

---

# *AIDA*<sub>Cyber</sub>: CONTRIBUCIONES EN CIBERSEGURIDAD Y CIBERCRIMEN

---

**David Camacho**

Departamento de Sistemas Informáticos  
Universidad Politécnica de Madrid  
Calle de Alan Turing, s/n, 28031 Madrid, Spain

[David.Camacho@upm.es](mailto:David.Camacho@upm.es)

<http://aida.etsisi.upm.es/>

[https://www.researchgate.net/profile/David\\_Camacho](https://www.researchgate.net/profile/David_Camacho)

<https://scholar.google.com/citations?hl=es&user=fpf6EDAAAAAJ#>

8 de julio de 2020

## RESUMEN

Este informe resume las principales contribuciones realizadas por el grupo de investigación de *Inteligencia artificial aplicada y análisis de datos* (AIDA) en el área de la **Inteligencia Artificial**, las **ciberseguridad** y el **cibercrimen**. Se describen el conjunto de publicaciones, así como sus principales contribuciones al estado del arte, relacionadas con la aplicación de técnicas de Inteligencia Artificial y Ciencia de Datos en el área de la ciberseguridad y del análisis de redes sociales. En concreto, se presentan las contribuciones relativas al análisis de malware, así como la utilización de técnicas de aprendizaje automático y computación basada en grafos para la detección (y prevención) de polarización y radicalización en redes sociales. En este resumen se presentan los principales trabajos publicados en las áreas mencionadas, realizándose además una breve descripción de algunas publicaciones seleccionadas por su relevancia.

**Palabras clave:** Malware · Computer Security · Radicalisation · Polarisation

### Principales publicaciones:

– Alejandro Martín, Héctor D. Menéndez, David Camacho. *MOCDroid: multi-objective evolutionary classifier for Android malware detection*. **Soft Computing**, Vol. 21.24 (2017): 7405-7415.

– Raul Lara-Cabrera, Antonio González-Pardo, K. Benouaret, N. Faci, Djamal Benslimane, David Camacho (2017). *Measuring the radicalisation risk in social networks*. **IEEE Access**, Vol. 5, pp. 10892-10900.

– Alejandro Martín, Raúl Lara-Cabrera, David Camacho. *Android malware detection through hybrid features fusion and ensemble classifiers: the AndroPyTool framework and the OmniDroid dataset*. **Information Fusion**. Vol. 52 (2019): 128-142.

– David Camacho, Ángel Panizo-Lledot, Gema Bello-Orgaz, Antonio González-Pardo, Erik Cambria. *The four dimensions of social network analysis: An overview of research methods, applications, and software tools*. **Information Fusion**. Vol. 63 (2020), pp. 1–33. DOI: 10.1016/j.inffus.2020.05.009

## 1. Introducción

A continuación, se describe el conjunto más significativo y relevante de publicaciones realizado por el grupo de investigación Applied Intelligence & Data Analysis Group (AIDA<sup>1</sup>), adscrito al Departamento de Sistemas Informáticos de la Universidad Politécnica de Madrid en el área del Análisis de Redes Sociales (SNA, del inglés *Social Network Analysis*).

### 1.1. El grupo de investigación

El grupo de investigación tiene su origen en la Universidad Autónoma de Madrid, donde fue creado en 2011 por el Dr. David Camacho. Posteriormente, en septiembre de 2019, el grupo se incorporaría de manera gradual al Departamento de Sistemas Informáticos de la Universidad Politécnica de Madrid, donde actualmente se encuentra desarrollando sus actividades de investigación y desarrollo de proyectos.

El Grupo AIDA está actualmente formado por un total de quince miembros, nueve doctores (tres externos a la Universidad Politécnica de Madrid), cinco estudiantes de doctorado, y un técnico de investigación. El equipo de investigación es un grupo multidisciplinar formado principalmente por Doctores e Ingenieros en Informática (10), Matemáticos (3), Físicos (1), y que cuenta con Psicólogos especializados en Criminología, Pedagogos, y Bioinformáticos, que participan en el desarrollo de diversas líneas de investigación multidisciplinarias.

