک رویکرد بیزی	یادگیری ماشین		✓	✓
۱۳	دسته‌بندی خودکار مقیاس بزرگ صفحات		✓	✓	✓
۱۴	نوار ابزار ضد فیشینگ بیزی(B-APT)			✓	
۱۵	شناسایی خودکار هدف فیشینگ از صفحه فیشینگ			✓	✓

روش‌شناسی تحقیق

این تحقیق از نظر هدف؛ از آنجا که بر اساس مطالعات گستردهٔ انجام شده، به بررسی روش‌های شناسایی حملهٔ فیشینگ و ارزیابی این روش‌ها از ابعاد مختلف مانند دقت و قابلیت به‌کارگیری روش در شناسایی حملهٔ فیشینگ می‌پردازد، از نوع کاربردی و از نظر ماهیت؛ از آنجا که به بررسی و ارزیابی روش‌های حمله فیشینگ می‌پردازد، از نوع تحقیقات توصیفی - تحلیلی است. در این تحقیق، ابتدا روش‌های موجودِ شناسایی وب‌سایت فیشینگ تشریح شد. سپس، بر اساس مطالعات انجام شده و تجارب محققان،

1.PhishTank

با پیشنهاد معیارهایی، روش‌های شناسایی وب‌سایت فیشینگ مورد ارزیابی قرار گرفت. روش جمع‌آوری اطلاعات در این تحقیق، مطالعات کتابخانه‌ای گسترده است.

یافته‌های تحقیق

الف) معیارهای ارزیابی روش‌های شناسایی حمله فیشینگ

بر اساس ارزیابی‌های ارائه شده در پژوهش‌های مختلف در حوزهٔ شناسایی حملهٔ فیشینگ مانند (خونجی، عراقی و جونز، ۲۰۱۳، ص ۲۵)، معیارهای زیر برای ارزیابی روش‌های شناسایی حمله فیشینگ ارائه و پیشنهاد می‌شود. این معیارها در سه دستهٔ زیر قرار می‌گیرند:

دسته اول: ویژگی‌ها

این دسته شامل ویژگی روش‌های شناسایی حملهٔ فیشینگ است که قابلیت‌های آن‌ها را نشان می‌دهد.

۱- شناسایی لحظهٔ صفر[1]: روش موردنظر توانایی شناسایی حملهٔ فیشینگ در لحظهٔ صفر را دارد یا خیر؟ منظور شناسایی آنی وب‌سایت‌هایی است که همان لحظه، توسط فیشر در دسترس قرار داده شده است.

۲- شناسایی فارمینگ[2]: 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 حملهٔ فیشینگ است یا خیر؟ منظور از هدف، می‌تواند همان وب‌سایت قانونی باشد که فیشر آن را جعل کرده است و یا اینکه، وب‌سایت فیشینگ در پی به دست آوردن چه نوع اطلاعاتی(مانند اطلاعات حساب بانکی و یا نام کاربری و رمز عبور یک ایمیل) است؟

۴- شناسایی اشیاء جاسازی شده: جعل برخی از وب‌سایت‌های قانونی توسط فیشرها با استفاده از اشیاء جاسازی شده 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 ساخته شده است یا خیر؟

1.Zero-Day
2.Pharming

۵- متمرکز بر سرویس دهنده: روش شناسایی، در سرویس دهنده متمرکز است یا خیر؟

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

۷- تبدیل به ابزار قابل استفاده: روش شناسایی به ابزاری قابل استفاده توسط کاربر یا یک سرویس دهنده 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 سرویس دهنده یا وب‌سایت مشکوک از جمله این موارد است.

۹- نیازمند ارتباط با کاربر: روش موردنظر برای شناسایی حملهٔ فیشینگ، وابسته به تعامل با کاربر است یا خیر؟ البته بدیهی است نمایش نتیجهٔ یک روش یا اعلام هشدار فیشینگ بودن یک وب‌سایت به کاربر، به عنوان وابستگی مطرح نمی‌شود.

