INTERNET 2026

The Eighteenth International Conference on Evolving Internet

ISBN: 978-1-68558-349-1

March 8th –12th, 2026

Valencia, Spain

INTERNET 2026 Editors

Eugen Borcoci, National University of Science and Technology POLITEHNICA Bucharest, Romania

INTERNET 2026

Forward

The Eighteenth International Conference on Evolving Internet (INTERNET 2026), held between March 8-th, 2026 and March 12-th, 2026 in Valencia, Spain, continued a series of events addressing challenges raised by the evolving Internet, making use of the progress in different advanced mechanisms and theoretical foundations. The gap analysis aims at mechanisms and features concerning the Internet itself, as well as special applications for software defined radio networks, wireless networks, sensor networks, or Internet data streaming and mining.

Originally designed in the spirit of interchange between scientists, the Internet reached a status where large-scale technical limitations impose rethinking its fundamentals. This refers to design aspects (flexibility, scalability, etc.), technical aspects (networking, routing, traffic, address limitation, etc.), as well as economics (new business models, cost sharing, ownership, etc.). The evolving Internet poses architectural, design, and deployment challenges in terms of performance prediction, monitoring and control, admission control, extendibility, stability, resilience, delay-tolerance, and interworking with the existing infrastructures or with specialized networks.

We take here the opportunity to warmly thank all the members of the INTERNET 2026 technical program committee, as well as all the reviewers. The creation of such a high-quality conference program would not have been possible without their involvement. We also kindly thank all the authors who dedicated much of their time and effort to contribute to INTERNET 2026. We truly believe that, thanks to all these efforts, the final conference program consisted of top-quality contributions. We also thank the members of the INTERNET 2026 organizing committee for their help in handling the logistics of this event.

We hope that INTERNET 2026 was a successful international forum for the exchange of ideas and results between academia and industry for the promotion of progress in the field of the evolving Internet.

INTERNET 2026 Chairs

INTERNET 2026 Steering Committee
Renwei (Richard) Li, Southeast University, China
Eugen Borcoci, National University of Science and Technology POLITEHNICA Bucharest, Romania
Terje Jensen, Telenor, Norway
Przemyslaw (Przemek) Pochec, University of New Brunswick, Canada
Parimala Thulasiraman, University of Manitoba – Winnipeg, Canada
Dirceu Cavendish, Kyushu Institute of Technology, Japan
Sonia Ben Rejeb, ISI/University El Manar, Tunisia

INTERNET 2026 Publicity Chairs
Francisco Javier Díaz Blasco, Universitat Politècnica de València, Spain
Ali Ahmad, Universitat Politècnica de València, Spain

José Miguel Jiménez, Universitat Politècnica de València, Spain
Sandra Viciano Tudela, Universitat Politècnica de València, Spain

INTERNET 2026
Committee

INTERNET 2026 Steering Committee

Renwei (Richard) Li, Southeast University, China
Eugen Borcoci, National University of Science and Technology POLITEHNICA Bucharest, Romania
Terje Jensen, Telenor, Norway
Przemyslaw (Przemek) Pochec, University of New Brunswick, Canada
Parimala Thulasiraman, University of Manitoba – Winnipeg, Canada
Dirceu Cavendish, Kyushu Institute of Technology, Japan
Sonia Ben Rejeb, ISI/University El Manar, Tunisia

INTERNET 2026 Publicity Chairs

Francisco Javier Díaz Blasco, Universitat Politècnica de València, Spain
Ali Ahmad, Universitat Politècnica de València, Spain
José Miguel Jiménez, Universitat Politècnica de València, Spain
Sandra Viciano Tudela, Universitat Politècnica de València, Spain

