EDITORES:

Manuel A. Serrano - Eduardo Fernández-Medina

Cristina Alcaraz - Noemí de Castro - Guillermo Calvo

Actas de las VI Jornadas Nacionales
(JNIC2021 LIVE)

INVESTIGACIÓN EN ciber segu ridad

Ediciones de la Universidad
de Castilla-La Mancha

Investigación en Ciberseguridad

Actas de las VI Jornadas Nacionales (JNIC2021 LIVE)

Online 9-10 de junio de 2021
Universidad de Castilla-La Mancha

Investigación en Ciberseguridad

Actas de las VI Jornadas Nacionales (JNIC2021 LIVE)

Online 9-10 de junio de 2021
Universidad de Castilla-La Mancha

Editores:
Manuel A. Serrano,
Eduardo Fernández-Medina,
Cristina Alcaraz
Noemí de Castro
Guillermo Calvo

Ediciones de la Universidad de Castilla-La Mancha

Cuenca, 2021

Edita: Ediciones de la Universidad de Castilla-La Mancha

Colección JORNADAS Y CONGRESOS n.º 34

Esta editorial es miembro de la UNE, lo que garantiza la difusión y comercialización de sus publicaciones a nivel nacional e internacional.

I.S.B.N.: 978-84-9044-463-4
D.O.I.: http://doi.org/10.18239/jornadas_2021.34.00

Hecho en España (U.E.) – *Made in Spain (E.U.)*

Bienvenida del Comité Organizador

Tras la parada provocada por la pandemia en 2020, las VI Jornadas Nacionales de Investigación en Ciberseguridad (JNIC) vuelven el 9 y 10 de Junio del 2021 con energías renovadas, y por primera vez en su historia, en un formato 100% online. Esta edición de las JNIC es organizada por los grupos GSyA y Alarcos de la Universidad de Castilla-La Mancha en Ciudad Real, y con la activa colaboración del comité ejecutivo, de los presidentes de los distintos comités de programa y del Instituto Nacional de Ciberseguridad (INCIBE). Continúa de este modo la senda de consolidación de unas jornadas que se celebraron por primera vez en León en 2015 y le siguieron Granada, Madrid, San Sebastián y Cáceres, consecutivamente hasta 2019, y que, en condiciones normales se habrían celebrado en Ciudad Real en 2020.

Estas jornadas se han convertido en un foro de encuentro de los actores más relevantes en el ámbito de la ciberseguridad en España. En ellas, no sólo se presentan algunos de los trabajos científicos punteros en las diversas áreas de ciberseguridad, sino que se presta especial atención a la formación e innovación educativa en materia de ciberseguridad, y también a la conexión con la industria, a través de propuestas de transferencia de tecnología. Tanto es así que, este año se presentan en el Programa de Transferencia algunas modificaciones sobre su funcionamiento y desarrollo que han sido diseñadas con la intención de mejorarlo y hacerlo más valioso para toda la comunidad investigadora en ciberseguridad.

Además de lo anterior, en las JNIC estarán presentes excepcionales ponentes (Soledad Antelada, del Lawrence Berkeley National Laboratory, Ramsés Gallego, de Micro Focus y Mónica Mateos, del Mando Conjunto de Ciberdefensa) mediante tres charlas invitadas y se desarrollarán dos mesas redondas. Éstas contarán con la participación de las organizaciones más relevantes en el panorama industrial, social y de emprendimiento en relación con la ciberseguridad, analizando y debatiendo el papel que está tomando la ciberseguridad en distintos ámbitos relevantes.

En esta edición de JNIC se han establecido tres modalidades de contribuciones de investigación, los clásicos artículos largos de investigación original, los artículos cortos con investigación en un estado más preliminar, y resúmenes extendidos de publicaciones muy relevantes y de alto impacto en materia de ciberseguridad publicados entre los años 2019 y 2021. En el caso de contribuciones de formación e innovación educativa, y también de transferencias se han considerado solamente artículos largos. Se han recibido para su valoración un total de 86

contribuciones organizadas en 26, 27 y 33 artículos largos, cortos y resúmenes ya publicados, de los que los respectivos comités de programa han aceptado 21, 19 y 27, respectivamente. En total se ha contado con una ratio de aceptación del 77%. Estas cifras indican una participación en las jornadas que continúa creciendo, y una madurez del sector español de la ciberseguridad que ya cuenta con un volumen importante de publicaciones de alto impacto.