### 1.2. Las líneas de investigación

Las actuales líneas de investigación, pueden dividirse en dos grandes subconjuntos: las de carácter básico, o fundamental, en el área de las ciencias de la computación, y las de carácter aplicado (más multidisciplinar). En concreto se mencionarán:

#### INVESTIGACIÓN BÁSICA:

- *Desarrollo de algoritmos y sistemas para Ciberseguridad*: aplicación de técnicas de Inteligencia Artificial (IA) y *Machine Learning* para el desarrollo de algoritmos y técnicas aplicables en problemas de detección de malware, estegoanálisis o la detección de ciberdelincuencia en redes sociales. Algunas de las técnicas de IA más relevantes que se han utilizado en este área son:
  - o *Aprendizaje Automático*: orientado al aprendizaje no supervisado (Clustering y Modelos Ocultos de Markov), supervisado (Clasificación, *ensemble learning*, *Boosting*), aplicación de computación evolutiva en clasificación (Neuroevolución) y en modelos de aprendizaje profundo (Redes Convolucionales y Deep Learning).
  - o *Computación Evolutiva*: enfocado en Algoritmos Evolutivos (mono y multi-objetivos) y Programación Genética.
  - o *Computación basada en enjambres*: enfocado al estudio de Algoritmos de Optimización basados en Colonias de Hormigas (ACO).
  - o *Computación basada en grafos*: enfocado al diseño de nuevas métricas y algoritmos de optimización aplicados entre otros a la detección de subgrafos.

#### INVESTIGACIÓN APLICADA

- Desarrollo de herramientas para el análisis y detección de malware.
- Implementación de técnicas de esteganografía y estegoanálisis.
- Detección y medida de radicalización en redes sociales.
- Polarización política.

## 2. AIDA y el análisis de malware

Esta línea de investigación tiene como objetivo principal el desarrollo de métodos para la detección de amenazas y ataques cibernéticos, principalmente aquellos en los que se realizan operaciones no deseadas ni permitidas en el dispositivo de un usuario. Estos métodos se basarán en modelos de Aprendizaje Automático por un lado de Clustering,

<sup>1</sup><http://aida.etsisi.upm.es>

para la búsqueda de patrones que identifiquen estas amenazas, y por otro lado de Clasificación, para determinar, en base a una información que caracterice un comportamiento concreto, si este tiene un carácter benigno o maligno.

El desarrollo de sistemas informáticos ha estado ligado, desde sus orígenes, a la aparición de distintos intentos para romper su seguridad aprovechando sus vulnerabilidades. Internet y la gran expansión de la tecnología, en la que hoy en día la mayoría de las personas en los países desarrollados cuentan con más de un dispositivo conectado a Internet, han hecho que estos ataques hayan aumentado de forma muy significativa. La importancia actual del Ciberespacio para la comunicación entre personas y con la administración pública, la realización de operaciones financieras o, de forma general, la transmisión de información, han hecho de este un campo muy amplio, pero también de muy difícil control. Dada esta importancia, actualmente forma parte de las líneas de actuación de la Estrategia de Seguridad Nacional [1].

El interés por la Ciberseguridad se remarca cuando se cuantifica el número de ataques y cómo avanza año tras año esta cifra. Por ejemplo, desde 2007 hasta 2014, los ataques basados en la web han crecido dramáticamente cerca de un 6.000 %, alcanzando más de 1.400 millones en 2014. Esta cifra muestra la importancia que cobra día tras día la Ciberseguridad y que origina grandes pérdidas económicas. Así, en 2013, el Ciberdelito ocasionó daños por valor de casi 6 millones de dólares únicamente en Estados Unidos [9]. Simultáneamente, se ha producido un aumento en la complejidad de estos ataques y en el tamaño de los mismos. En los últimos años se han utilizado Botnets para controlar ordenadores de forma remota, se han ejecutado ataques distribuidos de denegación de servicio DDoS [6] para detener el correcto funcionamiento de sistemas o se han empleado formas complejas de Malware [26].