INTERNET 2026 Technical Program Committee

Majed Alowaidi, Majmaah University, Saudi Arabia
Mohammad Alsulami, University of Connecticut, USA
Mário Antunes, Polytechnic of Leiria & INESC-TEC, Portugal
Andrés Arcia-Moret, Xilinx, Cambridge, UK
Marcin Bajer, ABB Corporate Research Center Krakow, Poland
Michail J. Beliatis, Research Centre for Digital Business Development | Aarhus University, Denmark
Laura Belli, University of Parma, Italy
Sonia Ben Rejeb, ISI/University El Manar, Tunisia
Driss Benhaddou, University of Houston, USA
Nik Bessis, Edge Hill University, UK
Maumita Bhattacharya, Charles Sturt University, Australia
Filippo Bianchini, Studio Legale Bianchini, Perugia, Italy
Eugen Borcoci, National University of Science and Technology POLITEHNICA Bucharest, Romania
Fernando Boronat Seguí, Universitat Politecnica de Valencia-Campus De Gandia, Spain
Christos Bouras, University of Patras, Greece
Lianjie Cao**,** Hewlett Packard Labs, USA
Dirceu Cavendish, Kyushu Institute of Technology, Japan
Hao Che, University of Texas at Arlington, USA
Albert M. K. Cheng, University of Houston, USA
Hongmei Chi, Florida A&M University, USA
Andrzej Chydzinski, Silesian University of Technology, Poland
Franco Cicirelli, ICAR-CNR, Italy
Victor Cionca, Munster Technical University, Ireland
Fábio M. Costa, Institute of Informatics (INF) | Federal University of Goiás (UFG), Brazil

Vittorio Curri, Politecnico di Torino, Italy
Monireh Dabaghchian, Morgan State University, USA
Luca Davoli, University of Parma, Italy
Noel De Palma, University Grenoble Alpes, France
Rubens de Souza Matos Junior, Instituto Federal de Sergipe, Brazil
Angel P. del Pobil, Jaume I University, Spain
Jun Duan, IBM T. J. Watson Research Center, USA
Said El Kafhali, Hassan 1st University, Settat, Morocco
Khalid Elbaamrani, Cadi Ayyad University, Marrakech, Morocco
Mohamed Elhadad, Abu Dhabi University, UAE
Ahmed Elwhishi, University of Doha for Science and Technology, Qatar
Hasan Farsijani, Shahid Beheshti University, Iran
Zongming Fei, University of Kentucky, USA
Felix Fischer, Hochschule Mittweida | University of Applied Sciences, Germany
Steffen Fries, Siemens AG, Germany
Song Fu, University of North Texas, USA
Marco Furini, University of Modena and Reggio Emilia, Italy
Antonino Galletta, University of Messina, Italy
Dimitrios Georgakopoulos, Swinburne University of Technology, Australia
Shadan Golestan, University of Alberta, Canada
Victor Govindaswamy, Concordia University Chicago, USA
Shuyang Gu, Texas A & M University-Central Texas, USA
Abdelhay Haqiq, Information Sciences School in Rabat, Morocco
Wladyslaw Homenda, Warsaw University of Technology, Poland
Fu-Hau Hsu, National Central University, Taiwan
Pengfei Hu, VMWare Inc, USA
Takeshi Ikenaga, Kyushu Institute of Technology, Japan
Oliver L. Iliev, FON University, Republic of Macedonia
Khondkar R. Islam, George Mason University, USA
Terje Jensen, Telenor, Norway
Sokratis K. Katsikas, Norwegian University of Science and Technology, Norway
Brian Kelley, University of Texas at San Antonio / 5G Program Management Office - JBSA 5G NextGen, USA
Ahmed Khaled, Northeastern Illinois University, USA
Rasool Kiani, University of Isfahan, Iran
Koteswararao Kondepu, Indian Institute of Technology Dharwad, India
Kishori Mohan Konwar, MIT / Broad Institute of MIT and Harvard, USA
Hovannes Kulhandjian, California State University, Fresno, USA
Mikel Larrea, University of the Basque Country UPV/EHU, Spain
Frédéric Le Mouël, INSA Lyon / University of Lyon, France
Kevin Lee, Deakin University, Australia
Kin K. Leung, Imperial College London, UK
Renwei (Richard) Li, Southeast University, China
Xin Li, Google, USA
Zhijing Li, Facebook, USA
Jinwei Liu, Florida A&M University, USA
Luís Miguel Lopes de Oliveira, Institute Polytechnic of Tomar, Portugal
Imad Mahgoub, Florida Atlantic University, USA

