ISSN 1870-4069

Detección de ciberataques a través del análisis de mensajes de redes sociales: revisión del estado del arte

Jorge Enrique Coyac-Torres, Grigori Sidorov, Eleazar Aguirre-Anaya

Instituto Politécnico Nacional,
Centro de Investigación en Computación,
México

B160512@sagitario.cic.ipn.mx, sidorov@cic.ipn.mx,
eaguirrea@ipn.mx

Resumen. Este artículo muestra que de la misma forma como se ha utilizado la información de redes sociales para extraer ciertos eventos específicos; por ejemplo, la aceptación de un producto en el mercado, preferencias políticas del público, eventos de salud, entre otros. Del mismo modo, existen investigaciones, que utilizan estos datos, para detectar diversos eventos o actos maliciosos que podrían poner en riesgo al ciberespacio. Por ende, este trabajo busca identificar las áreas de oportunidad que puedan existir para desarrollar futuras investigaciones de este tipo, a partir de la comparación de resultados obtenidos de cada uno de estos trabajos. Igualmente, demuestra como las herramientas de procesamiento del lenguaje natural, y los algoritmos de aprendizaje automático y profundo, han contribuido con estas tareas al ser implementadas, por cada autor, dentro de sus metodologías.

Palabras clave: Ciberataque, redes sociales, procesamiento del lenguaje natural, ciberseguridad, aprendizaje automático.

Cyberattacks Detection by Social Network Media Messages Analysis: State of the Art Review

Abstract. This paper shows that in the same way that information from social networks has been used to extract certain specific events; for example, the acceptance of a product in the market, political preferences, health events, among others. Similarly, there are investigations, which use this data to detect various malicious events or acts that could put cyberspace at risk. Therefore, this work identify the areas of opportunity that may exist for the development of future research of this type, from the comparison of results obtained by each of these works. Likewise, it demonstrates how natural language processing tools, and deep and

machine learning algorithms, have contributed to this type of task by being implemented, by each author, within their methodologies.

Keywords: Cyberattack, social networks, natural language processing , cybersecurity, machine learning.

1. Introducción

El poder de difusión de la información que tienen las redes sociales, principalmente para llegar a muchos lugares, se debe al impulso que le han otorgado las herramientas derivadas de las Tecnologías de la Información(TI). Pero si bien, cuando la tecnología evoluciona a pasos muy acelerados para brindar más facilidades y comodidades, de igual manera los riesgos y amenazas crecen con ella.

Este artículo muestra un panorama actualizado sobre aquellas investigaciones orientadas al análisis de la información contenida en mensajes de redes sociales, teniendo estas como objetivo en común, el detectar los eventos y/o actos que reflejen comportamientos maliciosos que amenacen el ciberespacio. Los resultados obtenidos, en cada una de las investigaciones, permitieron a sus autores la creación de metodologías y mecanismos que ayudaron a detectar y prevenir el riesgo de futuros ciberataques. Por lo tanto, este trabajo demuestra que las siguientes investigaciones justifican la idea de que, entre la diversidad de los temas discutidos en las redes sociales también se puede encontrar información, que después de aplicarle diversas técnicas de aprendizaje automático y de procesamiento del lenguaje natural, sirva como una herramienta de apoyo enfocada a exponer diversos eventos de ciberseguridad.

Tales investigaciones tomaron un enfoque de análisis similar a los que se han realizado en otras áreas; por ejemplo, la aceptación de algún nuevo producto en el mercado [4], el seguimiento de los eventos políticos [23], de salud [10], económicos [22], entre otros que precisamente hacen uso de la información contenida en los mensajes de las redes sociales. Pero con la diferencia de que los siguientes trabajos están orientados a eventos de ciberseguridad, como es la detección de los tipos de ciberataques, predicción de los eventos relacionados a la ejecución de un ciberataque, e incluso las posibilidades de usar una falla en software para poder ejecutar un ataque tomando ventaja de la misma.

Cabe resaltar que la mayoría de estos trabajos, no se hubieran logrado sin una metodología basada en el uso de herramientas del área de procesamiento del lenguaje natural, y apoyada en algunos casos por el uso de algoritmos de aprendizaje automático o aprendizaje profundo. Este artículo está organizado de la siguiente manera. La sección 2, presenta un análisis sobre las redes sociales y los eventos de ciberseguridad. En la sección 3, se encuentra una revisión de trabajos dirigidos a la detección de ciberataques empleando la información que es publicada en las redes sociales. Finalmente, la sección 4 es dedicada a nuestra conclusión.