Debido a la necesidad de asegurar la integridad, privacidad y seguridad del ciberespacio y, por tanto, de todas las comunicaciones y operaciones que se realizan dentro del marco del mismo, es necesario el estudio y desarrollo de métodos para detener o mitigar, dado el caso, los posibles efectos de los ataques cibernéticos. De forma general, se pretende desarrollar métodos para prevenir, detener y detectar este tipo de ataques dinámicamente, en tiempo real y de forma efectiva. En el desarrollo de estos métodos, jugarán un papel principal técnicas y algoritmos basados en Inteligencia Artificial.

El *Malware*, como definición de aquellos programas con intenciones maliciosas, supone un importante campo de estudio dentro del gran espacio que abarca la Ciberseguridad. A su vez, existen distintos enfoques para la detección de Malware, así como de plataformas susceptibles de ser atacados mediante Malware. En general, son dos los enfoques principales a la hora de afrontar un análisis de programas para determinar su origen benigno o maligno, el análisis estático y el análisis dinámico [8]. Por un lado, el análisis estático está enfocado en el uso de información contenida en el fichero donde reside el programa, sin llegar a ejecutar el software. Por otro lado, el análisis dinámico se centra en ejecutar el software, en un entorno seguro y controlado, y monitorizar las acciones que realiza, con el objetivo de buscar rasgos diferenciadores.

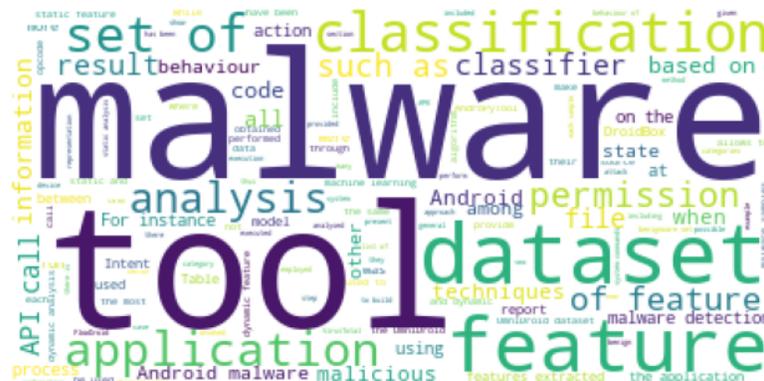


Figura 1: Imagen representativa de términos utilizados en *Android malware detection through hybrid features fusion and ensemble classifiers: the AndroPyTool framework and the OmniDroid dataset* [17].

A continuación, se detalla, en una serie de puntos, las principales contribuciones a estas líneas de investigación y que han supuesto un avance significativo al estado actual del arte en esta área. De forma general, estas contribuciones pueden agruparse en dos grandes líneas de trabajo y que hacen referencia a los dos enfoques, previamente explicados, a la hora de analizar un programa y determinar su naturaleza.

### 1. Análisis estático

- a) Estudio de los procedimientos ya existentes para realizar análisis estático de programas [19, 20].
- b) Estudio de las propiedades estáticas de programas que pueden permitir discernir su naturaleza [20, 15, 21]
- c) Búsqueda de fuentes de datos etiquetados de programas benignos y malignos que permitan la realización de posteriores análisis [17, 16].
- d) Análisis, mediante técnicas de Aprendizaje Automático, de las diferencias entre programas benigno y malignos [17, 18, 19, 12].
- e) Estudio de los métodos de clasificación existentes, su grado de precisión y su posible mejora para obtener métodos de gran precisión capaces de determinar la naturaleza de un programa [20].
- f) Aplicación de técnicas de algoritmos evolutivos para el desarrollo de métodos de clasificación de malware para Android [14, 13].