Zoubir Mammeri, IRIT - Paul Sabatier University, France
Philippe Merle, Inria Lille - Nord Europe, France
Jared Onyango Oluoch, University of Toledo, USA
Carlos Enrique Palau Salvador, Universitat Politecnica de Valencia, Spain
Fidel Paniagua Diez, Universidad Internacional de La Rioja - UNIR, Spain
Przemyslaw (Przemek) Pochec, University of New Brunswick, Canada
Mirko Presser, Aarhus University, Denmark
Shiyin Qin, Beihang University, China
Marek Reformat, University of Alberta, Canada
Domenico Rotondi, Grifo Multimedia Srl, Italy
Hooman Samani, University of Plymouth, UK
Sandeep Singh Sandha, University of California-Los Angeles, USA
José Santa, Technical University of Cartagena, Spain
Meghana N. Satpute, University of Texas at Dallas, USA
Irida Shallari, Mid Sweden University, Sweden
Mukesh Singhal, University of California, Merced, USA
Brad Smith, University of California Santa Cruz, USA
Francesco Betti Sorbelli, University of Perugia, Italy
Pedro Sousa, University of Minho, Portugal
Grażyna Suchacka, University of Opole, Poland
Álvaro Suárez Sarmiento, Universidad de Las Palmas de Gran Canaria, Spain
Diego Suárez Touceda, Universidad Internacional de La Rioja - UNIR, Spain
Bedir Tekinerdogan, Wageningen University & Research, The Netherlands
Parimala Thulasiraman, University of Manitoba - Winnipeg, Canada
Homero Toral Cruz, University of Quintana Roo (UQROO), Mexico
Mudasser F. Wyne, National University, USA
Ali Yahyaouy, Faculty of Sciences Dhar El Mahraz, Fez, Morocco
Ping Yang, State University of New York at Binghamton, USA
Zhicheng Yang, PingAn Tech - US Research Lab, USA
Ali Yavari, Swinburne University of Technology, Australia
Habib Zaidi, Geneva University Hospital, Switzerland
Huanle Zhang, University of California, Davis, USA

Copyright Information

For your reference, this is the text governing the copyright release for material published by IARIA.

The copyright release is a transfer of publication rights, which allows IARIA and its partners to drive the dissemination of the published material. This allows IARIA to give articles increased visibility via distribution, inclusion in libraries, and arrangements for submission to indexes.

I, the undersigned, declare that the article is original, and that I represent the authors of this article in the copyright release matters. If this work has been done as work-for-hire, I have obtained all necessary clearances to execute a copyright release. I hereby irrevocably transfer exclusive copyright for this material to IARIA. I give IARIA permission or reproduce the work in any media format such as, but not limited to, print, digital, or electronic. I give IARIA permission to distribute the materials without restriction to any institutions or individuals. I give IARIA permission to submit the work for inclusion in article repositories as IARIA sees fit.

I, the undersigned, declare that to the best of my knowledge, the article is does not contain libelous or otherwise unlawful contents or invading the right of privacy or infringing on a proprietary right.

Following the copyright release, any circulated version of the article must bear the copyright notice and any header and footer information that IARIA applies to the published article.

IARIA grants royalty-free permission to the authors to disseminate the work, under the above provisions, for any academic, commercial, or industrial use. IARIA grants royalty-free permission to any individuals or institutions to make the article available electronically, online, or in print.

IARIA acknowledges that rights to any algorithm, process, procedure, apparatus, or articles of manufacture remain with the authors and their employers.

I, the undersigned, understand that IARIA will not be liable, in contract, tort (including, without limitation, negligence), pre-contract or other representations (other than fraudulent misrepresentations) or otherwise in connection with the publication of my work.

Exception to the above is made for work-for-hire performed while employed by the government. In that case, copyright to the material remains with the said government. The rightful owners (authors and government entity) grant unlimited and unrestricted permission to IARIA, IARIA's contractors, and IARIA's partners to further distribute the work.