۱۰- وابستگی زبانی: روش موردنظر محدودیت زبان دارد یا خیر؟ به عبارت دیگر آیا روش موردنظر وابسته به زبان خاصی(مانند انگلیسی یا فارسی) هست یا خیر؟

دستهٔ دوم: نتایج آزمون

برای آزمایش یک روش شناسایی حملهٔ فیشینگ، می‌توان از مجموعهٔ داده[1] شامل نمونه‌های قانونی و فیشینگ استفاده کرد که در نهایت چهار دستهٔ احتمالی وجود خواهد داشت که مطابق جدول (۲)، این پراکندگی نمایش داده شده است:

جدول ۲ – پراکندگی نتایج دسته‌بندی روش‌های شناسایی (خونجی، عراقی و جونز، ۲۰۱۳، ص ۶)

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

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

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

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

1.Dataset

$NL \rightarrow L$ تعداد نمونه‌های **قانونی** که به درستی به عنوان **قانونی** دسته‌بندی شده‌اند.

بر این اساس، می‌توان معیارهای زیر را ارائه کرد:

۱- تأخیر[1]: روش شناسایی چقدر زمان برای انجام کار خود نیاز دارد.

۲- خطای مثبت[2] **یا FP**: با اجرای روش شناسایی فیشینگ، روی لیستی 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، نحوهٔ محاسبهٔ خطای مثبت را نشان می‌دهد.

$$FP = \frac{N_{L \rightarrow P}}{N_{L \rightarrow L} + N_{L \rightarrow P}}$$

۳- خطای منفی[3] **یا FN**: با اجرای روش شناسایی فیشینگ، روی فهرستی از وب‌سایت‌های فیشینگ، درصد خطاهایی که در شناسایی این‌گونه وب‌سایت‌ها به عنوان وب‌سایت قانونی شناخته‌شده، خطای منفی نامیده می‌شود(به عبارت دیگر سایت فیشینگ را به عنوان سایت قانونی تشخیص می‌دهد). فرمول زیر، نحوهٔ محاسبه خطای منفی را نشان می‌دهد.

$$FN = \frac{N_{P \rightarrow L}}{N_{P \rightarrow P} + N_{P \rightarrow L}}$$

دسته سوم: تکنیک‌های مورد استفاده

این دسته شامل تکنیک‌هایی است که روش‌های شناسایی حملهٔ فیشینگ استفاده می‌کنند که شامل موارد زیر می‌شوند:

۱- لیست سیاه[4]: 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.

1.Delay
2.False Positives
3.False Negatives
4.Blacklists

۲- لیست سفید[1]: 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[2]: 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 به دنبال رسیدن به نزدیک‌ترین جواب صحیح هستند.

۴- شباهت ظاهری[3]: 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 تکنیک‌های داده‌کاوی برای کار خود بهره می‌برد یا خیر؟

ب) ارزیابی روش‌های شناسایی حمله فیشینگ

ارزیابی روش‌های شناسایی حمله فیشینگ را می‌توان با اعتبار سنجی و مقایسه با سایر روش‌ها انجام داد.

برای اعتبار سنجی، می‌توان روش موردنظر را با دو مجموعه آزمود تا میزان خطای مثبت و منفی آن محاسبه شود. یک مجموعه شامل تعدادی صفحات قانونی است و مجموعهٔ دیگر شامل مجموعه‌ای از صفحات فیشینگ است. از نمونه‌های قانونی برای محاسبهٔ میزان خطای مثبت روش و از نمونه‌های فیشینگ برای محاسبهٔ میزان خطای منفی روش استفاده می‌شود.

برخی از مجموعه داده‌های موجود شامل موارد زیر هستند:

- فیش‌تانک[4](۲۰۱۶)
- گروه‌کاری ضدفیشینگ[5](۲۰۱۶)
- فهرست وب[6](۲۰۱۶)
- API مرور امن گوگل[1](۲۰۱۶)

1.Whitelists
2.Heuristics
3.Visual Similarity
4.PhishTank
5.APWG
6.DMOZ: The Directory of the Web

- راهنمای سایت مکافی[2] (۲۰۱۶)
- Iran.ir

۱- مقایسه با سایر روش‌ها: *با بهره‌گیری از مقایسه (خونجی، عراقی و جونز، ۲۰۱۳، ص ۲۵)، ارزیابی انجام‌شده بر اساس معیارهای معرفی شده، در جدول (۳) نمایش داده شده است.*

در این جدول، در مقابل نام روش، معیارهای ارزیابی جهت مقایسه ارائه شده است. در صورتی که روش موردنظر از تکنیک‌های لیست سیاه، لیست سفید، اکتشافی، شباهت ظاهری و یادگیری ماشین استفاده کند و همچنین، ویژگی‌های مختلف از جمله شناسایی لحظهٔ صفر، شناسایی فارمینگ، شناسایی هدف، شناسایی اشیاء جاسازی شده، 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 ✓ مشخص شده است. در ستون‌های خطای مثبت و خطای منفی نیز نتیجه آزمون هرکدام ارائه شده است.