### 2. Análisis dinámico

- a) Estudio de los procedimientos y los fundamentos del análisis dinámico de programas [23].
- b) Estudio de las plataformas (sistemas operativos y dispositivos) y su importancia a la hora de sufrir y prevenir ataques cibernéticos [15].
- c) Ejecución y análisis de los efectos producidos por la ejecución de software maligno. Concretamente en la comparación con los efectos producidos por software benigno y maligno [17, 23].
- d) Estudio de las vías de aplicación de métodos de aprendizaje automático para el análisis clasificación de Malware [23].
- e) Desarrollo de nuevos métodos capaces de detectar en tiempo real y de forma dinámica Malware que utilice formas actuales de ofuscación.
  - 1) Estudio de los efectos en memoria de la ejecución de programas benignos y malignos [15, 17].
  - 2) Análisis y clasificación de comportamientos de programas en tiempo real según su naturaleza benigna o maligna [22].
  - 3) Desarrollo de mecanismos de detección ágil y en tiempo real de detección de Malware [22].
  - 4) Estudio de los efectos dañinos producidos por Malware y desarrollo de métodos para minimizarlos [15].

Por otro lado el cibercrimen, o la ciberdelincuencia, puede ser considerara como otra de las áreas, o dominios de gran interés de la ciberseguridad.

### 3. AIDA y la radicalización en redes

Una segunda línea de contribuciones relacionadas con la ciberseguridad se circunscriben a un área denominado **ciberdelincuencia**. Dentro de este área se ha trabajado en el análisis y detección de riesgo de polarización y radicalización, principalmente desde una perspectiva de las redes sociales. Las principales contribuciones en el área de la ciberdelincuencia concretamente en el problema de la detección y prevención de la radicalización en redes sociales, así como otros problemas como la difusión de contenidos engañosos (fakenews, bulos) y desinformación se están abordando más recientemente.

Las principales contribuciones en este área se han realizado en:

- La detección de polarización, peligro de radicalización en redes [2, 3, 4, 7, 10, 11, 24, 27, 28]
- La detección de bulos, fake news o desinformación [25]

En este tipo de aplicaciones se emplean los métodos y técnicas de IA descritos en [] para la valoración del nivel de riesgo de radicalización de un individuo en redes, el análisis del discurso político en redes, o la utilización del lenguaje en ciertos dominios.

### 4. Principales publicaciones

Del anterior conjunto de publicaciones se describirán brevemente un subconjunto de las mismas debido al impacto obtenido, y a la relevancia de la investigación en el área del SNA:



- D. Camacho et al. (2020), "The four dimensions of social network analysis: An overview of research methods, applications, and software tools-[5]. Este artículo presenta tres contribuciones principales: 1) una revisión bibliográfica actualizada del estado del arte en el área del análisis de redes sociales; 2) se propone un nuevo conjunto de métricas basadas en cuatro características (o *dimensiones*) fundamentales del SNA; 3) por último, se realiza un análisis cuantitativo de un conjunto de herramientas software populares en SNA. En este artículo se puede encontrar una revisión actualizada del estado del arte en el área de la ciberdelincuencia y cibercrimen en redes sociales. En concreto, el trabajo propone la definición de cuatro **dimensiones** diferentes: *Pattern & Knowledge discovery*, *Information Fusion & Integration*, *Scalability*, y *Visualization*, que se utilizan para definir un conjunto de nuevas métricas (denominadas *grados*, o (*degrees*) con el fin de poder evaluar la madurez de cualquier tecnología software relacionada con el SNA.

## 5. Proyectos de investigación

El grupo desarrolla, o ha desarrollado recientemente, diferentes proyectos de investigación competitivos tanto nacionales como internacionales en este área. En concreto se mencionarán los siguientes proyectos:

- "Tracking tool based on social media for risk assessment on radicalisation (RiskTrack)". Unión Europea (JUST-2015-JCOO-AG-723180). 2016-2018.
- "Cybersecurity, Network Analysis and Monitoring for the Next Generation Internet (CYNAMON)". Comunidad de Madrid (P2018/TCS-4566). 2019-2023.
- "Ciberseguridad: datos, información, riesgos (CIBERDINE)". Comunidad de Madrid (S2013/ICE-3095). 2016-2018.
- "Nuevos Modelos de Cómputo Bioinspirado para Entornos Masivamente Complejos (DeepBio)". Ministerio de Economía, Industria y Competitividad (Excelencia). TIN2017-85727-C4-3-P. 2018-2021
- "Bioinspired Algorithms in Complex Ephemeral Environments (EphemeCH)". Ministerio de Economía, Industria y Competitividad (Excelencia). TIN2014-56494-C4-4-P. 2015-2018