2. Las redes sociales y eventos de ciberseguridad

Las redes sociales se han convertido en un gran medio masivo de difusión de la información, y actualmente han llegado a lugares donde la humanidad difícilmente se habría imaginado explorar. Diversos tópicos como política, entretenimiento, tecnología y muchos otros más son discutidos en ellas.

Toda esta diversidad de temas, ha permitido desarrollar varios trabajos dedicados al análisis de información generada por los mismos usuarios, teniendo a estos miembros participando en una forma similar a la de sensores de eventos sociales. Por lo tanto, la cantidad y calidad de la información contenida en las redes sociales depende, en gran medida, de la interacción que existe entre los integrantes de las mismas. Siendo los mensajes que se publican, referentes a los diversos sucesos que ahí se discuten, la forma de retro-alimentar esta fuente de información.

Han existido numerosos trabajos que se han basado en la extracción de información que existe en las redes sociales referentes a algún evento en específico; por ejemplo, el análisis del impacto en el mercado sobre algún nuevo producto [4], el seguimiento de alguna enfermedad dentro de una comunidad [10], el seguimiento de preferencias durante elecciones políticas [23], y varios más que le prestan atención a la plétora de eventos que puedan seguir surgiendo en una sociedad tan globalizada al día de hoy. Por ende, entre todos esos trabajos, existen otros cuya atención está dirigida al hallazgo de información del área de ciberseguridad, y posteriormente, utilizan todos esos datos para crear herramientas o mecanismos de prevención sobre futuros eventos y riesgos que atenten en contra del ciberespacio.

La preocupación ante el riesgo de ser víctimas de distintos tipos de ciberataques, y que estos puedan generar daños a infraestructuras y servicios de cómputo de usuarios, negocios, empresas e incluso instituciones de gobierno, ha generado un gran interés por parte de muchos investigadores el poder anticiparse a estos eventos. Estudios como el de Saidi [19] han expuesto la importancia de un análisis de las comunidades e información contenida de las redes sociales, concluyendo que un modelado y análisis semántico pueden ayudar a extraer información para proporcionar una visión profunda de las operaciones de grupos clandestinos de terroristas cibernéticos.

La siguiente sección muestra aquellos trabajos que se han concentrado en la detección de diferentes tipos de ciberataques. Todos ellos utilizando la información contenida en las redes sociales, y apoyándose del uso de técnicas y algoritmos del área de procesamiento del lenguaje natural, aprendizaje automático o aprendizaje profundo.

3. Detección de ciberataques en las redes sociales

Antes que nada, es importante aclarar que el análisis de la información contenida en mensajes de redes sociales, para la detección de ciberataques, no puede ser comparado con las herramientas y mecanismos comúnmente usados

en el área de ciberseguridad; por ejemplo, los antivirus, *firewalls*, Sistemas de Detección de Intrusos (o por sus siglas en inglés *IDS*), Sistemas de Prevención de Intrusos (o por sus siglas en inglés *IPS*), entre otros.

Los siguientes trabajos tienen como meta descubrir amenazas que puedan existir en el ciberespacio, pero de una forma alternativa. Por ende, estos no pretenden de alguna manera ser considerados como herramientas para mitigar los diferentes tipos de ciberataques. Más bien, sólo cumplen con la función de avisar o alertar del hallazgo de diversas formas de este tipo de amenazas.

Este artículo expone cuatro enfoques diferentes sobre la detección de ciberataques, los cuales son: **detección de eventos relacionados a ciberataques**, **detección de tipos de ciberataques**, **interés en vulnerabilidades de software para ejecutar un ciberataque**, y finalmente **ciberataques en las redes sociales**. Las siguientes subsecciones muestran trabajos relevantes que son parte de esta revisión del estado del arte.

3.1. Detección de eventos relacionados a ciberataques

Este apartado considera los trabajos enfocados en la predicción y extracción de posibles eventos que están relacionados a un tipo de ciberataque, como son el secuestro de cuentas (Hijacking), la denegación de servicio distribuido (o de sus siglas en inglés DDoS), entre otros.

