Journal of Electronical & Cyber Defence
Vol. 2, No. 2, 2014, Serial No.6

Spam Management in Social Networks by Content Rating

S. Ghesmati[1] , A. Yazdian Varjani*[2]
1,2- Tarbiat Modares University
(Reccive: 2013/07/29, Accept: 2014/04/10)

Abstract

The high-speed growth of social networks has led to publication of inappropriate contents and spams. Spam posts fill out user's home pages and cause time wasting and also traffic load, by phishing, attackers attempt to acquire sensitive information from a victim in SNs. The other problem in these networks is Adults only posts that users face in their profiles. Inappropriate contents posted by the SN user's friends or spammers would be shown in their home pages or spammers walls. In this paper a new method based on content rating (CR) to identify and manage spam in social networks, is presented. The method is applied for post rating including spam, phishing and Adults only (+18) contents in order to manage home page posts (labeled with spam logo, hide or show) according to users feedbacks. The proposed method has been implemented on open source social network software. Data has been gathered during about two months operation of this SN platform and has been rated and applied in post section of home pages/walls of each SN users. The proposed content rating method has shown significant identification of spam and phishing in this social network.

Keywords:

Social network, Spam, Content rating, Security

***Corresponding Author Email:** yazdian@modares.ac.ir

مجله علمی- پژوهشی "پدافند الکترونیکی و سایبری"

سال دوم، شماره ۲، تابستان ۱۳۹۳، ص ۶۲-۵۳

مدیریت هرزنامه در شبکه‌های اجتماعی با استفاده از برچسب‌گذاری محتوا

سیمین قسمتی[۱]، علی یزدیان ورجانی[۲*]

۱- کارشناسی ارشد، فناوری اطلاعات، دانشکده فنی و مهندسی دانشگاه تربیت مدرس

۲- دکترا، برق، دانشکده برق و کامپیوتر دانشگاه تربیت مدرس

(دریافت: ۹۲/۷/۵ ، پذیرش: ۹۳/۱/۲۱)

چکیده

همزمان با رشد سریع شبکه‌های اجتماعی، انتشار مطالب نامناسب و هرزنامه در این شبکه‌ها مشاهده می‌شود. هرزنامه‌ها سبب می‌شوند که صفحات شخصی کاربران با انبوهی از این پیام‌ها مواجه شده و این امر سبب اتلاف وقت کاربران و بار ترافیکی بالا در هنگام مشاهده این شبکه‌ها می‌شود. علاوه‌براین، ارسال پیام‌های فریب‌کارانه به‌دلیل برقراری ارتباطات آسان در این شبکه‌ها، سبب به‌دست آوردن اطلاعات حساس کاربران می‌شود؛ از دیگر این مشکلات، می‌توان به قرارگیری محتوای مربوط به بزرگسالان اشاره کرد که کاربر بدون علم به این موضوع، در صفحه شخصی خود با چنین پیام‌هایی مواجه می‌شود. مطالب نامناسبی که توسط دوستان کاربران یا ارسال‌کنندگان هرزنامه ارسال می‌شود، در صفحه‌خانه کاربران و صفحه ارسال‌کنندگان هرزنامه قابل مشاهده است. در این مقاله، روشی جدید مبتنی بر برچسب‌گذاری محتوا به‌منظور شناسایی و مدیریت هرزنامه در شبکه‌های اجتماعی ارائه شده است. این روش برای برچسب‌گذاری مطالب هرزنامه، فریب‌کارانه و مربوط به بزرگسالان (۱۸+) به‌کار گرفته شده است تا بدین وسیله، بتوان پست‌های صفحه‌خانه کاربر را بر اساس بازخورد کاربران مدیریت کرده و آنها را برچسب گذاری و پنهان نمود. روش پیشنهادی بر روی یک نرم‌افزار شبکه اجتماعی متن باز پیاده‌سازی شده است. داده‌ها در حدود ۲ ماه از کارکرد این شبکه اجتماعی جمع‌آوری و پست‌ها، برچسب‌گذاری شده و روش پیشنهادی بر روی پست‌های صفحه‌خانه کاربران و صفحه ارسال‌کنندگان هرزنامه اعمال شده است. روش مدیریت محتوای پیشنهادی، کاهش قابل توجهی از مطالب هرزنامه و فریب‌کارانه را در این شبکه نشان داده است.

واژه‌های کلیدی: شبکه اجتماعی، هرزنامه، برچسب‌گذاری محتوا، امنیت

۱. مقدمه

رشد روزافزون طرفداران شبکه‌های اجتماعی، این شبکه‌ها را به‌عنوان یکی از مسائل مهم روز مطرح ساخته است. شبکه‌های اجتماعی و امکانات مهیج آنها، مخاطبان و علاقمندان بسیاری را در سنین مختلف به‌سوی خود جلب نموده است[۱ و ۲]. اغلب این شبکه‌ها، به‌دلیل حجم اطلاعات زیادی که در دسترس آنها می‌باشد، حیاتی به‌شمار می‌آیند و به‌همین دلیل، امنیت و حفظ حریم خصوصی در چنین شبکه‌هایی از نگرانی‌های کنونی است [۱ و ۳].

تاکنون الزامات امنیتی و حریم خصوصی شبکه‌های اجتماعی به‌خوبی درک نشده است [۲]. یکی از مشکلات این شبکه‌ها، ارسال هرزنامه به تعداد زیادی از کاربران است. ارسال‌کنندگان هرزنامه می‌توانند با قرار دادن هرزنامه‌ها در صفحه شخصی خود یا با ارسال پست و نظرات هرز در این شبکه‌ها، سبب ایجاد مشکل برای سایر کاربران شوند [۴ و ۵]. در صورتی‌که راه حل مناسبی برای این مشکل ارائه نشود، ممکن است این مشکل سبب بازپس زدن چنین شبکه‌هایی از سوی کاربران شده و اعتماد آنها را کاهش دهد. اهمیت این موضوع سبب شده است که کمیسیون حفظ داده و حریم خصوصی، در سی‌امین کنفرانس خود، ده راهکار را جهت حفظ حریم خصوصی در شبکه‌های اجتماعی ارائه نماید [۶].