El formato online de esta edición de las jornadas nos ha motivado a organizar las jornadas de modo más compacto, distinguiendo por primera vez entre actividades plenarias (charlas invitadas, mesas redondas, sesión de formación e innovación educativa, sesión de transferencia de tecnología, junto a inauguración y clausura) y sesiones paralelas de presentación de artículos científicos. En concreto, se han organizado 10 sesiones de presentación de artículos científicos en dos líneas paralelas, sobre las siguientes temáticas: detección de intrusos y gestión de anomalías (I y II), ciberataques e inteligencia de amenazas, análisis forense y cibercrimen, ciberseguridad industrial, inteligencia artificial y ciberseguridad, gobierno y riesgo, tecnologías emergentes y entrenamiento, criptografía, y finalmente privacidad.

En esta edición de las jornadas se han organizado dos números especiales de revistas con elevado factor de impacto para que los artículos científicos mejor valorados por el comité de programa científico puedan enviar versiones extendidas de dichos artículos. Adicionalmente, se han otorgado premios al mejor artículo en cada una de las categorías. En el marco de las JNIC también hemos contado con la participación de la Red de Excelencia Nacional de Investigación en Ciberseguridad (RENIC), impulsando la ciberseguridad a través de la entrega de los premios al *Mejor Trabajo Fin de Máster en Ciberseguridad* y a la *Mejor Tesis Doctoral en Ciberseguridad*. También se ha querido acercar a los jóvenes talentos en ciberseguridad a las JNIC, a través de un CTF (Capture The Flag) organizado por la Universidad de Extremadura y patrocinado por Viewnext.

Desde el equipo que hemos organizado las JNIC2021 queremos agradecer a todas aquellas personas y entidades que han hecho posible su celebración, comenzando por los autores de los distintos trabajos enviados y los asistentes a las jornadas, los tres ponentes invitados, las personas y organizaciones que han participado en las dos mesas redondas, los integrantes de los distintos comités de programa por sus interesantes comentarios en los procesos de revisión y por su colaboración durante las fases de discusión y debate interno, los presidentes de las sesiones, la Universidad de Extremadura por organizar el CTF y la empresa Viewnext por patrocinarlo, los técnicos del área TIC de la UCLM por el apoyo con la plataforma de comunicación, los voluntarios de la UCLM y al resto de organizaciones y entidades patrocinadoras, entre las que se encuentra la Escuela Superior de Informática, el Departamento de Tecnologías y Sistemas de Información y el Instituto de Tecnologías y Sistemas de Información, todos ellos de la Universidad de Castilla-La Mancha, la red RENIC, las cátedras (Telefónica e Indra) y aulas (Avanttic y Alpinia) de la Escuela Superior de Informática, la empresa Cojali, y muy especialmente por su apoyo y contribución al propio INCIBE.

Manuel A. Serrano, Eduardo Fernández-Medina
Presidentes del Comité Organizador
Cristina Alcaraz
Presidenta del Comité de Programa Científico
Noemí de Castro
Presidenta del Comité de Programa de Formación e Innovación Educativa
Guillermo Calvo Flores
Presidente del Comité de Transferencia Tecnológica

Índice General

Comité Ejecutivo 11
Comité Organizador 12
Comité de Programa Científico 13
Comité de Programa de Formación e Innovación Educativa 15
Comité de Transferencia Tecnológica 17

Comunicaciones

Sesión de Investigación A1: Detección de intrusiones y gestión de anomalías I 21
Sesión de Investigación A2: Detección de intrusiones y gestión de anomalías II 55
Sesión de Investigación A3: Ciberataques e inteligencia de amenazas 91
Sesión de Investigación A4: Análisis forense y cibercrimen 107
Sesión de Investigación A5: Ciberseguridad industrial y aplicaciones 133
Sesión de Investigación B1: Inteligencia Artificial en ciberseguridad 157
Sesión de Investigación B2: Gobierno y gestión de riesgos 187
Sesión de Investigación B3: Tecnologías emergentes y entrenamiento en ciberseguridad 215
Sesión de Investigación B4: Criptografía 235
Sesión de Investigación B5: Privacidad 263
Sesión de Transferencia Tecnológica 291
Sesión de Formación e Innovación Educativa 301