Ritter [16] demostró que un gran número de eventos relacionados a fallos en ciberseguridad son mencionados en Twitter, logrando crear un extractor de esos eventos a través de un método semi-supervisado aplicado al flujo de publicaciones en esta red social. Mientras que Khandpur [11] desarrolló un método que de forma dinámica podía extraer aquellos eventos relacionados a ciberataques reportados y discutidos en Twitter, detectando un rango de tres diferentes tipos de amenazas.

Los diferentes trabajos de Hernández [9] [8] plantean que se puede predecir la respuesta de grupos específicos involucrados en Hacktivismo (acrónimo de *hacker* y activismo) cuando los sentimientos son lo suficientemente negativos hacia un usuario, ambos métodos reúnen un *corpus* de tweets, emplean análisis de sentimientos, y finalmente usan diferentes herramientas estadísticas en cada trabajo para predecir la posible aparición del ciberataque.

Pero no sólo Twitter ha demostrado ser una red social recurrente para la obtención de información. También se han analizado mensajes de otras redes sociales provenientes de foros *Hacker* que residen en la *Deep/Dark* web. Tal es el trabajo de Goyal [6], donde empleó herramientas de aprendizaje profundo (Redes neuronales profundas) y series de tiempo para predecir ciberataques con información de diversos foros Hacker.

Deliu [5] concluyó con un análisis comparativo entre Máquinas de Vectores de Soporte (o por sus siglas en inglés SVM) y Redes Neuronales Convolucionales (o por sus siglas en inglés CNN) para la detección de estos eventos dentro de foros similares a los previos. Encontrando resultados significativamente parecidos entre ambas herramientas de aprendizaje.

Finalmente, la tabla 1 muestra una comparación de los resultados obtenidos por los trabajos previamente mencionados. Aquí se observa que los algoritmos más recientes, del área de aprendizaje automático y profundo, han contribuido en la mejora de resultados, según la medida F1 en cada uno de los trabajos. Del mismo modo, se puede visualizar que entre más específico es el ciberataque a detectar, la medida F1 tiende a incrementar su valor.

Tabla 1. Comparación de resultados entre las investigaciones enfocadas a la detección de eventos relacionados a ciberataques.

Investigación	**Objetivo**	**Algoritmos**	**Medida F1**
Hernández [8]	Eventos de ciberseguridad	Regresión lineal	0.442
Goyal [6]	Malware, e-mail malicioso, URL malicioso	Redes neuronales profundas	0.653
Ritter [16]	DDoS, Hijacking, Fuga de datos	SVM + regularización L2	0.676
Khandpur [11]	DDoS, Hijacking, Fuga de datos	Convolución de kernels	0.71
Deliu [5]	Malware, Spam, DDoS	CNN , SVM	0.976

3.2. Tipos de ciberataques

Los trabajos aquí presentados están orientados a detectar y clasificar distintas formas de ciberataques que son distribuidos en las redes sociales, y que pueden engañar al usuario para que interactúe con el contenido que existe dentro de los mensajes de estas redes; por ejemplo, acceder a sitios con malware a través de una URL, ataques del tipo ingeniería social, Spam, descarga de contenido malicioso, entre otros. De tal manera que estas amenazas tienen como objetivo afectar y desestabilizar el sistema o recurso que le pertenece al usuario.

En Liao [12] presentan un framework llamado *iACE* cuya función es extraer Indicadores de Compromiso (o por sus siglas en inglés IoC [15]) a partir de textos no estructurados. El concepto IoC se define como una descripción, sobre alguna actividad y/o artefacto malicioso, que es relevante a un incidente de ciberseguridad, y que este puede ser identificado a través del análisis de sus patrones de comportamiento. Con base en lo anterior, el autor propone el uso de técnicas de procesamiento del lenguaje natural(NER/ER) para la recolección de información proveniente de fuentes públicas como blogs, artículos y otros medios escritos de acceso público, y crear una serie de términos que permitan definir características únicas de diferentes ciberataques.

Un trabajo dedicado a la detección de *Phishing* fue el de Wooi [20], donde desarrollaron un mecanismo de alerta de seguridad en tiempo real que se activa al encontrar este tipo de amenazas en los mensajes de Twitter. Dicha investigación utilizó un modelo de clasificación derivado del aprendizaje automático "Random Forest".