از آنجا که محتوای نامناسب بر کاربران تاثیر منفی می‌گذارد، راه حل‌های زیادی در مقابله با این آسیب‌پذیری شبکه‌ها وجود دارد که برچسب‌گذاری محتوا را می‌توان یکی از آنها دانست. جداسازی وب مربوط به بزرگسالان (۱۸+) و سنین کمتر از ۱۸ سال نیز، یکی از عوامل مهمی است که باید مد نظر قرار گیرد. در برخی کشورها همچون ایران و چین، کاربران با فیلترینگ سایت‌ها مواجه هستند

* رایانامه نویسنده مسئول: yazdian@modares.ac.ir

که برچسب‌گذاری محتوا می‌تواند سبب حل چنین مشکلی شود. برخی تحقیقات، از برچسب‌گذاری محتوا در مطالب موجود در اینترنت استفاده نموده‌اند [۷ و ۸]. هرزنامه در محیط‌هایی چون شبکه اجتماعی، به انتشار محتوای انبوه و ناخواسته از طریق کاربردهای Web 2.0 گفته می‌شود [۹] که اغلب با مقاصد تجاری هستند. تعریف هرزنامه، از دیدگاه کاربر می‌باشد و ممکن است مطلبی از دید یک کاربر، مفید و از دید کاربر دیگر، هرزنامه باشد [۱۰]. پیام‌های فریب‌کارانه، نوعی از مهندسی اجتماعی هستند که مهاجم سعی در به‌دست آوردن اطلاعات حساس قربانی از آن طریق دارد. مهاجمان در این موارد خود را به‌عنوان مرجعی معتبر برای کاربر معرفی می‌نمایند [۱۱ و ۱۲].

طبق تعریف هیئت مدیره برچسب‌گذاری سرگرمی‌های نرم‌افزاری[1] (ESRB)، محتوای مربوط به بزرگسالان (۱۸+)، به محتوایی اطلاق می‌شود که به‌دلیل وجود محتوای ترسناک یا مربوط به قمار و ...، مشاهده آن تنها برای افراد بالای ۱۸ سال مجاز می‌باشد. به‌کارگیری برچسب‌گذاری محتوا در محیط‌هایی چون شبکه‌های اجتماعی، می‌تواند علاوه بر جلوگیری از انتشار هرزنامه، جایگزین فیلترینگ شده و از مشاهده مطالب هرزنامه و نامناسب جلوگیری نماید.

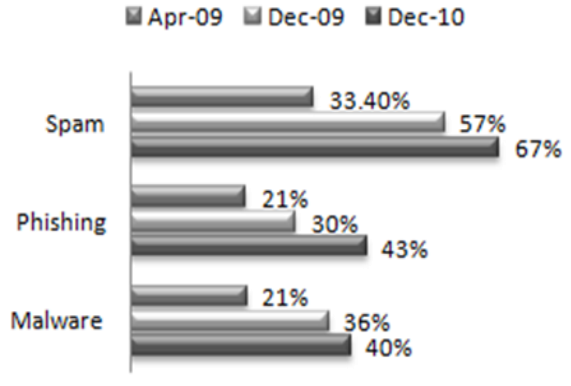
در این مقاله، از برچسب‌گذاری محتوا برای پیام‌های هرزنامه، فریب‌کارانه و مطالب مربوط به بزرگسالان (۱۸+) استفاده شده است که در ادامه به توضیح این مفاهیم پرداخته می‌شود. در بخش۲، هرزنامه در شبکه‌های اجتماعی مورد بحث قرار گرفته است؛ در بخش ۳ توضیحی در رابطه با برچسب‌گذاری محتوا ارائه می‌گردد. بخش ۴ به توضیح روش پیشنهادی پرداخته و نتایج و جمع‌بندی نیز به‌ترتیب در بخش‌های ۵ و ۶ ارائه می‌گردد.

۲. هرزنامه در شبکه‌های اجتماعی

از دست رفتن حریم خصوصی، تهدیدی برای کاربران شبکه‌های اجتماعی محسوب می‌شود. در سال ۲۰۱۰، محققان دریافتند که اطلاعات خصوصی بیش از ۱۰۰ میلیون کاربر شبکه اجتماعی فیس‌بوک از طریق موتورهای جستجو قابل دستیابی است [۱۳]. همچنین کاربران با تهدیداتی چون هرزنامه، نرم‌افزارهای مخرب و مهندسی اجتماعی مواجه هستند [۱۴]. هابر و همکاران، در تحقیقات خود، احتمال حمله‌ای با نام دوست در میان (FITM) را در فیس‌بوک اثبات کرده‌اند و نشان داده‌اند که با زمان کم و منابع سخت‌افزاری ساده، به‌راحتی می‌توان تعداد زیادی هرزنامه را در این شبکه منتشر نمود و با ۴۰۰۰ حمله، $۳/۰۳\times۱۰^{۵}$ کاربر را در خطر دریافت هرزنامه قرار داد [۴]. ابونیمه و همکاران، تحقیقی در مقیاس بزرگ در رابطه با پست‌های مخرب و هرزنامه در فیس‌بوک انجام داده‌اند که در آن از دیفنسیو[2] (یک نرم‌افزار کاربردی در فیس‌بوک) استفاده نموده‌اند. تحلیل داده‌ها در این تحقیق نشان می‌دهد که در حدود ۹٪ پست‌ها در فیس‌بوک، هرزنامه و در حدود ۳٪ پست‌ها، لینک‌های مخرب می‌باشد [۱۴]. لیانگ و همکاران، بر اساس اعتبار به‌دست‌آمده از روابط اجتماعی کاربر، سیستمی جهت مسدود نمودن هرزنامه طراحی نموده‌اند. این سیستم با توجه به واکنش و پاسخ کاربران، خود را به‌روز می‌نماید تا در آینده پاسخ بهتری در مقابل نامه‌های ناخواسته داشته باشد. این تحقیق بر روی نمونه‌های فیس‌بوک پیاده‌سازی شده است [۱۵]. ونگ[3]، سیستمی جهت تشخیص پیام‌های هرزنامه در توییتر طراحی کرده است. روابط دنبال‌کنندگان و دوستان در این شبکه با استفاده از مدل گراف اجتماعی مورد بررسی قرار گرفته است. در این سیستم، از سیاست هرزنامه در توییتر، استفاده از سیستم‌های تشخیص هرزنامه مبتنی بر محتوای پیام و مبتنی بر گراف کمک گرفته شده است. با استفاده از یک خزنده وب و دسته‌بندی بیزین، به شناسایی رفتارهای غیرعادی پرداخته شده است که در این تحقیق، شناسایی هرزنامه با صحت ۸۹٪ بوده است [۱۶].