Premios RENIC 343

Patrocinadores 349

Comité Ejecutivo

Juan Díez González	INCIBE
Luis Javier García Villalba	Universidad de Complutense de Madrid
Eduardo Fernández-Medina Patón	Universidad de Castilla-La Mancha
Guillermo Suárez-Tangil	IMDEA Networks Institute
Andrés Caro Lindo	Universidad de Extremadura
Pedro García Teodoro	Universidad de Granada. Representante de red RENIC
Noemí de Castro García	Universidad de León
Rafael María Estepa Alonso	Universidad de Sevilla
Pedro Peris López	Universidad Carlos III de Madrid

Comité Organizador

Presidentes del Comité Organizador

Eduardo Fernández-Medina Patón	Universidad de Castilla-la Mancha
Manuel Ángel Serrano Martín	Universidad de Castilla-la Mancha

Finanzas

David García Rosado	Universidad de Castilla-la Mancha
Luis Enrique Sánchez Crespo	Universidad de Castilla-la Mancha

Actas

Antonio Santos-Olmo Parra	Universidad de Castilla-la Mancha

Difusión

Julio Moreno García-Nieto	Universidad de Castilla-la Mancha
José Antonio Cruz Lemus	Universidad de Castilla-la Mancha
María A Moraga de la Rubia	Universidad de Castilla-la Mancha

Webmaster

Aurelio José Horneros Cano	Universidad de Castilla-la Mancha

Logística y Organización

Ignacio García-Rodriguez de Guzmán	Universidad de Castilla-la Mancha
Ismael Caballero Muñoz-Reja	Universidad de Castilla-la Mancha
Gregoria Romero Grande	Universidad de Castilla-la Mancha
Natalia Sanchez Pinilla	Universidad de Castilla-la Mancha

Comité de Programa Científico

Presidenta

Cristina Alcaraz Tello	Universidad de Málaga

Miembros

Aitana Alonso Nogueira	INCIBE
Marcos Arjona Fernández	ElevenPaths
Ana Ayerbe Fernández-Cuesta	Tecnalia
Marta Beltrán Pardo	Universidad Rey Juan Carlos
Carlos Blanco Bueno	Universidad de Cantabria
Jorge Blasco Alís	Royal Holloway, University of London
Pino Caballero-Gil	Universidad de La Laguna
Andrés Caro Lindo	Universidad de Extremadura
Jordi Castellà Roca	Universitat Rovira i Virgili
José M. de Fuentes García-Romero de Tejada	Universidad Carlos III de Madrid
Jesús Esteban Díaz Verdejo	Universidad de Granada
Josep Lluis Ferrer Gomila	Universitat de les Illes Balears
Dario Fiore	IMDEA Software Institute
David García Rosado	Universidad de Castilla-La Mancha
Pedro García Teodoro	Universidad de Granada
Luis Javier García Villalba	Universidad Complutense de Madrid
Iñaki Garitano Garitano	Mondragon Unibertsitatea
Félix Gómez Mármol	Universidad de Murcia
Lorena González Manzano	Universidad Carlos III de Madrid
María Isabel González Vasco	Universidad Rey Juan Carlos I
Julio César Hernández Castro	University of Kent
Luis Hernández Encinas	CSIC
Jorge López Hernández-Ardieta	Banco Santander
Javier López Muñoz	Universidad de Málaga
Rafael Martínez Gasca	Universidad de Sevilla
Gregorio Martínez Pérez	Universidad de Murcia

David Megías Jiménez	Universitat Oberta de Cataluña
Luis Panizo Alonso	Universidad de León
Fernando Pérez González	Universidad de Vigo
Aljosa Pasic	ATOS
Ricardo J. Rodríguez	Universidad de Zaragoza
Fernando Román Muñoz	Universidad Complutense de Madrid
Luis Enrique Sánchez Crespo	Universidad de Castilla-La Mancha
José Soler	Technical University of Denmark-DTU
Miguel Soriano Ibáñez	Universidad Politécnica de Cataluña
Victor A. Villagrá González	Universidad Politécnica de Madrid
Urko Zurutuza Ortega	Mondragon Unibertsitatea
Lilian Adkinson Orellana	Gradiant
Juan Hernández Serrano	Universitat Politécnica de Cataluña

Comité de Programa de Formación e Innovación Educativa

Presidenta

Noemí De Castro García	Universidad de León

Miembros

Adriana Suárez Corona	Universidad de León
Raquel Poy Castro	Universidad de León
José Carlos Sancho Núñez	Universidad de Extremadura
Isaac Agudo Ruiz	Universidad de Málaga
Ana Isabel González-Tablas Ferreres	Universidad Carlos III de Madrid
Xavier Larriva	Universidad Politécnica de Madrid
Ana Lucila Sandoval Orozco	Universidad Complutense de Madrid
Lorena González Manzano	Universidad Carlos III de Madrid
María Isabel González Vasco	Universidad Rey Juan Carlos
David García Rosado	Universidad de Castilla - La Mancha
Sara García Bécares	INCIBE