جدول ۳ – خلاصهٔ مقایسهٔ روش‌های شناسایی وب‌سایت فیشینگ

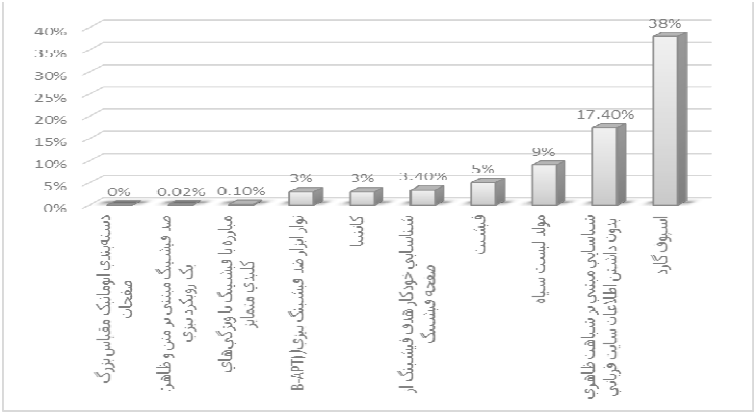
ردیف	روش شناسایی	تکنیک مورد استفاده: لیست سیاه	لیست سفید	اکتشافی	شباهت ظاهری	یادگیری ماشین	ویژگی‌ها: شناسایی لحظهٔ صفر	شناسایی فارمینگ	شناسایی هدف	شناسایی اشیاء جاسازی شده	متمرکز بر سرویس‌دهنده	متمرکز بر ایستگاه کاری	تبدیل به ابزار جهت استفاده	عدم نیاز به ارتباطات شبکه	عدم نیاز به ارتباط با کاربر	عدم وابستگی زبانی	نتیجه آزمون: خطای مثبت	خطای منفی
۱	فیش نت	✓		✓								✓	✓	✓	✓	✓	٪۵	٪۳
۲	اسپوف گارد			✓			✓					✓	✓	✓	✓		٪۳۸	٪۹
۳	CID			✓				✓			✓				✓	✓	-	-
۴	فیش گارد			✓			✓					✓	✓	✓	✓		-	-

1.GoogleSafeBrowsing
2.McAfee Site Advisor

ادامهٔ جدول ۳ – خلاصهٔ مقایسهٔ روش‌های شناسایی وب‌سایت فیشینگ

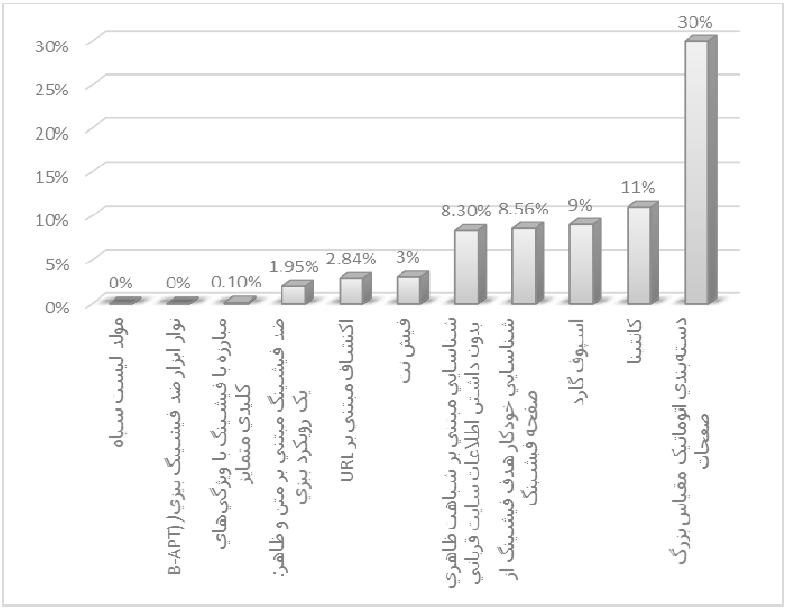
۵	کانتینا			✓			✓					✓	✓		✓		٪۳	٪۱۱
۶	مولد لیست سیاه			✓			✓					✓	✓		✓	✓	٪۹	٪۰
۷	شناسایی صفحهٔ فیشینگ و کشف هدف			✓			✓		✓			✓			✓		-	-
۸	اکتشاف مبتنی بر URL			✓			✓				✓	✓			✓	✓	-	٪۲.۸۴
۹	جلوگیری از حملات فیشینگ با استخراج رتبهٔ صفحه، اعتبار و منبع کد			✓			✓					✓			✓		-	-
۱۰	شناسایی مبتنی بر شباهت 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	✓	✓	✓		✓	✓					✓			✓	✓	٪۰.۰۲-۰	٪۱.۹۵-۰
۱۳	دسته‌بندی خودکار مقیاس بزرگ صفحات	✓	✓	✓		✓					✓	✓	✓		✓		٪۰	٪۳۰-۱۶
۱۴	نوار ابزار ضد فیشینگ بیزی (B-APT)		✓			✓	✓						✓		✓		٪۳	٪۰
۱۵	شناسایی خودکار هدف فیشینگ از صفحهٔفیشینگ				✓	✓	✓		✓			✓			✓		٪۳.۴	٪۸.۵۶