Shu [21] usó un modelo de regresión logística para predecir el comportamiento de un ciberataque, ante amenazas como correo malicioso, URL maliciosa y distribución de malware. Previamente, aplicaron técnicas de análisis de sentimientos, al flujo de mensajes de Twitter, para determinar la probabilidad de encontrar un ciberataque.

Madisetty [13] enunció un enfoque basado en conjuntos de redes neuronales para la detección de Spam en Twitter. El modelo empleó, a nivel de tweet, características de contenido del usuario y de n-gramas. Después, diferentes arquitecturas de CNN fueron agregadas a una red neuronal, para que finalmente pudieran obtener una detección de Spam. Consiguiendo una medida F de 0.894.

Igualmente, se han realizado investigaciones para identificar información que esté relacionada a diferentes tipos de ciberataques, tomando como fuente los foros de discusión y blogs que residen en la *Deep/Dark* web.

En Grisham [7] crearon un método para identificar mensajes que contienen información o características de distribución de aplicaciones *malware*, y que expone a los usuarios que están esparciendo estas amenazas. En su investigación hacen uso de arquitecturas de Redes Neuronales Recurrentes (o por sus siglas en inglés RNN) y análisis de las conexiones entre usuarios usando grafos. Sus resultados muestran que frecuentemente los archivos esparcidos están en formato .zip, y dirigidos a aplicaciones Android. Demostrando que, la mayoría de usuarios responsables de difundir este contenido, son aquellos que poseen alguna posición administrativa en este tipo de foros.

Finalmente, la tabla 2 muestra una comparación de los resultados obtenidos en cada uno de los trabajos anteriores. Como se puede observar, la medida F1 tiende a subir cuando los ciberataques a detectar son específicos a un tipo. Por otra parte, esta medida incrementó aún más cuando los algoritmos se apoyaron de herramientas del área del procesamiento del lenguaje natural, tal y como fue demostrado en el trabajo de Liao [12]. Pero aún está pendiente, el realizar investigaciones que consideren otros tipos de amenazas como detección de diferentes tipos de ataques de ingeniería social, detección de cuentas dedicas a la difusión de estas amenazas, entre otras.

Tabla 2. Comparación de resultados entre las investigaciones enfocadas a la detección de ciberataques que se difunden en mensajes de redes sociales.

Investigación	**Objetivo**	**Algoritmos**	**Medida F1**
Liao [12]	Indicadores de compromiso	NER/ER, SVM	0.95
Wooi [20]	Phishing	Random Forest	0.95
Madissety [13]	Spam	CNN	0.893
Grisham [7]	Distribución de malware	RNN	0.87
Shu [21]	Malware, URL malicioso, e-mail malicioso	Regresión logística + PCA	0.644

3.3. Interés en vulnerabilidades para uso de *exploits*

Los **exploits** [17] son un tipo de *malware*. Se trata de programas maliciosos que contienen datos o códigos ejecutables que se aprovechan de las vulnerabilidades del software instalado en un equipo local o remoto. Estos han llamado la atención de usuarios maliciosos para comprometer, extraer información e incluso tomar control de los equipos atacados. Dichas fallas de seguridad en el software de infraestructura de cómputo y comunicaciones, hacen que los activos de los usuarios se conviertan en víctimas de estos ciberataques.

Para este tipo de ciberataques existen un par de conceptos que se mencionan a lo largo de esta subsección. El primero es **exploit de Prueba de Concepto** (o del inglés **PoC exploit**), el cual es desarrollado como parte de un proceso para exponer e informar sobre una vulnerabilidad en un software, teniendo como objetivo el demostrar la existencia de este fallo de seguridad en el código del mismo.

El segundo concepto es **exploit de Mundo Real** (o del inglés **real-world exploit**), estos exploits son los que se utilizan para generar ciberataques a equipos de cómputo, redes e infraestructuras de comunicación de las víctimas, aprovechando las fallas de seguridad en software de los equipos para lograr un acto perjudicial al usuario.

La diferencia entre los exploits de *PoC* y de *real-world* es que, los primeros no tienen la intención de generar un daño, más bien son utilizados para alertar al proveedor de su fallo de seguridad en software, y que se tiene que crear una solución. Sin embargo, los de *real-world*, si genera un daño y muchas veces son desarrollados a partir de la información de los exploits *PoC*.