Comité de Transferencia Tecnológica

Presidente

Guillermo Calvo Flores	INCIBE

Miembros

José Luis González Sánchez	COMPUTAEX
Marcos Arjona Fernández	ElevenPaths
Victor Villagrá González	Universidad Politécnica de Madrid
Luis Enrique Sánchez Crespo	Universidad de Castilla – La Mancha

Sesión de Investigación A4: Análisis forense y cibercrimen
http://doi.org/10.18239/jornadas_2021.34.29

A Review of "Bringing Order to Approximate Matching: Classification and Attacks on Similarity Digest Algorithms"

Miguel Martín-Pérez
Universidad de Zaragoza, Spain
miguelmartinperez@unizar.es

Ricardo J. Rodríguez
Universidad de Zaragoza, Spain
rjrodriguez@unizar.es

Frank Breitinger
University of Liechtenstein, Liechtenstein
frank.breitinger@uni.li

***Abstract*—Fuzzy hashing or similarity hashing (a.k.a. bytewise approximate matching) converts digital artifacts into an intermediate representation to allow for efficient (fast) identification of similar objects, e.g., for deny-listing. Over the past decade, new algorithms have been developed and released to the digital forensics community. When releasing algorithms (e.g., as part of a scientific article), they are frequently compared with other algorithms to outline the benefits and sometimes also the weaknesses of the proposed approach. However, given the wide variety of algorithms and approaches, it is impossible to provide direct comparisons with all existing algorithms. In this paper, we present the first classification of approximate matching algorithms which allows for an easier description and comparisons. Therefore, we first reviewed existing literature to understand the techniques various algorithms use and to familiarize ourselves with the common terminology. Our findings allowed us to develop a categorization relying heavily on the terminology proposed by NIST SP 800-168. In addition to the categorization, this paper presents an abstract set of attacks against algorithms and why they are feasible. Lastly, we detail the characteristics needed to build robust algorithms to prevent attacks. We believe that this paper helps newcomers, practitioners, and experts alike to better compare algorithms, understand their potential, as well as characteristics and implications they may have on forensic investigations.**

***Index Terms*—Similarity digest algorithm, Approximate matching, Fuzzy hashing, Similarity hashing, Bytewise, Classification scheme**

Tipo de contribución: *Investigación ya publicada en* "Bringing Order to Approximate Matching: Classification and Attacks on Similarity Digest Algorithms," Forensic Science International: Digital Investigation, 2021 [6].

I. Extended abstract

According to NIST SP 800-168, "approximate matching is a promising technology designed to identify similarities between two digital artifacts" [1]. This identification of similarities between two or more artifacts can happen on three different levels of abstraction: *bytewise*, when the comparison relies on the raw sequence of bytes that form the digital artifacts; *syntactic*, when the internal structures of the digital artifacts under analysis are used instead of merely byte sequences; or *semantic*, when the comparison relies on contextual attributes to interpret the digital artifacts and estimate their similarity. Furthermore, algorithms may either compare artifacts directly (e.g., Levenshtein distance or Hamming distance), or they may first convert them into an intermediate representation (e.g., a fingerprint, hash, digest) that can then be compared. This latter case is often referred to as fuzzy hashing or similarity hashing and aims at complementing cryptographic hash functions by allowing for the identification of *similar* objects instead of *completely identical* objects.

In this paper we focus on algorithms/literature that operate on the byte-level[1] and utilize an intermediate representation, i.e., a digest/fingerprint. We define these kinds of algorithms as *similarity digest algorithms* (SDA)[2]. These algorithms gained popularity around 2006 when `ssdeep` was published [2]. Over the years, many more algorithms have been proposed such as `sdhash` [3], `mrsh-v2` [4] or `TLSH` [5], to name a few.