۱.۲. انتشار هرزنامه در شبکه‌های اجتماعی

شکل ۱ گزارش Sophos [۱۷] را در مورد حملات هرزنامه، نامه‌های فریب‌کارانه و نرم‌افزارهای مخرب در شبکه‌های اجتماعی نشان می‌دهد که بر طبق آن، این حملات در سال‌های اخیر رشد قابل توجهی داشته‌اند؛ به‌طوری‌که در دسامبر ۲۰۱۰، حملات هرزنامه به ۶۷٪ رسیده است [۱۷].



شکل ۱. گزارش حملات هرزنامه، نامه‌های فریب‌کارانه و نرم‌افزارهای مخرب در شبکه‌های اجتماعی [۱۴].

1. Entertainment Software Rating Board (http://www.esrb.org/ratings/ratings_guide.jsp)

2. Defencive
3. Wang

در ادامه به برخی راه‌های استفاده از شبکه‌های اجتماعی به‌منظور ارسال هرزنامه اشاره می‌شود:

- ارسال هرزنامه از طریق پیام خصوصی به کاربر؛ بدین‌وسیله، هرزنامه تنها توسط کاربر قابل مشاهده است [۴].
- ارسال هرزنامه از طریق نظرات یا امکاناتی چون feed؛ بدین‌وسیله، هرزنامه توسط کاربر و دوستان وی قابل مشاهده است [۴، ۵ و ۱۸].
- ارسال هرزنامه از طریق اعلام وضعیت(status) یا امکاناتی چون Tweet به افرادی که در فهرست دوستان ارسال کننده هرزنامه می‌باشند. همچنین، قرار دادن محتویات هرزنامه در صفحه شخصی فرستندگان هرزنامه و دعوت از دیگر کاربران به روش‌های مختلف جهت مشاهده صفحه شخصی و یا دعوت از کاربران با استفاده از صفحه‌های شخصی جذاب؛ در این حالت، هرزنامه تنها توسط قربانی قابل مشاهده است [۵ و ۱۸].
- ایجاد مشاغل جعلی یا گروه‌هایی با مطالب جذاب در شبکه‌های اجتماعی و جذب کاربران جهت به‌دست آوردن اطلاعات آنها و ارسال هرزنامه، نامه‌های فریب‌کارانه یا نرم‌افزارهای مخرب [۱۹].
- ایجاد شبکه‌های اجتماعی جعلی یا شبکه‌هایی با مقاصد تجاری و جمع‌آوری اطلاعات شخصی، آدرس ایمیل و علایق کاربران [۱۴].
- به‌دست آوردن آدرس ایمیل‌ها از طریق صفحه‌هایی که قابل مشاهده برای عموم است [۱۴]. با استفاده از موتورهای جستجو، می‌توان آدرس‌های ایمیل صفحه‌های شخصی را درو نمود. در صورتی که آدرس ایمیل قربانی به‌دست آید، حملات از طریق ایمیل صورت گرفته و شبکه‌های اجتماعی قادر به کشف چنین هرزنامه‌هایی نخواهند بود [۴].
- یافتن کاربران مختلف از طریق امکانات جستجو در شبکه‌های اجتماعی و ارسال هرزنامه به آنها [۲۰].
- به‌دست آوردن ایمیل و اطلاعات کاربران از طریق مهندسی اجتماعی [۴، ۵ و ۱۸]؛ از آنجا که برقراری ارتباط در شبکه‌های اجتماعی آسان‌تر می‌باشد، با ایجاد روابط دوستانه، می‌توان به اطلاعات کاربر از جمله آدرس ایمیل او پی برد.
- حملات Hijack session، استفاده از آسیب‌پذیری‌های Cross site scripting) XSS)، جاسازی نرم‌افزارهای مخرب در قالب به‌روزرسانی Flash player به‌منظور دزدیدن کوکی کاربر و دست‌یابی به محتویات صفحه شخصی قربانی از جمله ایمیل او [۴، ۲۱ و ۲۲].
- حملات Hijack session به‌منظور اضافه نمودن هکر به فهرست دوستان قربانی؛ بدین وسیله علاوه بر اینکه اطلاعات دوستان قربانی به‌دست می‌آید، می‌توان به دوستان قربانی نیز هرزنامه ارسال نمود. بدین ترتیب، گیرندگان، هرزنامه را از طریق دوست خود دریافت خواهند نمود [۴].
- استفاده از روبات ارسال هرزنامه برای ارسال دعوت‌نامه، پست و نظرات هرز به‌صورت خودکار [۵].
- دزدیدن رمز عبور کاربران و قرار دادن نظرات هرز توسط او در صفحه‌های شخصی دیگران [۵].
- استفاده از اطلاعات قرارگرفته در صفحه شخصی کاربران توسط افراد شخص سوم و به‌کارگیری آنها در ارسال هرزنامه [۵].

۳. برچسب‌گذاری محتوا

برچسب‌گذاری، توضیحی مختصر در رابطه با محتوا بیان می‌دارد. حفاظت کودکان و امنیت فرهنگی در اینترنت، از اهداف برچسب‌گذاری محتوا است. شبکه‌های اجتماعی و سایت‌های اینترنتی، سرویس‌هایی هستند که وظیفه بزرگی در مدیریت محتوای مناسب دارند. رشد مطالب موجود در اینترنت علاوه بر مطالب مفید، سبب قرارگیری مطالب نامناسب، مضر و غیرقانونی در دید عموم شده است که این امر، امنیت فرهنگی جامعه را مورد مخاطره قرار می‌دهد. در نتیجه، به‌منظور جلوگیری از آسیب‌های ایجادشده توسط این مطالب، باید برچسب‌گذاری درستی در مورد محتوای مطالب موجود در اینترنت انجام شود تا بتوان محیط سالم و مناسبی را ایجاد نمود [۷].

۱.۳. برچسب‌گذاری هرزنامه در شبکه اجتماعی

با توجه به اهمیت شبکه‌های اجتماعی، استفاده از برچسب‌گذاری محتوا می‌تواند تاثیر به‌سزایی در سلامت چنین شبکه‌هایی داشته باشد. این موضوع تاکنون در شبکه‌های اجتماعی مورد استفاده قرار نگرفته است و از آنجا که در ادبیات موضوع نیز روش‌هایی همچون یافتن هرزنامه در ایمیل پیشنهاد شده است، به‌کارگیری چنین روشی می‌تواند به‌عنوان ایده‌ای جدید مطرح شود. از طرفی، چون به‌کارگیری این روش نظر کاربران را دربر دارد، نسبت به روش‌های

خودکار تشخیص هرزنامه مفیدتر است. به‌کارگیری برچسب‌گذاری محتوا در محیط‌هایی چون شبکه‌های اجتماعی می‌تواند علاوه بر جلوگیری از انتشار هرزنامه، جایگزین فیلترینگ شده و از مشاهده مطالب هرزنامه و نامناسب جلوگیری نماید. علاوه‌براین می‌توان از بازخورد کاربران به یکدیگر در شبکه‌های اجتماعی در سیستم تشخیص هرزنامه استفاده نمود و میزان اعتبار هر کاربر را مشخص کرد [۲۳].