بر اساس مقایسهٔ انجام شده در جدول(۳)، مقایسهٔ روش‌ها بر اساس نـرخ خطـای مثبـت آن‌ها در نمودار(۱) ارائه شده است:



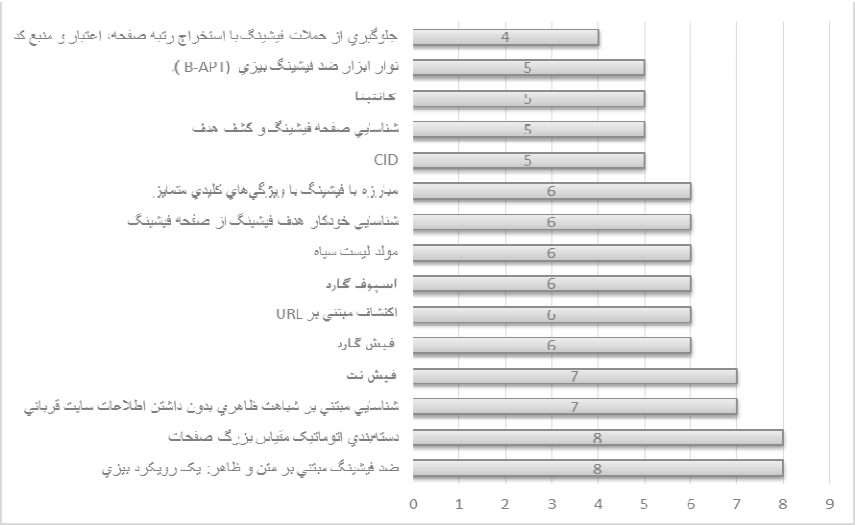
نمودار ۱ – مقایسه روش‌ها بر اساس نرخ خطای مثبت

مطابق جدول(۳)، مقایسه روش‌ها بر اساس نرخ خطای منفی آن‌هـا در نمـودار(۲) ارائـه شده است:



نمودار ۲ – مقایسهٔ روش‌ها بر اساس نرخ خطای منفی

مطابق جدول(۳)، مقایسهٔ روش‌ها بر اساس تعداد ویژگی‌ها و تکنیک‌های مورد استفادهٔ آن‌ها در نمودار(۳) ارائه شده است:



نمودار ۳– مقایسهٔ روش‌ها بر اساس تعداد ویژگی‌ها و فن‌های مورد استفادهٔ آن‌ها

نتیجه‌گیری

در این مقاله به حملهٔ فیشینگ و اینکه از مهم‌ترین چالش‌های موجود در بانکداری اینترنتی، خطر این نوع حملات و کلاه‌برداری‌های اینترنتی است، اشاره شد. همچنین، به دلیل خسارت‌های فراوانی که از حملهٔ فیشینگ به مشتریان و سازمان‌ها وارد می‌شود و علاوه بر پول و زمان مشتریان؛ اعتماد آنان به برنامه‌ها و خدمات برخط بانک‌ها نیز از دست رفته و اعتبار سازمان‌ها کم‌رنگ می‌شود، به ضرورت شناخت روش‌های شناسایی حملات فیشینگ و انتخاب روش مناسب تأکید شد. در این مقاله، با ارائه معیارهایی، روش‌های ارائه شده برای شناسایی وب‌سایت‌های فیشینگ، مورد ارزیابی قرار گرفتند. نتایج ارزیابی روش‌های شناسایی وب‌سایت‌های فیشینگ موجود، حاکی از آن است روش‌هایی که از تکنیک‌های مختلف شناسایی در کنار هم استفاده می‌کنند و همچنین، اکثر ویژگی‌های صفحات را بررسی می‌کنند، در شناسایی حمله، موفقیت بیشتری حاصل کرده‌اند. بدین صورت که اکثر روش‌های شناسایی حمله، دارای نرخ خطای مثبت پایین و در مقابل، دارای نرخ بالای خطای منفی هستند. در حالی که روش ضد فیشینگ مبتنی بر متن و ظاهر، یک رویکرد بیزی که اکثر ویژگی‌های روش‌های شناسایی حمله را

داراست و از اکثر تکنیک‌های موجود استفاده می‌کند، دارای نرخ خطای مثبت و منفی نزدیک به صفر است که در نهایت، نشان‌دهندهٔ عملکرد بهتر این روش نسبت به سایرین است. مزیت تحقیق حاضر به تحقیقات انجام شده در این حوزه، شامل دسته‌بندی روش‌های شناسایی وب‌سایت‌های فیشینگ، در نظر گرفتن روش‌های بیشتر در ارزیابی و همچنین، ارزیابی روش‌ها با معیارهای بیشتر است.

پیشنهادها

موضوع‌هایی که می‌توانند به عنوان تحقیقات مرتبط و تکمیلی در آینده انجام شوند، عبارت‌اند از:

- بررسی و تحلیل هزینه‌های محاسباتی و مصرف انرژی روش‌های شناسایی حملهٔ فیشینگ؛
- وزن دهی مناسب به هرکدام از معیارها برای ارزیابی روش‌های شناسایی حملهٔ فیشینگ؛
- دسته‌بندی معیارها بر اساس اهداف مهاجمان و وزن دهی مناسب به هرکدام از معیارها مطابق با آن برای ارزیابی چندگانهٔ روش‌های شناسایی حملهٔ فیشینگ.

منابع

منابع انگلیسی

- Chen, K.-T., Chen, J.-Y., Huang, C.-R., & Chen, C.-S. (2009). Fighting phishing with discriminative keypoint features. *Internet Computing. 13*, pp. 56 –63. IEEE. Retrieved from: doi: *10.1109/MIC.2009.59*
- Chou, N., Ledesma, R., Teraguchi, Y., & Mitchell, J. C. (2004). Client-side defense against web-based identity theft. *The Internet Society.* NDSS. Retrieved from: *https://crypto.stanford.edu/SpoofGuard/webspoof.pdf*
- Dunham, K. (2009). *Mobile Malware Attacks and Defense.* Retrieved from: *http://www.sciencedirect.com/science/book/9781597492980*
- Hara, M., Yamada, A., & Miyake, Y. (2009.). Visual similarity-based phishing detection without victim site information. 30 – 36. Retrieved from: doi:*10.1109/CICYBS.2009.4925087*
- Khonji, M., Iraqi, Y., & Jones, A. (2013). Phishing Detection: A Literature Survey. *IEEE COMMUNICATIONS SURVEYS & TUTORIALS, 15*(4).