Table of Contents

Medical IoT Devices and Systems: Challenges and Opportunities 1
Dirceu Cavendish, Daiki Nobayashi, and Takeshi Ikenaga

Studying the Practicality of Theoretically Proven Online Routing/Scheduling Algorithms for Autonomous Logistics Systems 6
Su Dong, Parimala Thulasiraman, and Pak Ching Li

Private Information in Public Data - A Forensic Analysis of Unaffected User Image Metadata Provided by Online Platforms 12
Felix Fischer, Elisabeth Ziesch, and Dirk Labudde

Medical IoT Devices and Systems: Challenges and Opportunities

Dirceu Cavendish, Daiki Nobayashi, Takeshi Ikenaga
Faculty of Engineering
Kyushu Institute of Technology
Fukuoka, Japan
e-mail: {cavendish@net.ecs, nova@ecs, ike@ecs}.kyutech.ac.jp

***Abstract*—Rapid adoption of Internet of Things (IoT) devices within different verticals has stressed the need to strengthen security of such devices and systems. In particular, medical IoT devices must satisfy not only security, but also safety and privacy requirements. In this paper, we provide a snapshot of modern medical device systems, and underline a balanced approach to security of medical IoT systems, taking into consideration not only security risk mitigation measures but also safety and privacy of such systems. We further advocate a data driven health care delivery framework.**

Keywords*—*Medical wearables; Biosensors; Biomarkers; Personal Health Information; Risk controls; Software as a Medical Device.

I. INTRODUCTION

The advanced state of communication networks of today has impacted many industry verticals, from financial to automotive to power transmission systems. In the medical world, the pervasive connectivity of the Internet, combined with advances in wearable sensors and data mining, is challenging traditional health care delivery models across the globe. The evolution of medical systems, however, must follow tight regulatory guardrails in order to ensure safety and privacy of patients.

In this work, we discuss challenges and opportunities of modern medical IoT devices, in this rapidly evolving "all connected" and data driven world. Specifically, we highlight security, privacy, and safety of medical IoT (m-IoT) devices. Our ultimate goal is to expose specific research areas that ultimately will enable a data driven health care delivery system.

The paper is organized as follows. Related work is included in Section II. Section III describes current medical ecosystem and its various components: Cloud, controllers, medical IoT devices, including biosensors. Section IV addresses security, safety, and privacy of medical devices and systems, with emphasis to challenges and opportunities. Section V addresses how medical data enables a novel data oriented healthcare model that benefits both healthcare providers as well as patients, via use of Artificial Intelligence/Machine Learning (AI/ML) techniques applied to healthcare. Section VI summarizes our studies and addresses promising medical IoT device research directions.

II. RELATED WORK

There has been a large proliferation of research work related to medical IoT devices recently. Most of them deal with one aspect among safety, security, privacy of these devices. For instance, Nanni et al. [1] addresses the cybersecurity vulnerabilities of medical IoT devices, providing a taxonomy as well as statistical data on various types of products. Granlund et al. [2] analyses cybersecurity and safety of medical devices from a regulatory perspective. Yaqoob et al. [3] proposes an Integrated Safety, Security, and Privacy framework to evaluate medical devices, applying it to risk evaluation of an infusion pump. They provide an excellent introduction to medical devices design controls regulated by U.S. Food and Drug Administration (FDA-US) and Medical Device Regulation (MDR-EU) regulatory bodies. Our work takes a more comprehensive approach, aiming at not only describing security, safety, and privacy aspects of m-IoT devices and system, and their interplay when comes to device requirements, but also describing specific challenges and research opportunities in realizing a data driven healthcare framework supported by a modern medical system.

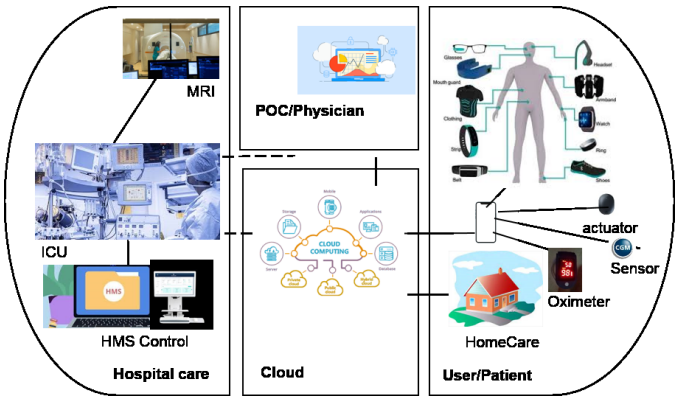


Figure 1. Modern Medical System.