۲.۳. روش پیشنهادی به‌منظور مدیریت هرزنامه در شبکه‌های اجتماعی

در این روش، کاربر امکان برچسب‌گذاری پست‌های ارسال‌شده از دوستان خود را دارد که در صفحه شخصی دریافت می‌کند؛ به این ترتیب که می‌تواند هر پست را به‌عنوان هرزنامه، فریب‌کارانه یا بالای ۱۸ سال انتخاب کند. البته طبق تعاریف سایت esrb.org، برچسب‌های دیگری نیز در رابطه با سایت‌ها و فیلم‌ها وجود دارد، اما این سه مورد برای استفاده در شبکه اجتماعی انتخاب گردید؛ زیرا این قابلیت برای برچسب‌گذاری سایت‌های مختلف استفاده می‌شود و در سایت icra.org قابل ثبت می‌باشد. برچسب‌گذاری با سه نوع هرزنامه، فریب‌کارانه و بزرگسالان (۱۸+) ایجاد شده و برای هر یک از این گزینه‌ها، ستونی در پایگاه داده در جدول پست‌ها تخصیص داده شده است. نحوه کار به این ترتیب است که به‌ازای هر رأی، یک شمارنده برای آن پست در نظر گرفته می‌شود. هر رأی هرزنامه، فریب‌کارانه یا مربوط به بزرگسالان (۱۸+) در ستون‌ها ثبت شده و بزرگ‌ترین رأی در بین این ستون‌ها انتخاب شده و لوگوی مربوطه به‌جای نمایش متن در صفحه کاربر نشان داده خواهد شد. در صورتی

```
read post
if post is spam
    set post as spam in DB
else if post is phishing
    set post as phishing in DB
else post is +18
    set post as +18 in DB
compare post type in DB
select the highest type
hide post content
show post type logo
if logo is clicked
    show post
```

شکل ۲. شبه‌برنامه روش پیشنهادی

که بر روی لوگو کلیک شود، متن برچسب‌گذاری‌شده نشان داده می‌شود. در ضمن، پیش از ثبت رأی در پایگاه داده، بررسی می‌شود که کاربر پیش از این در رابطه با آن پست، رأی ارسال ننموده باشد، تا بدین ترتیب هر کاربر تنها یک حق رأی در رابطه با هر پست داشته باشد. شکل ۲، شبه‌برنامه[1] روش پیشنهادی را نشان می‌دهد.

شکل ۳ چگونگی برچسب‌گذاری پست‌ها از سوی کاربران و شکل ۴ نمای تصویری یک پست برچسب‌گذاری‌شده به‌عنوان هرزنامه را نمایش می‌دهد. به‌منظور این که ارسال‌کنندگان هرزنامه متوجه پنهان شدن پست‌های خود در شبکه نشوند، تمامی پست‌های ارسال شده توسط کاربر بدون هیچ شرطی در صفحه شخصی ایشان نمایش داده می‌شود؛ اما در رابطه با پست‌های دوستان کاربر، شرط

شکل ۳. نحوه برچسب‌گذاری پست‌ها در شبکه اجتماعی

1- pseudocode

شکل ۴. نمای تصویری پست برچسب‌گذاری شده به‌عنوان هرزنامه

بررسی برچسب‌گذاری محتوا قبل از نمایش آن بررسی می‌شود تا پست‌های نامناسب سایر کاربران، در صفحه شخصی آنها به‌صورت برچسب‌گذاری‌شده نمایش داده شود. از آنجا که برخی کاربران تمایل دارند تمامی‌پست‌ها را بدون توجه به برچسب‌گذاری مشاهده نمایند، در قسمت تنظیمات شبکه اجتماعی مورد نظر، امکان نمایش به دو صورت (با برچسب گذاری و بدون برچسب‌گذاری هرزنامه) قرار داده شده است.

۴. آزمایش

نسخه اولیه شبکه اجتماعی، تحقیق از یک نرم‌افزار متن باز[1] بوده است که پس از سفارشی‌سازی و برگردان آن به زبان فارسی و افزودن پیمانه‌های مورد نیاز تحقیق، به‌منظور آزمایش روش پیشنهادی بر روی آدرس www.irancsirt.ir قرار گرفته و در مدت ۵۴ روز- از تاریخ ۹۰/۱۰/۲۷ تا ۹۰/۱۲/۲۱- با ۱۰۰ کاربر مورد آزمایش قرار گرفته است. این افراد در شبکه اجتماعی پیاده‌سازی‌شده عضو شده‌اند. از این افراد خواسته شده که دوستان خود را دنبال نموده و شروع به فعالیت‌هایی چون ارسال پست و نظر نمایند.

۱.۴. سناریو

۹۸ کاربر مجاز و ۲ نفر به‌عنوان ارسال‌کننده هرزنامه به فعالیت خود در این شبکه پرداخته‌اند. فرستندگان هرزنامه، تمامی افراد شبکه را دنبال نموده‌اند. پس از یک ماه از شروع کار شبکه اجتماعی، قابلیت‌های افزوده‌شده به‌منظور برچسب‌گذاری محتوا، به نیمی از کاربران به‌عنوان کاربران تست، توضیح داده شده است.

۲.۴. مجموعه داده‌ها

در مدت آزمایش، ۵۰۲ پست و ۱۰۸ نظر در این شبکه ارسال شده است که از این تعداد، ۱۸۵ پست هرز و ۳۱۶ پست مجاز بوده است. همچنین ۱۸ نظر هرز و ۹۰ نظر مجاز بوده است که بر طبق آن، ۳۶٫۸۵٪ پست‌ها هرزنامه و ۶۲٫۹۴٪ آنها مجاز بوده‌اند. همچنین ۱۷٫۵۹٪ نظرات هرزنامه و ۸۲٫۴۰٪ آنها مجاز بوده‌اند.