In order to compare algorithms, the community mostly focuses on obvious metrics such as runtime efficiency or precision and recall rates. However, due to the various design decisions researchers and practitioners have made during the development, we argue that a finer granular comparison is necessary as there may be instances where precision and recall are insufficient. For instance, some implementations have difficulties handling extremely small files, while others are susceptible if the difference in file size between two objects is too large (e.g., 5 MiB vs. 5 GiB). Consequently, this paper has the following contributions:

- The first categorization for SDA, allowing the community to better discuss and compare the various existing algorithms. Categorizations are useful for scientific fields, as they allow structuring a domain.
- A comprehensive discussion of the algorithms with respect to the categorization and its implication for practitioners.
- A discussion of the categorization with respect to why these characteristics are important and how practitioners may contribute from it, describing an abstract set of attacks.
- In addition, we also provide insights on the desirable properties to build a robust SDA against attacks.

In order to develop the classification, we have identified the six phases of a SDA. Five of these phases can be grouped in an "artifact processing and digest generation phase", while the other is devoted to digest comparison. Each phase consists of various dimensions and procedures

[1] Inputs/artifacts are treated as a byte stream and are processed without any interpretation of the data.

[2] In this paper, we use the SDA interchangeably as a singular and plural acronym.

Algorithm	Feature generation				Feature Processing		Feature Selection		
	Length	*Support Function*	*Intersection*	*Cardinality*	*Mapping Function*	*Bit Reduction*	*Selection Function*	*Domain*	*Coverage*
dcfldd	Static (512)	None	No	Variable ($L/512$)	Hashing	None (128)	None	(n/a)	Full
Nilsimsa	Static (3)	None	Yes	Variable ($6L$)	Hashing	None (8)	None	(n/a)	Full
ssdeep	Dynamic ($L/64$)	Trigger function	No	Fixed (64)	Hashing	Ratio (6/32)	None	(n/a)	Full
md5bloom	Static (512)	None	No	Variable ($L/512$)	Hashing	Ratio (40/128)	None	(n/a)	Full
MRS hash	Dynamic (234)	Trigger function	No	Variable ($L/234$)	Hashing	Ratio (44/128)	None	(n/a)	Full
SimHash	Static (1)	None	Yes	Variable ($8L$)	Identifier	None (8)	Block matching	Feature	Partial
sdhash	Static (64)	None	Yes	Variable (L)	Hashing	Ratio (55/160)	Minimum probability	Feature	Partial
MRSH-V2	Dynamic (320)	Trigger function	No	Variable ($L/320$)	Hashing	Ratio (55/64)	None	(n/a)	Full
mvHash-B	Static ($20, 50$)	None	Yes	Variable ($8L$)	Encoding	Ratio (1/32)	Block similarity	Feature	Full
TLSH	Static (3)	None	Yes	Variable ($6L$)	Hashing	None (8)	None	(n/a)	Full
saHash	Static (1)	None	Yes	Variable ($4L$)	None	None (8)	None	(n/a)	Full
LZJD	Dynamic ($1 + \log_{256} L$)	Unique	No	Variable ($L/(1 + \log_{256} L)$)	Hashing	None (128)	Minimum value	Processed feature	Partial
FbHash	Static (7)	None	Yes	Variable (L)	Hashing	None (64)	None	(n/a)	Full

Table I
CLASSIFICATION OF SIMILARITY DIGEST ALGORITHMS ACCORDING TO OUR PROPOSED CLASSIFICATION SCHEME (*feature generation*, *feature processing*, AND *feature selection* PHASES).

Algorithm	Digest generation				Feature Deduplication		Digest comparison			
	Digest Size	*Storing Structure*	*Order*	*Requirements*	*Type*	*Occurrence*	*Requirements*	*Output Score*	*Score Trend*	*Space Sensitivity*
dcfldd	Input dependent	Processed feature concatenation	Absolute	None	None	(n/a)	None	Interval	Ascending	Total
Nilsimsa	Fixed	Counter	Processed feature-aware	None	Consecutive	Comparison	Minimum commonality	Interval	Ascending	None
ssdeep	Input dependent with max	Processed feature concatenation	Absolute	Minimum features	Consecutive	Comparison	Minimum commonality, Similar input size	Interval	Ascending	Total
md5bloom	Input dependent	Set concatenation	Set-absolute	None	In-Scope	Generation	None	Interval	Ascending	Partial
MRS hash	Input dependent	Set concatenation	Set-absolute	None	In-Scope	Generation	None	Interval	Ascending	Partial
SimHash	Fixed	Counter	None	None	None	(n/a)	Similar input size	Half-bounded	Descending	None
sdhash	Input dependent	Set concatenation	Set-absolute	Diversity	In-Scope	Generation	Minimum amount	Interval	Ascending	Partial
MRSH-V2	Input dependent	Set concatenation	Set-absolute	None	In-Scope	Generation	Minimum amount	Interval	Ascending	Partial
mvHash-B	Input dependent	Set concatenation	Set-absolute	Diversity	In-Scope	Generation	Similar input size	Interval	Ascending	Partial
TLSH	Fixed	Counter	Processed feature-aware	None	None	(n/a)	None	Half-bounded	Descending	None
saHash	Fixed	Counter	Processed feature-aware	None	None	(n/a)	None	Binary value	(n/a)	Total
LZJD	Fixed	Set	None	None	None	(n/a)	None	Interval	Ascending	None
FbHash	Fixed	Counter	Processed feature-aware	Document frequency	None	(n/a)	None	Interval	Ascending	None