III. MODERN MEDICAL SYSTEM

Figure 1 illustrates the complex medical Ecosystem nowadays. We divide the Ecosystem in four sub-systems:

- Hospital Care: Main hospital complex infrastructure, where major health procedures such as surgeries are delivered. The infrastructure contains heavy/large healthcare diagnosis and medical procedure machines, such as Magnetic Resonance Imaging (MRI) and Intensive Care Unit (ICU) Pumps. Hospital Managent System (HMS) is typically part of the infrastructure of a modern hospital.
- Point of Care (POC): This is where Physicians such as Health Care Providers (HCP) and specialists deliver healthcare services in clinics.
- Medical Cloud: This subsystem supports several medical services, from heavy machinery controls (manufacturer

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

Copyright (c) IARIA, 2026. ISBN: 978-1-68558-349-1

equipment management) to physician data delivery and reports to patient data collection.
- Patient location: This subsystem encompasses patient location where care is delivered, such as a private home, or a nursing home.

Within the patient sub-system, the concept of home health care is well underway, propelled by multiple wearables and sensors currently available. In this work, we focus on sensors and wearables that are connected to HCPs, hospitals, or/end Cloud services, hereafter referred simply as medical IoT devices.

A. Current Healthcare Delivery

In most developed countries, health care is provided by a two tier system: Primary care physicians provide preventative care via patient periodic visits, typically once or twice a year, with participation of specialists once abnormality is detected; reactive care, where specialists and hospitals provide treatments for specific diseases, once detected. One problem is that preventative procedures are not always able to detect diseases at an early enough stage, decreasing probability of cure outcomes. Another issue is that the laboratory tests are performed at the snapshot time of the point of care visit. Continuous tracking of patient data, such as biomarkers (see III-B) is rarely performed.

B. Biomarkers

Biomarkers are measurable indicators of a biological state or condition. Indicators may include molecules, genes, proteins, etc. Biomarkers may be extracted from blood, urine, tissues, etc. Fig. 2 illustrates a few biomarkers.

Biomarkers can be classified according to their purpose and usage:

- Disease detection and diagnosis: Used for helping detect health conditions, even before symptoms may appear.
- Prognosis: Used for tracking severity and progression of a health condition.
- Monitoring: Used for measuring treatment response and effectiveness. Drugs side effects and consequences are managed by this type of biomarker.
- Treatment personalization: Used for tailoring to a patient a specific treatment.
- Drug development: Used for supporting clinical trials for innovative treatments.
- Risk assessment: Used for identifying individuals and group types with high chance of developing a disease.

In terms of tracking biomarkers, we differentiate between slow varying and fast varying biomarkers. Slow biomarkers may be tracked with laboratory exams few times a year. Fast varying biomarkers' readings may change significantly within few hours or days. These biomarkers require constant monitoring. Hence, biosensors is the most appropriate alternative for tracking.

Comp. Metabolic Panel (14)

Test	Current Result and Flag		Previous Result and Date		Units	Reference Interval
▲ Glucose[63]	107	High	95	02/25/2025	mg/dL	70-99
BUN[53]	15		19	02/25/2025	mg/dL	8-27
▼ Creatinine[53]	0.72	Low	0.73	02/25/2025	mg/dL	0.76-1.27
eGFR	105		104	02/25/2025	mL/min/1.73	>59
BUN/Creatinine Ratio	21		26	02/25/2025		10-24
Sodium[53]	138		138	02/25/2025	mmol/L	134-144
Potassium[53]	4.1		4.3	02/25/2025	mmol/L	3.5-5.2
Chloride[53]	103		102	02/25/2025	mmol/L	96-106
Carbon Dioxide, Total[53]	21		25	02/25/2025	mmol/L	20-29
Calcium[53]	9.1		9.0	02/25/2025	mg/dL	8.6-10.2
Protein, Total[53]	6.9		7.0	02/25/2025	g/dL	6.0-8.5
Albumin[53]	4.3		4.2	02/25/2025	g/dL	3.8-4.9
Globulin, Total	2.6		2.8	02/25/2025	g/dL	1.5-4.5
Bilirubin, Total[53]	0.8		0.9	02/25/2025	mg/dL	0.0-1.2
Alkaline Phosphatase[53]	84		76	02/25/2025	IU/L	44-121
▲ AST (SGOT)[53]	55	High	66	04/29/2025	IU/L	0-40

Figure 2. Biomarkers.