۳.۴. آزمایش برچسب‌گذاری محتوا در شبکه اجتماعی

به‌منظور آزمایش برچسب‌گذاری محتوا، در مدت تعیین‌شده برای فعالیت کاربران، امکان برچسب‌گذاری محتوا در سمت سرور بر روی شبکه اجتماعی قرار نگرفته است. بنابراین، در میان پست‌ها، کاربران با پست‌های هرزنامه و فریب‌کارانه نیز مواجه شده‌اند. پس از ۵۴ روز، کدهای سمت سرور تغییر داده شده و امکان برچسب‌گذاری محتوا برای کاربران تست اضافه شده است. از ۲۵۱ پست صفحه‌های شخصی کاربران تست، ۹۶ پست هرزنامه بوده که در حدود ۳۸٫۲۴٪ است و ۲۶ پست فریب‌کارانه بوده که در حدود ۱۰٫۳۵٪ می‌باشد. پست‌ها در بین ساعات ۱۲ تا ۱۴ روز ۹۰/۱۲/۲۰ توسط کاربران تست برچسب‌گذاری شده و از نمایش آنها در شبکه‌اجتماعی جلوگیری شده است که در ادامه به بررسی گزارشات رسیده از برچسب‌گذاری محتوا پرداخته شده و میزان کاهش هرزنامه و تاثیر برچسب‌گذاری محتوا مورد سنجش قرار می‌گیرد.

۵. نتایج

پس از برچسب‌گذاری محتوا توسط کاربران تست، از ۹۶ پست هرزنامه، ۲۶۷ گزارش هرزنامه دریافت گردید که با این روش، ۸۷٫۵٪ هرزنامه‌ها شناسایی شده است. از ۲۶ پست فریب‌کارانه ارسال‌شده در شبکه، ۴۱ گزارش فریب‌کارانه از سوی کاربران دریافت شده است و ۹۶٫۱۵٪ پست‌های فریب کارانه شناسایی شده است. با استفاده از روش برچسب‌گذاری، پست‌های هرزنامه و فریب‌کارانه پنهان شده است و متن پست با لوگوی هرزنامه و فریب‌کارانه جایگزین شده است (جدول ۱).

جدول ۱. گزارش پست‌های هرزنامه و فریب‌کاران توسط کاربران تست

تعداد پیام هرزنامه در صفحه شخصی	تعداد هرزنامه گزارش شده	تعداد پیام برچسب گذاری و پنهان شده	درصد کاهش هرزنامه
۹۶	۲۶۷	۸۴	٪۸۷٫۵
۲۶	۴۱	۲۵	٪۹۶٫۱۵

۱.۵. بررسی عملکرد راه‌حل پیشنهادی در تشخیص هرزنامه

در این قسمت به ارزیابی روش مدیریت هرزنامه در شبکه اجتماعی پیاده‌سازی‌شده پرداخته می‌شود.

۱- نرم‌افزار متن باز شبکه اجتماعی از آدرس زیر دریافت شده است:
http://code.google.com/p/forunpaf/downloads

۱.۱.۵. معیار ارزیابی

دسته مثبت، نشان‌دهنده هرزنامه و دسته منفی، نشان‌دهنده نامه معتبر باشد. در دسته‌بندی دودویی، معمولا چهار حالت پیش می‌آید که در جدول ۲ نشان داده شده‌اند. این چهار حالت مختلف عبارت‌اند از:

- **منفی درست:** پیام معتبری که به درستی در دسته معتبر قرار گرفته است.
- **منفی کاذب:** هرزنامه‌ای که به اشتباه در دسته معتبر قرار گرفته است.
- **مثبت کاذب:** پیام معتبری که به اشتباه در دسته هرزنامه قرار گرفته است.
- **مثبت درست:** هرزنامه‌ای که به درستی در دسته هرزنامه قرار گرفته است.

جدول ۲. حالات ارزیابی تشخیص هرزنامه

برچسب دسته‌بندی کننده	برچسب واقعی نمونه: منفی	برچسب واقعی نمونه: مثبت
منفی	منفی درست	منفی کاذب
مثبت	مثبت کاذب	مثبت درست

جدول ۳. ماتریس نتایج پیام‌های هرزنامه

برچسب دسته‌بندی کننده	برچسب واقعی نمونه: منفی	برچسب واقعی نمونه: مثبت
منفی	٪۹۹/۲۴	٪۱۰/۴۱
مثبت	٪۱	٪۸۹/۵۸

جدول ۴. ماتریس نتایج پیام‌های فریب‌کارانه

برچسب دسته‌بندی کننده	برچسب واقعی نمونه: منفی	برچسب واقعی نمونه: مثبت
منفی	٪۱۰۰	٪۳/۸
مثبت	٪۰	٪۹۶/۱۵

طبق جدول ۳، نتایج به‌دست‌آمده از گزارش هرزنامه در پست‌ها از این قرار است: نرخ منفی درست ٪۹۹/۲۴، منفی کاذب ٪۱۰/۴۱، مثبت کاذب ٪۱ و مثبت درست ٪۸۹/۵۸.

طبق جدول ۴، نتایج به‌دست‌آمده از گزارش فریب‌کارانه در پست‌ها بدین گونه است: نرخ منفی درست ٪۱۰۰، منفی کاذب ٪۳/۸، مثبت کاذب ٪۰ و مثبت درست ٪۹۶/۱۵.

در طول آزمایش، هیچ‌گونه پستی با محتوای مربوط به بزرگسالان (۱۸+) بر روی شبکه قرار نگرفت که به‌همین دلیل، آماری از نتایج حاصل از آن نمی‌توان بیان نمود.

۲.۱.۵. یادآوری، دقت، صحت و خطا

فرمول‌های ۱ تا ۴ چگونگی محاسبه مقادیر یادآوری، دقت، صحت و خطا را نشان می‌دهد.

صحت :
$$Acc = \frac{n_{L\to L} + n_{S\to S}}{N_L + N_S} \quad (۱)$$

نرخ خطا :
$$Err = \frac{n_{L\to S} + n_{S\to L}}{N_L + N_S} \quad (۲)$$

یادآوری :
$$SR = \frac{n_{S\to S}}{n_{S\to S} + n_{S\to L}} \quad (۳)$$

دقت :
$$SP = \frac{ns\to S}{n_{S\to S} + n_{L\to S}} \quad (۴)$$

که در این فرمول‌ها، S به معنی هرزنامه، L به معنی پیام معتبر، n_S تعداد هرزنامه- که به درستی به‌عنوان هرزنامه شناخته شده است -n_L تعداد هرزنامه- که به‌عنوان نامه معتبر دسته‌بندی شده‌اند-، N_L تعداد نامه معتبری که قرار است دسته‌بندی شوند و N_S تعداد هرزنامه‌ای که قرار است دسته‌بندی شوند، می‌باشد.