## Biografía

David Camacho es Profesor Titular en el Departamento de Sistemas Informáticos de la Universidad Politécnica de Madrid (España) y dirige el Grupo de Inteligencia Aplicada y Análisis de Datos (AIDA). Recibió su doctorado en Ingeniería Informática por la Universidad Carlos III de Madrid en 2001. Sus intereses de investigación incluyen la inteligencia artificial, el aprendizaje automático, la minería de datos, la computación evolutiva, el análisis de redes sociales, la inteligencia de enjambre, entre otros. Ha publicado más de 300 artículos de investigación (revistas, conferencias, capítulos de libro, etc.), participado en más de 40 proyectos de investigación competitivos (tanto de carácter nacional como internacional), ha impartido más de 50 charlas invitadas y editado decenas de números especiales en revistas y actas de congresos. Actualmente forma parte del consejo editorial de diversas revistas internacionales, entre las que se mencionarán: [Information Fusion](#), [Journal of Ambient Intelligence and Humanized Computing](#), [International Journal of Bio-Inspired Computation](#), [Expert systems](#), o [Evolutionary Intelligence](#), entre otras.

## Referencias

- [1] Félix Arteaga. La estrategia de seguridad nacional 2013. *Comentario Elcano*, 37:2013, 2013.
- [2] Mahmoud Barhamgi, Raúl Lara-Cabrera, Djamel Benslimane, and David Camacho. Ontology uses for radicalisation detection on social networks. In *International Conference on Intelligent Data Engineering and Automated Learning*, pages 3–8. Springer, Cham, 2018.
- [3] Mahmoud Barhamgi, Abir Masmoudi, Raul Lara-Cabrera, and David Camacho. Social networks data analysis with semantics: application to the radicalization problem. *Journal of Ambient Intelligence and Humanized Computing*, pages 1–15, 2018.
- [4] David Camacho, Irene Gilpérez-López, Antonio Gonzalez-Pardo, Alvaro Ortigosa, and Carlota Urruela. Risktrack: a new approach for risk assessment of radicalisation based on social media data. In *CEUR Workshop Proceedings*, 2016.
- [5] David Camacho, Angel Panizo-LLedot, Gema Bello-Orgaz, Antonio Gonzalez-Pardo, and Erik Cambria. The four dimensions of social network analysis: An overview of research methods, applications, and software tools. *Information Fusion*, 63:1–33, 2020.