TABLE I. BIOSENSORS

Biosensor	Measures	Conditions
Oximeter [19]	Lungs oxygen saturation level heart rate	Pneumonia Blood Clots Covid-19
Glucose monitors [18]	Glucose of interstitial skin	Diabetes Obesity Insulinomia
Cholesterol monitors smart contact lenses [14]	Cholesterol in tear fluid	Hyperlipidemia Cardiovascular diseases
Electronic tattoos [17]	Heart rate Blood pressure	Cardiovascular diseases
Smartwatch/ring	steps sleep data	Sleep abnormalities

C. Biosensors

Recent development of microelectronics have made possible the design and manufacturing of several biosensors. Table I illustrates the most popular biosensors, their data and usage.

D. Data Driven Healthcare

Modern data storage systems have allowed the gathering of a multitude of health related data, whether on health care sites or in Cloud storage services [4]. In particular, data lake technologies are widely used for health care data storage.

We can classify healthcare data into few categories according to specific purposes:

- Descriptive data: Patient Health Information (PHI), such as age, gender, ethnicity; Biomarker's data 2; Environment data: patient location, work activities, workout activities (sports); Data outcomes: admission statistics; mortality rates; infection rates
- Assisted diagnosis: Databases holding specific illnesses data, to support research, diagnosis, and clinical trials. Patient anonymized data feed into data lakes. AI/Machine Learning techniques attempt to improve diagnosis time and accuracy.
- Predictive/prescriptive Analytics: Focuses on patient outcomes to diseases, mortality rate, medications' reactions and side effects. It may be used for decisions about multiple viable therapies, sometimes associated with patient genetics.

IV. Security, Safety, and Privacy of Medical Systems

In a complex medical system (Fig. 1), each component presents security vulnerabilities and security threats. In addi-

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

tion, safety threats may arise from communication issues with outside world. Finally, privacy issues may be present via data logging and hardware address information availability.

A. Edge medical devices

Device implants, such as pacemakers, require calibration, and management by an external controller. Lack of encryption and authentication between the controller and device makes it feasible the injection of harmful commands. The limited hardware and software capabilities of these devices makes it a challenge the support of a full security stack.

Biomarkers/sensors also suffer from limited hardware and software resources. Firmware authentication and secure updates are typically lacking, risking data tampering and Man In The Middle (MITM) attacks. [1] provides a taxonomy and statistical data on vulnerabilities of various wearable medical devices.

Many of these devices have as controllers applications running on smartphones, nowadays. These controllers are vulnerable to Smartphone OS vulnerabilities, tracked as Common Vulnerability Scoring System lists for specific Operating Systems (OS) versions and known vulnerabilities. Another security challenge is the lack of special security hardening OS features - essentially, medical applications are treated with the same level of security requirements as any other application on a smartphone. Typically, these medical applications lack authentication and attestation verification, even though authentication mechanisms and attestation services are available Application attestation consists in accessing the security health of an application and the device it is running at, e.g. Google Play Integrity [5].

Communication of medical sensors with controller typically takes place over Bluetooth Low Energy (BLE) wireless technology. While BLE provides a security layer, additional security protocols may be in order [6]. For instance, a medical device controller should prevent itself from pairing with a foreign BLE device. Also, a Denial of service attack may be possible if foreign controllers continually attempts to BLE pair with a medical device. Availability of hardware addresses, such as BLE MAC address may pose privacy threats if the address can be traced to user PHI information.