Retrieved from:doi:10.1109/SURV.2013.032213.00009
- Likarish, P., Dunbar, D., & Hansen, T. E. (2008). B-apt: Bayesian anti-phishing toolbar. *IEEE International Conference on Communications*, (pp. 1745 –1749). Retrieved from: doi:*10.1109/ICC.2008.335*
- Joshi, Yogesh ; Saklikar ,Samir ; and Saha, Subir . (2008). Phishguard: A browser plug-in for protection from phishing. *2nd International Conference on Internet Multimedia Services Architecture and Applications.* IMSAA. Retrieved from: doi:*10.1109/IMSAA.2008.4753929*
- Liu, G., Qiu, B., & Wenyin, L. (2010). Automatic detection of phishing target from phishing webpage., (pp. 4153 –4156). 20th International Conference on Pattern Recognition (ICPR). Retrieved from: doi:*10.1109/ICPR.2010.1010*
- Nguyen, L. A., To, B. L., Nguyen, H. K., & Nguyen, M. H. (2014). A Novel Approach for Phishing Detection Using URL-Based Heuristic. Retrieved from:doi:*10.1109/ComManTel.2014.6825621*
- Panda, R., & Tiwari, R. (2014). Protection from Phishing Attacks by Exploiting Page Rank,Reputation and Source Code of the Webpage. Retrieved from http://goo.gl/3QhLCv
- Phelps, T. A., & Wilensky, R. (2000). Robust Hyperlinks and Locations. *DLib Magazine*, *6.* Retrieved from doi: *10.1045/july2000-wilensky*
- Prakash, P., Kumar, M., Kompella, R. R., & Gupta, M. (2010). Phishnet: predictive blacklisting to detect phishing attacks. *IEEE Press* (pp. 346–350). Piscataway, NJ, Proceedings of the 29th conference on Information communications. Retrieved from: doi:*10.1109/INFCOM.2010.5462216*
- Salton, G., & McGill, M. (1986). Introduction to Modern Information Retrieval. New York, NY. Retrieved from: *http://dl.acm.org/citation.cfm?id=576628*
- Sharifi, M., & Siadati, S. H. (2008). A Phishing Sites Blacklist Generator. Retrieved from doi:*10.1109/AICCSA.2008.4493625*
- Sheng, S., Wardman, B., Warner, G., Cranor, L. F., Hong, J., & Zhang, C. (2009). An empirical analysis of phishing blacklists. *Proceedings of the 6th Conference in Email and Anti-Spam.* CA: CEAS. Retrieved from: *http://repository.cmu.edu/cgi/viewcontent.cgi?article=1286&context=hcii*
- Singh, P., & Patil, M. D. (2014). Identification of Phishing Web Pages and Target Detection. *3*(2). Retrieved from: *http://ijarcet.org/wp-content/uploads/IJARCET-VOL-3-ISSUE-2-260-263.pdf*
- Whittaker, C., Ryner, B., & Nazif, M. (2010). Large-scale automatic classification of phishing pages. *10.* NDSS. Retrieved from *http://www.internetsociety.org/sites/default/files/whit.pdf*
- Zhang, H., Liu, G., Chow, T., & Liu, W. (2011). Textual and visual contentbased anti-phishing: A bayesian approach. *22*, pp. 1532 –1546. IEEE Transactions on Neural Networks. Retrieved from: doi:*10.1109/TNN.2011.2161999*

- Zhang, Y., Hong, J. I., & Cranor, L. F. (2007). Cantina: a content-based approach to detecting phishing web sites. *Proceedings of the 16th international conference on World Wide Web*, (pp. 639–648). New York, NY, USA. Retrieved from: *http://www2007.org/papers/paper557.pdf*
- Zhou, C. V., Leckie, C., Karunasekera, S., & Peng, T. (2008). A self-healing self-protecting collaborative intrusion detection architecture to traceback fast-flux phishing domains. *NOMS Workshops.* IEEE. Retrieved from: doi:*10.1109/NOMSW.2007.50*

منابع اینترنتی

- APWG. (2016). (Anti-Phishing Working Group: Phishing Activity Trends Report) Retrieved from: *http://www.antiphishing.org*
- Cloudmark. (2016, December). Retrieved from: *http://www.cloudmarkdesktop.com*
- DMOZ. (2016). Retrieved from: *http://rdf.dmoz.org/rdf/*
- Gartner. (2016). (Gartner) Retrieved from: *http://www.gartner.com*
- GoogleDevelopers. (2016, December). *Google Developers*. Retrieved from: *https://developers.google.com/safe-browsing/*
- GoogleSafeBrowsing. (2014, December). *Google Safe Browsing*. Retrieved from:*http://www.google.com/tools/firefox/safebrowsing/*
- McAfeeSiteAdvisor. (2014, December). Retrieved from: *https://www.siteadvisor.com/final/index.html*
- Netcraft. (2014, December). Retrieved from: *http://www.netcraft.com/*
- PhishTank. (2016). Retrieved from: *http://www.phishtank.com*
- Symantec. (2014). *Internet Security Threat Report.* Retrieved from: *https://www.symantec.com/security-center/threat-report*
- TrustWatch. (2016, December). Retrieved from: *http://www.geotrust.com/comcasttoolbar/*