Los siguientes trabajos están enfocados a identificar información que existe en las redes sociales para la detección de exploits que pudieran representar un futuro ciberataque.

En Sabottke [18] crearon una técnica que permite detectar la existencia de posibles exploits del tipo *real-world*, a partir de información disponible en la red social Twitter. El número de características obtenidas para aplicar en un modelo de aprendizaje automático (Maquina de Vectores de Soporte) fueron 38. Entre esas características se consideró el texto y estadísticas de los tweets, un marcador llamado CVSS (de sus siglas en inglés Common Vulnerability Score System) e información de una base de datos de vulnerabilidades NVD (de sus siglas en inglés National Vulnerability Database).

En Mittal [14] crearon un *framework* llamado CyberTwitter, que tiene la función de analizar información contenida en mensajes de Twitter, para hallar las vulnerabilidades que representen una amenaza, y poder alertar al usuario de las mismas. Donde emplearon técnicas del procesamiento del lenguaje natural como NER, manejo de sentencias RDF (de sus siglas en inglés Resource Description Framework) para la representación de conceptos de ciberseguridad a través de ontologías. Creando al final su sistema de alerta con la utilización de reglas SWRL (de sus siglas en inglés Semantic Web Rule Language).

Por otro lado, Chen [3] desarrolló un método para analizar información de los mensajes de Twitter, y poder predecir cuándo las vulnerabilidades serán usadas

en un ciberataque del tipo exploit *real-world* o *PoC*. Empleó un análisis basado en un grafo llamado CVE-autor-Tweet, para descubrir el conjunto de características a utilizar en su modelo de predicción. Al final propone dos modelos, uno para la clasificación llamado FEEU, con una medida F1 promedio de 0.514. Y el otro para la regresión, llamado FRET, siendo este último el que indica cuando será ejecutado el exploit. Según sus resultados, puede predecir el uso de exploits desde que aparece la vulnerabilidad, 37.5 días antes para el exploit *PoC*, y 11.9 días para exploits *real-world*.

Un estudio que consideró información de redes sociales y foros de la *Deep/Dark* web fue realizado por Almukaynizi [2], donde abordaron la probabilidad de usar un exploit como un problema de clasificación binario; positivo si tiende a usarse o negativo en caso contrario. Emplearon un modelo de clasificación *Random Forest*, logrando resultados promedio de 0.57 para *precision*, 0.93 para *recall*, y 0.67 para una medida F1.

Por último, la tabla 3 muestra una comparación de resultados obtenidos por cada uno de los trabajos previos. Como se puede observar, el algoritmo de aprendizaje automático, *Random Forest*, mostró la mejor medida F1. Aunque Mittal [14] no presentó un valor de esta medida, y sólo reportó la cantidad de mensajes correctamente detectados dentro de un corpus de tweets. En ese trabajo, se argumenta que sus resultados fueron mejorando conforme implementaba herramientas del procesamiento del lenguaje natural.

Tabla 3. Comparación de resultados entre las investigaciones enfocadas al interés en las vulnerabilidades para su uso en exploits.

Investigación	Objetivo	Algoritmos	Medida F1
Sabottke [18]	exploits PoC, exploits real-world	SVM	0.45
Chen [3]	exploits PoC, exploits real-world	FEEU, FRET	0.514
Almukaynizi [2]	Vulnerabilidades discutidas	Random Forest	0.67
Mittal [14]	Vulnerabilidades discutidas	NER, SWRL	—

3.4. Ciberataques en las redes sociales

Este apartado hace referencia a trabajos, cuyo objetivo es la detección de ciberataques dirigidos de una cuenta a otra dentro de la misma red social. Por ejemplo, el robo o falsificación de identidad, aumento o pérdida de reputación, control de cuentas para uso en ataques tipo *Botnet*, entre otras nuevas formas de ciberataques adaptadas a estos sitios de comunicación.

En Al-Qurishi [1] crearon un sistema de predicción de ejecución de un ciberataque tipo *Sybil*, analizando información contenida en Twitter. En esta amenaza el atacante puede crear múltiples cuentas falsas o incluso instalar *malware* en el equipo de la víctima, para que posteriormente ese equipo realice *likes* o *reviews*, y el usuario víctima no tenga conocimiento de esto.