Finally, cryptographic material, such as certificates and cryptographic keys, needs to be protected on medical devices and sensors. One option is to use hardware areas that protect from reading/writing access outside device firmware. However, that in itself poses a challenge in lifecycle management of these cryptographic material. For instance, a medical device certificate rotation becomes difficult, especially because the medical device may not have direct connectivity to cloud services.

B. Medical Backend System

1) Medical Cloud Services: Cloud services have been growing steadily in most business areas, as attested by widely used Cloud Infrastructure such as Microsoft Azure and Amazon Web Services (AWS). From a regulatory standpoint, medical cloud services must follow the same safety, security, and privacy regulatory requirements as medical devices. However, due to the backend nature of Cloud services, the challenges are diverse from medical devices.

Modern Cloud Infrastructures have advanced security mechanisms, such as multiple cloud regions with hot backups to prevent service disruption in case of outages, load balancers with Denial Of Service preventative techniques, and gateways that arbitrate incoming and outgoing service and data transit tightly. They rely heavily on Public Key Infrastructure (PKI) certificates, which in itself requires proper management. Medical cloud services and infrastructure require design and deployment of cloud security mechanisms described above. In other words, cloud services are held to a high standard of security requirements (see IV-C).

Cloud privacy regulations [16] controls the collection of user data and its processing in the Cloud. Firstly, the principle of data minimization must be followed, which basically means that every piece of data collected must have a purpose to being gathered, and this purpose must not only be disclosed but also be consented by the user to be collected for that specific purpose. This explicit consent poses difficulties to comply with other regulatory obligations, such as post-marketing surveillance mechanisms to root cause analyze post deployment issues. A second regulatory challenge is the place at which the data is collected. Specific Cloud regions and database locations must ensure that jurisdiction boundaries are respected - collecting European user data into a US cloud region is not permitted. Another aspect of data privacy in the cloud is its capability of auditing data processing. Data access must be traceable; data accuracy must be maintained; data export to owner and to other systems (data portability) must be guaranteed. Data retention and erasure (right to be forgotten [16]) must be supported. Implementation of such tight data control features by the cloud poses many practical challenges.

2) Hospital Infrastructure: Hospital infrastructure typically houses ”legacy” equipment with obsolete Operating Systems that contain plenty of well known vulnerabilities. In addition, many such equipment have hardcoded credentials and no strong data encryption/protection, allowing for outside cyber break in, code injection, and ultimately ransomware attacks. These legacy equipment must be isolated from connectivity with other sub-systems and the Internet, if possible. A hospital management system may provide isolation of sub-systems. In addition, it must support a strong Identity and Access Management (IAM) service, with PHI and medical data access on a strictly and per needed basis.

C. Medical Device Regulatory Landscape

1) Medical Devices Safety: Medical devices' safety are tightly regulated by international standards, from design until end of life. [7] specifies a quality management system for the design, verification and validation of medical devices. It mandates a thorough documentation of the design process, from requirements to their verification and validation. Such documentation is referred to as Design History Files, and it is auditable by regulatory bodies such as FDA in the United

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

States. In addition, [8] specifies a risk management procedure where safety risks are identified, quantified, and mitigated during the design of medical devices.

Still within patient's safety, [12] and [13] address software risks impacting patient's health.

2) Medical Devices Security: Medical device security framework leverages various general security standards, governed by security organizations such as National Institute of Standards and Technology (NIST). Specific to medical devices, [9] describes a framework to access cybersecurity risks associated with medical devices. In addition, [10] describes an approach to conduct risk assessments. Finally, [11] provides a framework to track and manage security vulnerabilities of networked devices.

3) Medical Devices Privacy: The two major regulation standards on privacy are Health Insurance Portability and Accountability Act (HIPPA) [15] and General Data Protection Regulation (GDPR) [16] in the United States and Europe, respectively. These standards define Personal Health Information (PHI) as any data that may be traced to an individual identity, and attempt to protect such data from misuse. In particular, GDPR privacy framework brings transparency to medical data owner regarding their medical data collection and processing. The privacy framework mandates user consent in the usage of such data, and further supports user right to a copy of their data, as well as its deletion, referred to as right to be forgotten. In addition, the framework calls for anonymization techniques to protect user privacy, as well as minimization of data collection, forbidding data gathering without specific purposes consented by the user.