Table II
CLASSIFICATION OF SIMILARITY DIGEST ALGORITHMS ACCORDING TO OUR PROPOSED CLASSIFICATION SCHEME (*digest generation*, *feature deduplication*, AND *digest comparison* PHASES).

which themselves are based on characteristics (i.e., its values). For instance, the *feature generation phase* has, among other dimensions/procedures, *length*, *support function* and *intersection*. Every dimension can have different characteristics, such as *static* or *dynamic* (length) and *trigger function* or *unique* (support function). The classification of SDA according to our proposed classification scheme is given in Tables I and II.

Regarding the set of attacks against SDA, we have distinguished between two types of attacks:

- **Attacks against the similarity score**, which divides into *Reduction of Similarity* attacks (the input is crafted to minimize the similarity score when it is compared against other input) and *Emulation of Similarity* attacks (the artifact is manipulated to yield a high similarity score to a another [non-similar] artifact).
- **Attacks against impeding the last phases of an SDA**, which splits into *Impeding the Digest Generation Phase* attack (a deliberate modification of the input in such a way that the SDA is unable to generate a similarity digest due to insufficient conditions) and *Impeding the Digest Comparison Phase* attack (an adversary crafts an input so that the similarity digest generated cannot be compared or the similarity score computed is always very low – i.e., no similar).

As the last contribution of the paper, we have highlighted the properties needed to build a robust SDA against these attacks. For the sake of space, we have deliberately omitted a more detailed description of these properties in this paper. Further details are given in [6].

The full version of this paper (with a full description of the classification dimensions, as well as a detailed explanation on attacks and on building a robust SDA) was published in [6].

ACKNOWLEDGEMENTS

The research was supported in part by the Spanish Ministry of Science, Innovation and Universities under grant MEDRESE-RTI2018-098543-B-I00 and by the University, Industry and Innovation Department of the Aragonese Government under *Programa de Proyectos Estratégicos de Grupos de Investigación* (DisCo research group, ref. T21-20R). It was also supported by the Spanish National Cybersecurity Institute (INCIBE) "Ayudas para la excelencia de los equipos de investigación avanzada en ciberseguridad", grant numbers INCIBEC-2015-02486 and INCIBEI-2015-27300.

REFERENCES

[1] F. Breitinger, B. Guttman, M. McCarrin, V. Roussev, and D. White, "Approximate Matching: Definition and Terminology," National Institute of Standards and Technology, techreport NIST Special Publication 800-168, May 2014.

[2] J. Kornblum, "Identifying almost identical files using context triggered piecewise hashing," *Digital Investigation*, vol. 3, pp. 91–97, 2006, the Proceedings of the 6th Annual Digital Forensic Research Workshop (DFRWS '06).

[3] V. Roussev, "Data Fingerprinting with Similarity Digests," in *Advances in Digital Forensics VI*, K.-P. Chow and S. Shenoi, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 207–226.

[4] F. Breitinger and H. Baier, "Similarity Preserving Hashing: Eligible Properties and a New Algorithm MRSH-v2," in *Digital Forensics and Cyber Crime*, M. Rogers and K. C. Seigfried-Spellar, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 167–182.

[5] J. Oliver, C. Cheng, and Y. Chen, "TLSH – A Locality Sensitive Hash," in *2013 Fourth Cybercrime and Trustworthy Computing Workshop*. IEEE, 2013, pp. 7–13.

[6] M. Martín-Pérez, R. J. Rodríguez, and F. Breitinger, "Bringing Order to Approximate Matching: Classification and Attacks on Similarity Digest Algorithms," *Forensic Science International: Digital Investigation*, 2021.