Siendo el objetivo del ataque en redes sociales, el aumentar o reducir la reputación de una cuenta de usuario. El trabajo está integrado por tres módulos, que son, la recolección de datos, mecanismo de extracción de características, y un modelo de regresión profunda haciendo uso de Tensorflow para esta tarea. Logrando un resultado del 86 %, tal y como muestra la tabla 4.

Tabla 4. Resultado de la investigación enfocada a ciberataques entre cuentas de la misma red social.

Investigación	**Objetivo**	**Algoritmos**	**Medida F1**
Al-Qurish [1]	Sybil	Modelo de regresión profunda	0.86

Como se puede observar, sólo se ha identificado un trabajo que está relacionado a la detección de un ciberataque, y cuyo objetivo es controlar o afectar la cuenta de un usuario específico desde otra cuenta de la misma red social.

4. Conclusión

Aunque existen diversas amenazas en las redes sociales como el robo de identidad, esparcimiento de noticias falsas, mensajes de ingeniería social para engañar al usuario (fraudes, extorsiones, etc.), entre otras. Este artículo presentó aquellas investigaciones que están orientadas a la detección y predicción de ciberataques que tienen la intención de alterar, controlar, manipular, dañar o afectar los servicios digitales, equipos de cómputo y comunicaciones de las víctimas. Por otra parte, los resultados obtenidos en cada uno de los trabajos mostraron que el uso de algoritmos de aprendizaje automático, aprendizaje profundo y herramientas del procesamiento del lenguaje natural, contribuyen a una mejor detección de este tipo de amenazas. Por último, entre los enfoques mostrados en secciones anteriores, se considera que hay oportunidades de investigación, principalmente en dos de ellos. El primero es para mejorar el valor F1 con relación a la predicción del uso de vulnerabilidades en exploits. Y para el segundo enfoque, se tiene que trabajar en investigaciones con relación a la detección de ciberataques entre cuentas de la misma red social. Igualmente, para futuras investigaciones se tiene que seguir considerando, dentro de las metodologías, el implementar las herramientas más recientes de aprendizaje automático y profundo y del procesamiento del lenguaje natural, con la intención de mejorar la efectividad de resultados obtenidos de las futuras soluciones.

Agradecimientos. Los autores agradecemos a el Instituto Politécnico Nacional, pero en especial a el Centro de Investigación en Computación y a CONACYT, por el apoyo recibido durante la realización de este artículo. Agradecemos a apoyo de proyectos SIP 20200859 y 20200797 y Conacyt A1-S-47854.

Referencias

1. Al-Qurishi, M., Alrubaian, M., Rahman, S.M.M., Alamri, A., Hassan, M.M.: A prediction system of sybil attack in social network using deep-regression model. Future Generation Computer Systems 87, 743–753 (2018), http://www.sciencedirect.com/science/article/pii/S0167739X17300821
2. Almukaynizi, M., Grimm, A., Nunes, E., Shakarian, J., Shakarian, P.: Predicting cyber threats through hacker social networks in darkweb and deepweb forums. In: Proceedings of the 2017 International Conference of The Computational Social Science Society of the Americas. CSS 2017, Association for Computing Machinery, New York, NY, USA (2017), https://doi.org/10.1145/3145574.3145590
3. Chen, H., Liu, R., Park, N., Subrahmanian, V.: Using twitter to predict when vulnerabilities will be exploited. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 3143–3152. KDD '19, Association for Computing Machinery, New York, NY, USA (2019), https://doi.org/10.1145/3292500.3330742
4. Das, S., Behera, R.K., kumar, M., Rath, S.K.: Real-time sentiment analysis of twitter streaming data for stock prediction. Procedia Computer Science 132, 956–964 (2018), http://www.sciencedirect.com/science/article/pii/S1877050918308433, international Conference on Computational Intelligence and Data Science
5. Deliu, I., Leichter, C., Franke, K.: Extracting cyber threat intelligence from hacker forums: Support vector machines versus convolutional neural networks. In: 2017 IEEE International Conference on Big Data (Big Data). pp. 3648–3656 (2017)
6. Goyal, P., Hossain, K.T., Deb, A., Tavabi, N., Bartley, N., Abeliuk, A., Ferrara, E., Lerman, K.: Discovering signals from web sources to predict cyber attacks (2018)
7. Grisham, J., Samtani, S., Patton, M., Chen, H.: Identifying mobile malware and key threat actors in online hacker forums for proactive cyber threat intelligence. In: 2017 IEEE International Conference on Intelligence and Security Informatics (ISI). pp. 13–18 (2017)
8. Hernandez-Suarez, A., Sanchez-Perez, G., Toscano-Medina, K., Martinez-Hernandez, V., Perez-Meana, H., Olivares Mercado, J., Sanchez, V.: Social sentiment sensor in twitter for predicting cyber-attacks using l1 regularization. Sensors 18, 1380 (04 2018)
9. Hernández, A., Sanchez, V., Sánchez, G., Pérez, H., Olivares, J., Toscano, K., Nakano, M., Martinez, V.: Security attack prediction based on user sentiment analysis of twitter data. In: 2016 IEEE International Conference on Industrial Technology (ICIT). pp. 610–617 (2016)
10. Khan, M.A.H., Iwai, M., Sezaki, K.: A robust and scalable framework for detecting self-reported illness from twitter. In: 2012 IEEE 14th International Conference on e-Health Networking, Applications and Services (Healthcom). pp. 303–308 (2012)
11. Khandpur, R.P., Ji, T., Jan, S., Wang, G., Lu, C.T., Ramakrishnan, N.: Crowdsourcing cybersecurity: Cyber attack detection using social media. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. pp. 1049–1057. CIKM '17, Association for Computing Machinery, New York, NY, USA (2017), https://doi.org/10.1145/3132847.3132866
12. Liao, X., Yuan, K., Wang, X., Li, Z., Xing, L., Beyah, R.: Acing the ioc game: Toward automatic discovery and analysis of open-source cyber threat intelligence. In: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. pp. 755–766. CCS '16, Association for Computing Machinery, New York, NY, USA (2016), https://doi.org/10.1145/2976749.2978315