طبق نتایج به‌دست‌آمده برای پست‌های هرزنامه، مقدار یادآوری ٪۸۹/۵۸، دقت ٪۹۸/۸۵، صحت ٪۹۵/۱۹ و خطا ٪۴/۸ بوده است و از نتایج به‌دست‌آمده در رابطه با پست‌های فریب‌کارانه، مقدار یادآوری ٪۹۶/۱۵، دقت ٪۱۰۰ ، صحت ٪۹۹/۳۷ و خطا ٪۰/۶۲ بوده است

۲.۵. مقایسه پست‌های هرزنامه و فریب‌کارانه قبل و پس از به‌کارگیری روش برچسب‌گذاری محتوا

جدول ۶، مقایسه پست‌های هرزنامه و فریب‌کارانه قبل و پس از به‌کارگیری روش برچسب‌گذاری محتوا برای کاربران تست را نشان می‌دهد. پیش از به‌کارگیری برچسب‌گذاری محتوا، پست‌های هرز در

جدول ۵. نتایج حاصل از برچسب‌گذاری محتوا

برچسب‌گذاری پست	یادآوری	دقت	صحت	خطا
هرزنامه	٪۸۹٫۵۸	٪۹۸٫۸۵	٪۹۵٫۱۹	٪۴٫۸۰
فریب‌کارانه	٪۹۶٫۱۵	٪۱۰۰	٪۹۹٫۳۷	٪۰٫۶۲
بالای ۱۸ سال	–	–	–	–

جدول ۶. مقایسه پست‌های هرزنامه و فریب‌کارانه قبل و پس از به‌کارگیری روش برچسب‌گذاری محتوا

	هرزنامه	فریب‌کارانه
پیش از به‌کارگیری برچسب‌گذاری	٪۳۸٫۲۴	٪۱۰٫۳۵
پس از به‌کارگیری از برچسب‌گذاری	٪۴٫۷۸	٪۰٫۳۹

حدود ٪۳۵٫۴۵ و پست‌های فریب‌کارانه در حدود ٪۱۱٫۵ بوده است. پس از به‌کارگیری روش برچسب گذاری محتوا، پست‌های هرز به حدود ٪۴٫۷۸ و پست‌های فریب‌کارانه به حدود ٪۰٫۳۹ رسیده است. باید توجه نمود که کاهش نمایش پست های هرزنامه و فریب کارانه سبب کاهش بار ترافیکی در صفحه شخصی کاربران شده است. افزایش اعتماد کاربران و ایجاد محیطی امن، از دیگر مزایای این روش می‌باشد.

۳.۵. تحلیل نتایج با استفاده از آزمون فرض

در این قسمت، نتایج به‌دست‌آمده از کاربران تست، پیش از به‌کارگیری روش برچسب گذاری محتوا و پس از آن با آزمون میانگین‌های نمونه‌های جفت بررسی می‌شوند.

۱.۳.۵. آزمون میانگین‌های نمونه‌های جفت

اگر بخواهیم اندازه دو صفت را در هر فرد نمونه به‌دست آوریم، می‌گوییم یک نمونه جفت داریم که در اینجا، میزان هرزنامه قبل و پس از به‌کارگیری روش برچسب‌گذاری محتوا در کاربران تست است. در جدول ۷، مقادیر ستون N، تعداد داده‌های هر یک از نمونه‌ها را نشان می‌دهد. مقادیر ستون mean، میانگین‌های حسابی دو نمونه هستند. مقادیر ستون Std. Deviation، انحراف معیار نمونه‌ها و مقادیر ستونstd. Error mean، انحراف معیار میانگین نمونه‌ها را نشان می‌دهد.

جدول ۷. مقایسه آماری پست‌های فریب‌کارانه کاربران قبل و بعد از روش برچسب‌گذاری محتوا

	Mean	N	Std. Error Mean	Std. Deviation
هرزنامه قبل از برچسب‌گذاری محتوا	۱٫۹۲	۵۰	۰٫۳۶۲	۲٫۲۶۵
هرزنامه بعد از برچسب‌گذاری محتوا	۰٫۲۰	۵۰	۰٫۰۷۶	۰٫۵۳۵
فریب‌کارانه قبل از برچسب‌گذاری محتوا	۰٫۵۲	۵۰	۰٫۰۹۶	۰٫۵۳۵
فریب‌کارانه بعد از برچسب‌گذاری محتوا	۰٫۰۲	۵۰	۰٫۰۲۰	۰٫۱۴۱

جدول ۸. نتایج آزمون t

			هرزنامه	فریب کارانه
Paired Differences		Mean	۱٫۷۲۰	۰٫۵۰۰
		Std. Deviation	۲٫۱۳۸	۰٫۶۴۷
		Std. Error Mean	۰٫۳۰۲	۰٫۰۹۱
	95% Confidence Interval of the Difference	Lower	۱٫۱۱۲	۰٫۳۱۶
		Upper	۲٫۳۲۸	۰٫۳۱۶
		t	۵٫۶۸۷	۵٫۴۶۶
		Df	۴۹	۴۹
		Sig. (2-tailed)	۰۰۰٫۰	۰۰۰٫۰

در جدول ۸ فرض می‌کنیم $x_{11}, x_{21},\ldots,x_{n1}$ مقادیر اولین صفت و $x_{12}, x_{22}, \ldots, x_{n2}$ مقادیر دومین صفت باشند. میانگین اولین صفت را در نمونه تست با $\overline{x_1}$ و میانگین دومین صفت را در نمونه تست با $\overline{x_2}$ نشان می‌دهیم. فرض می‌کنیم i=1,2….,n و $di=x_{i1}-x_{i2}$ تفاضل دو صفت باشد. مقدار $\bar{d}=\overline{x1}-\overline{x2}$ در سطر mean نوشته شده است و اگر میانگین دو صفت در جامعه را با u_1 و u_2 نشان دهیم، آنگاه کران پایین و کران بالای فاصله اطمینان u_1-u_2 به‌ترتیب در سطرهای upper و lower آمده است. فرض $H_0: u_1-u_2=0$ در مقابل $H_1:u_1-u_2\neq0$ قرار دارد که از آماره $\bar{d}/(sd/\sqrt{n})$ استفاده می‌شود که مقدار آن در سطر t نشان داده شده است. درجه آزادی این آماره در سطر Df آمده است. سطح معنی‌داری این آزمون در سطر Sig.(2tailed) نوشته شده است. نتایج به‌دست‌آمده، فرض H_0 را حداکثر با اطمینان ٪۹۵ برای پست‌های هرز و فریب‌کارانه رد می‌کند.