- [6] Carol Fung, Jie Zhang, Issam Aib, and Raouf Boutaba. Trust management and admission control for host-based collaborative intrusion detection. *Journal of Network and Systems Management*, 19(2):257–277, 2011.
- [7] Irene Gilpérez-López, Javier Torregrosa, Mahmoud Barhamgi, and David Camacho. An initial study on radicalization risk factors: Towards an assessment software tool. In *2017 28th international workshop on database and expert systems applications (DEXA)*, pages 11–16. IEEE, 2017.
- [8] Nwokedi Idika and Aditya P Mathur. A survey of malware detection techniques. *Purdue University*, 48:2007–2, 2007.
- [9] Eric Jardine. Global cyberspace is safer than you think: real trends in cybercrime, 2015.
- [10] Raúl Lara-Cabrera, Antonio Gonzalez-Pardo, and David Camacho. Statistical analysis of risk assessment factors and metrics to evaluate radicalisation in twitter. *Future Generation Computer Systems*, 93:971–978, 2019.
- [11] Raul Lara-Cabrera, Antonio Gonzalez Pardo, Karim Benouaret, Noura Faci, Djamel Benslimane, and David Camacho. Measuring the radicalisation risk in social networks. *IEEE Access*, 5:10892–10900, 2017.
- [12] Alejandro Martín, Alejandro Calleja, Héctor D Menéndez, Juan Tapiador, and David Camacho. Adroit: Android malware detection using meta-information. In *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 1–8. IEEE, 2016.
- [13] Alejandro Martín and David Camacho. Evolving the architecture and hyperparameters of dnns for malware detection. In *Deep Neural Evolution*, pages 357–377. Springer, Singapore, 2020.
- [14] Alejandro Martín, Félix Fuentes-Hurtado, Valery Naranjo, and David Camacho. Evolving deep neural networks architectures for android malware classification. In *2017 IEEE Congress on Evolutionary Computation (CEC)*, pages 1659–1666. IEEE, 2017.
- [15] Alejandro Martín, Julio Hernandez-Castro, and David Camacho. An in-depth study of the jisut family of android ransomware. *IEEE Access*, 6:57205–57218, 2018.
- [16] Alejandro Martín, Raúl Lara-Cabrera, and David Camacho. A new tool for static and dynamic android malware analysis. *Data Science and Knowledge Engineering for Sensing Decision Support*, pages 509–516, 2018.
- [17] Alejandro Martín, Raúl Lara-Cabrera, and David Camacho. Android malware detection through hybrid features fusion and ensemble classifiers: the andropytool framework and the omnidroid dataset. *Information Fusion*, 52:128–142, 2019.
- [18] Alejandro Martín, Héctor D Menéndez, and David Camacho. Genetic boosting classification for malware detection. In *2016 IEEE Congress on Evolutionary Computation (CEC)*, pages 1030–1037. IEEE, 2016.
- [19] Alejandro Martín, Héctor D Menéndez, and David Camacho. String-based malware detection for android environments. In *International Symposium on Intelligent and Distributed Computing*, pages 99–108. Springer, Cham, 2016.
- [20] Alejandro Martín, Héctor D Menéndez, and David Camacho. Studying the influence of static api calls for hiding malware. In *Conference of the Spanish Association for Artificial Intelligence*, pages 363–372. Springer, Cham, 2016.
- [21] Alejandro Martín, Héctor D Menéndez, and David Camacho. Moeandroid: multi-objective evolutionary classifier for android malware detection. *Soft Computing*, 21(24):7405–7415, 2017.
- [22] Alejandro Martín, Víctor Rodríguez-Fernández, and David Camacho. Candyman: Classifying android malware families by modelling dynamic traces with markov chains. *Engineering Applications of Artificial Intelligence*, 74:121–133, 2018.
- [23] Alejandro Martín García. Machine learning techniques for android malware detection and classification. 2019.
- [24] Angel Panizo-LLedot, Javier Torregrosa, Gema Bello-Orgaz, Joshua Thorburn, and David Camacho. Describing alt-right communities and their discourse on twitter during the 2018 us mid-term elections. In *International Conference on Complex Networks and Their Applications*, pages 427–439. Springer, Cham, 2019.
- [25] Marialaura Previti, Victor Rodriguez-Fernandez, David Camacho, Vincenza Carchiolo, and Michele Malgeri. Fake news detection using time series and user features classification. In *International Conference on the Applications of Evolutionary Computation (Part of EvoStar)*, pages 339–353. Springer, Cham, 2020.
- [26] Bruce Schneier. The vulnerabilities market and the future of security. *Forbes*, May, 30, 2012.
- [27] Javier Torregrosa, Irene Gilpérez-López, Raul Lara-Cabrera, David Garriga, and David Camacho. Can an automatic tool assess risk of radicalization online? a case study on facebook. In *2017 European Intelligence and Security Informatics Conference (EISIC)*, pages 165–165. IEEE, 2017.
- [28] Javier Torregrosa, Joshua Thorburn, Raúl Lara-Cabrera, David Camacho, and Humberto M Trujillo. Linguistic analysis of pro-isis users on twitter. *Behavioral Sciences of Terrorism and Political Aggression*, pages 1–15, 2019.