13. Madisetty, S., Desarkar, M.S.: A neural network-based ensemble approach for spam detection in twitter. IEEE Transactions on Computational Social Systems 5(4), 973–984 (2018)
14. Mittal, S., Das, P.K., Mulwad, V., Joshi, A., Finin, T.: Cybertwitter: Using twitter to generate alerts for cybersecurity threats and vulnerabilities. In: 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). pp. 860–867 (2016)
15. Peréz, D.: Iocs, una palabra de moda, un tema caliente. pero, ¿realmente conocemos sus capacidades? https://www.pandasecurity.com/spain/mediacenter/seguridad/iocs-y-sus-capacidades/ (2016)
16. Ritter, A., Wright, E., Casey, W., Mitchell, T.: Weakly supervised extraction of computer security events from twitter. In: Proceedings of the 24th International Conference on World Wide Web. pp. 896–905. WWW '15, International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE (2015), https://doi.org/10.1145/2736277.2741083
17. the Robot, M.: ¿qué son los exploits y por qué son una amenaza? https://www.kaspersky.es/blog/exploits-problem-explanation/6520/ (2015)
18. Sabottke, C., Suciu, O., Dumitraş, T.: Vulnerability disclosure in the age of social media: Exploiting twitter for predicting real-world exploits. In: Proceedings of the 24th USENIX Conference on Security Symposium. pp. 1041–1056. SEC'15, USENIX Association, USA (2015)
19. Saidi, F., Trabelsi, Z., Salah, K., Ghezala, H.B.: Approaches to analyze cyber terrorist communities: Survey and challenges. Computers and Security 66, 66–80 (2017), http://www.sciencedirect.com/science/article/pii/S0167404817300068
20. Seow Wooi, L., Sani, N.F.M., Abdullah, M.T., Yaakob, R., Sharum, M.: An effective security alert mechanism for real-time phishing tweet detection on twitter. Computers & Security 83 (02 2019)
21. Shu, K., Sliva, A., Sampson, J., Liu, H.: Understanding Cyber Attack Behaviors with Sentiment Information on Social Media, pp. 377–388 (06 2018)
22. Urolagin, S.: Text mining of tweet for sentiment classification and association with stock prices. In: 2017 International Conference on Computer and Applications (ICCA). pp. 384–388 (2017)
23. Younus, A., Qureshi, M., Saeed, M., Touheed, N., O'Riordan, C., Pasi, G.: Election trolling: Analyzing sentiment in tweets during Pakistan elections 2013. pp. 411–412 (04 2014)