۴.۵. رضایت کاربران

برای اطلاع از رضایت کاربران، پرسشنامه‌ای با ۹ سؤال تهیه گردیده و در اختیار کاربران قرار داده شده است. این پرسشنامه، بسته پاسخ بوده است که ۴ سوال اول، اطلاعاتی در مورد پاسخ‌دهنده و ۵ سوال بعدی با پاسخ‌های: ۵ (خیلی زیاد)، ۴ (زیاد)، ۳ (متوسط)، ۲ (کم) و ۱ (خیلی کم) با استفاده از مقیاس لیکرت ارائه شده است. نمونه جامعه انتخابی به‌صورت احتمالی و تصادفی، ساده انتخاب شده‌اند و ۳۰ پرسشنامه، بین کاربران نمونه توزیع شده است.

۱.۴.۵. سؤالات پرسشنامه رضایت کاربران

۹ سؤال پرسشنامه عبارت‌اند از:

- سن
- جنسیت
- میزان تحصیلات
- به‌طور متوسط چه میزان در ماه، از شبکه‌های اجتماعی آنلاین استفاده می‌کنید؟
- به نظر شما برچسب‌گذاری محتوا به چه میزان در مدیریت هرزنامه در شبکه اجتماعی موثر بوده است؟
- به‌کارگیری برچسب‌گذاری محتوا تا چه میزان در کاهش نمایش پیام‌های هرز، فریب‌کارانه و مربوط به بزرگسالان (۱۸+) در شبکه اجتماعی موثر است؟
- به‌کارگیری برچسب‌گذاری محتوا تا چه میزان در افزایش اعتماد شما به شبکه اجتماعی موثر بوده است؟
- به نظر شما روش به‌کارگرفته‌شده در تحقیق، در عدم نمایش هرزنامه در شبکه اجتماعی تا چه میزان در صرفه‌جویی وقت کاربران موثر بوده است؟
- به نظر شما روش به‌کارگرفته‌شده به‌منظور عدم نمایش محتوای هرزنامه و بارگذاری آن در نمایه کاربر، تا چه میزان بر کاهش بار ترافیکی سایت موثر است؟

۲.۴.۵. نتایج پرسشنامه

در این تحقیق، از نرم‌افزار SPSS نسخه ۱۸/۰ برای آمار توصیفی پرسشنامه رضایت کاربران استفاده شده است. سن افراد پاسخ‌گو در پرسشنامه بدین ترتیب است: ۱۰٪ زیر ۲۰ سال، ۶۶/۷٪ بین ۲۰ تا ۳۰ سال، ۱۰٪ بین ۳۱ تا ۴۰ سال، ۱۰٪ بین ۴۱ تا ۵۰ سال و ۳/۳٪ بالای ۵۰ سال. غالب افراد شرکت‌کننده را افراد بین ۲۰ تا ۳۰ سال تشکیل می‌دهند و ۷۰٪ شرکت‌کنندگان مونث و ۳۰٪ مذکر می‌باشند که جنس مونث، غالب است.

میزان تحصیلات افراد پاسخ‌گو در پرسشنامه بدین‌ترتیب است: ۱۳/۳٪ دیپلم، ۴۰٪ لیسانس، ۴۰٪ فوق لیسانس، ۶/۷٪ دکترا که جمعیت غالب را افراد با مدرک لیسانس و فوق لیسانس تشکیل می‌دهند.

میزان استفاده از شبکه‌های اجتماعی در طول یک ماه بر حسب ساعت برای افراد پاسخ‌گو در پرسشنامه بدین‌ترتیب است: ۱۶/۷٪ کمتر از یک ساعت، ۲۳/۳٪ بین ۱ تا ۱۰ ساعت، ۶/۷٪ بین ۱۱ تا ۲۰ ساعت، ۱۰٪ بین ۲۱ تا ۳۰ ساعت و ۴۳٪ بیش از ۳۰ ساعت از این شبکه‌ها استفاده می‌نمایند که جمعیت غالب را افراد استفاده کننده از شبکه‌های اجتماعی با بیش از ۳۰ ساعت در ماه تشکیل می‌دهند.

جدول۹، میزان مدیریت هرزنامه، میزان کاهش هرزنامه، میزان افزایش اعتماد کاربران، میزان صرفه‌جویی در زمان و میزان کاهش بار ترافیکی با استفاده از برچسب‌گذاری محتوا در شبکه‌های اجتماعی را نشان می‌دهد. از این میان،۵۳/۳٪ تاثیر مدیریت هرزنامه را زیاد، ۶۶/۷٪ تاثیر میزان کاهش هرزنامه را زیاد، ۵۰٪ تاثیر میزان افزایش اعتماد را زیاد، ۶۳/۳٪ تاثیر میزان صرفه‌جویی در زمان را زیاد و ۵۶/۷٪ تاثیر میزان کاهش بار ترافیکی را زیاد دانسته‌اند.

جدول ۹. آمار به‌دست‌آمده از پرسشنامه رضایت کاربران

	درصد مدیریت هرزنامه	درصد کاهش هرزنامه	درصد افزایش عتماد کاربران	درصد صرفه‌جویی در زمان	درصد کاهش بار ترافیکی
خیلی کم					
کم		۶/۷			۳/۳
متوسط	۳۳/۳	۱۰/۰	۲۶/۷	۱۶/۷	۲۰/۰
زیاد	۵۳/۳	۶۶/۷	۵۰/۰	۶۳/۳	۵۶/۷
خیلی زیاد	۱۳/۳	۱۶/۷	۲۳/۳	۲۰/۰	۲۰/۰

۶. نتیجه

در این مقاله روشی جدید مبتنی بر برچسب‌گذاری محتوا به‌منظور شناسایی و مدیریت هرزنامه در شبکه‌های اجتماعی ارائه

شد. این روش برای برچسب‌گذاری مطالب هرزنامه، فریب‌کارانه و مربوط به بزرگسالان(۱۸+) به‌کار گرفته شد تا بدین‌وسیله، بتوان پست‌های صفحه خانه کاربر را بر اساس بازخورد کاربران مدیریت کرده، آنها را برچسب‌گذاری و پنهان نمود. روش پیشنهادی بر روی یک نرم‌افزار شبکه اجتماعی متن باز پیاده‌سازی شد. داده‌ها در حدود ۲ ماه از کارکرد این شبکه اجتماعی جمع‌آوری و پست‌ها برچسب‌گذاری شده است و روش پیشنهادی بر روی پست‌های صفحه خانه کاربران و صفحه ارسال‌کنندگان هرزنامه اعمال شده است. نتایج نشان می‌دهد که روش مدیریت محتوای پیشنهادی، سبب کاهش قابل توجه نمایش مطالب هرزنامه و فریب‌کارانه و کاهش بار ترافیکی به‌دلیل عدم نمایش پست‌های هرزنامه و فریب‌کارانه در این شبکه اجتماعی شده است.

۷. مراجع

[1] Ch. Zhang, J. Sun, , X. Zhu and Y. Fang, "Privacy and Security for Online Social networks: Challenges and Opportunities," IEEE Network, vol. 24, no.4, pp. 13-18, Jul/Aug 2010, doi: 10.1109/MNET.2010.5510913.

[2] G. Ahn, M, Shehab and A. Squicciarin, "Security and Privacy in Social Networks," IEEE Internet computing, vol. 15, no. 3, pp. 10-12, available at http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5755600, May/June 2011, doi: 10.1109/MIC.2011.66.

[3] D. Irani, S. Webb, C. Pu and K. Li, "Modeling Unintended Personal-Information Leakage from Multiple Online Social Networks," IEEE Internet computing, vol. 15, no.3, pp. 13-19, May/June 2011, doi: 10.1109/MIC.2011.25.

[4] M. Huber, M. Mulazzani, G. Kitzler, S. Goluch and E. Weippl, "Friend-in-the-middle Attacks: Exploiting Social Networking Sites for Spam," IEEE Internet computing, vol. 15, no.3, pp. 28 – 34, May/June 2011, doi: 10.1109/MIC.2011.24.

[5] M.L. Lobina, D. Giusto, D. Mula and E. Maggio, "Anti-spam laws at the times of social networks: the European approach" http://www.dimt.it/File/Giusto_Lobina_Maggio_Mula%20-%20Spam.pdf. 2010.

[6] Data Protection and Freedom of Information Commissioner of the State of Berlin (Germany), "Resolution on Privacy Protection in Social Network Services," 30th International Conference of Data Protection and Privacy Commissioners Strasbourg. http://www.ed- ps.europa.eu/EDPSWEB/webdav/shared/Documents/Cooperation/Conference_int/08-10-17_Strasbourg _social_network_EN.pdf. 2008.

[7] Y. Liu., G. Zhao, and T. Wang, "Architecture Research and Design for the Internet Content Rating Implementation System," Proc. 2nd IEEE workshop. Education Technology and Computer Science (ETCS), pp. 369-401, March. 2010, doi: 10.1109/ETCS.2010.419.

[8] L Yan, and W. Tai, "Technical Standard System Development for the Internet Content Rating Service," Proc. 3rd IEEE workshop. Intelligent Systems and Applications (ISA), pp. 1-5, May. 2011, doi: 10.1109/ISA.2011.5873326.

[9] P. Hayati, V. Potdar, A.Talevski, N.Firoozeh, S. Sarenche, and E.A. Yeganeh, "Definition of Spam 2.0: New Spamming Boom" 4th IEEE International Conference on Digital Ecosystems and Technologies (DEST), pp. 580-584, April 2010.

[10] Y. Li, B. Fang, L. Guo, and S. Wang, "Research of a novel anti-spam technique based on users feedback and improved naive Bayesian approach" International conference on Networking and Services, 2006. ICNS '06., pp. 86–86, July2006.

[11] T. Jagatic, N. Johnson, M. Jakobsson, and F. Menczer, "Social Phishing" http://efiko.org/material/Social%20Phishing%20by%20Tom%20Jagatic%20et.%20al.pdf . 2005

[12] K. Coronges, R.Dodge, C. Mukina, Z. Radwick, J. Shevchik, and E. Rovira, "The Influences of Social Networks on Phishing Vulnerability" 45th Hawaii International Conference on System Science (HICSS), pp. 2366- 2373, Jan. 2012.

[13] N. Bilton, "Researcher Releases Facebook Profile Data" http://bits.blogs.nytimes.com/2010/07/28/100-million-facebook-ids-compiled-online. 2010.

[14] S. Abu Nimeh, T. Chen and O. Alzubi, "Malicious and Spam Posts in Online Social Networks," IEEE computer, vol. 44, no. 9, pp. 23–28, Sep. 2011, doi: 10.1109/MC.2011.222.

[15] Ch. Liang, Y. Chen, G. Liao and B. Cheng, "Anti-spam Email System in Facebook," Proc. IEEE Symp. Computer, pp. 183-186, Dec. 2010, doi: 10.1109/COMPSYM. 2010.5685522.

[16] A. H. Wang, "Don't follow me: Spam Detection in Twitter," Proc. Conf. Security and Cryptography (SECRYPT), pp. 1-10. July. 2010.

[17] Sophos, "Sophos Security Threat Report reveals increase in social networking security threats," http://www.sophos.com/en-us/press-office/press-releases/2011/01/threat-report-2011.aspx. 2011.

[18] G. Stringhini, Ch. Kruegel and G. Vigna, "Detecting Spammers on Social Networks," iseclab.org/papers/acsac10-socialnets.pdf. 2010.

[19] Security parks, "Do not falling victim of social networking spam," http://www.securitypark.co.uk/security _article 262665.html. 2009.

[20] L. Munson, "5 Types Of Social Networking Scam – #5 Spam," http://www.security-faqs.com/5-types-of-social-networking-scam-5-spam.html. 2008.

[21] W. Luo, J. Liu, J. Liu, and Ch. Fan, "An Analysis of Security in Social Networks," Proc. Eighth Conf. Dependable Autonomic and Secure Computing, pp. 648-651, Dec. 2009, doi: 10.1109/DASC.20-09.100.

[22] S. Liss and S. S. Columnist, "As Facebook and Twitter grow, so do opportunities to rip people off," http://articles.sun-sentinel.com/2010-07-31/news/fl-slcol-social-media-scams-201007-31_1_social-media-scam-facebook. 2010.

[23] H. Y. Lam and D. Y. Yeung, "A Learning Approach to Spam Detection based on Social Networks," www.cse.ust.hk/~dyy-eung/paper/pdf/yeung.ceas2007.pdf. 2007.