As mentioned before, modern networked medical devices must protect information exported to controllers and Cloud infrastructure from privacy threats. For instance, hardware addresses such as MAC address must be at a minimum obfuscated if it needed to be exported outside the device.

4) Balancing Regulatory Standards: Although compliance with all three regulatory areas, safety, security, and privacy is the ultimate goal, requirements in different areas are sometimes conflicting,as represented in Fig. 3. The overlapping areas of the ovals represent non-conflicting requirements. One can see that only a small set of requirements may be aligned across all three areas.

Examples of tradeoff are:

- A pacemaker controller that has been detected to be security compromised may be required to remain in connection to the pacemaker until a suitable replacement or security fix may be available, for patient safety. In essence, therapy interruption may be more dangerous to the patient than allowing operation upon a security issue detection.
- Device ID/address (such as Media Access Control - MAC address) may be required to be collected, despite privacy concerns, in order to mitigate risk hazards, or security threats, such as connecting to a foreign medical device.

We advocate an integrated risk management system, as per

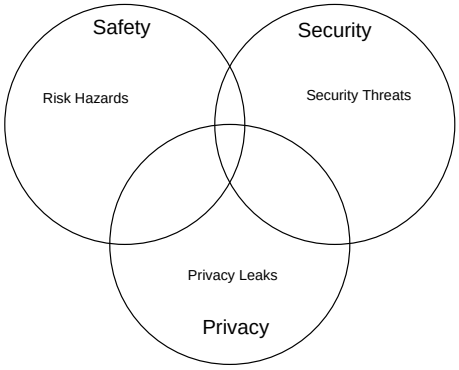


Figure 3. Safety, Security, And Privacy Risks.

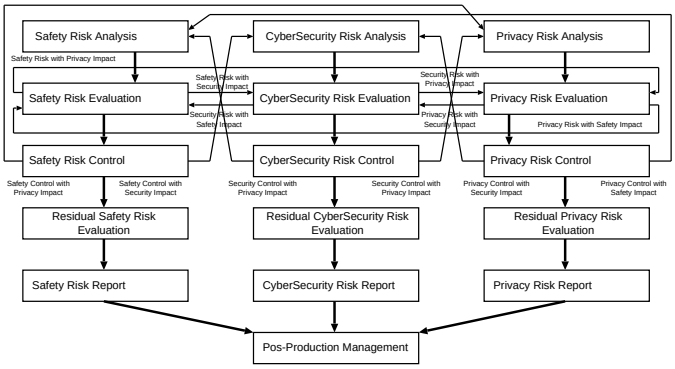


Figure 4. Integrated Risk Methodology.

Fig. 4, extending International Organization for Standardization (ISO) 14971 [8] into cybersecurity and privacy areas.

The figure displays how safety, cybersecurity and privacy risks must be evaluated in a cross impact manner. It also calls for residual risk analysis taking into account cross area risk controls. A Bayesian framework, similar to the one proposed in [3] may be used to quantify hazard probabilities and risk evaluation.

V. Medical Data Oriented Healthcare

Recent advances in medical data collection and large data processing are pushing health care delivery into a new data driven paradigm. Today, physicians and nurses are able to track patients medical data multi-day evolution via Cloud services. In addition, today Cloud data collection may be combined with advanced data processing techniques in search for abnormalities. We then may conceive a health care model at which physician's visit may be triggered by cloud services notifications directly from a patient to the physician's office. For instance, blood glucose can be tracked to detect pre-diabetes condition, and correlated with other data such as patient weight and Body Mass Index (BMI). This personalized data driven diagnoses allows a more preventative health care model, rather than a reactive model where medical data is obtained only upon symptoms appearance. Challenges to build a data driven diagnosis system is that physician's workflows may be different, depending on the specialty and nature of the

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

disease. For instance, an orthopedic joint problem typically requires some imaging processing, where oncology conditions may be detected by extensive lab tests.

In addition to diagnosis, illnesses treatment decision may be driven by tracking specific drugs' side effects within a large patient population, as well as treatment's effectiveness within a group of patients with similar genetic and environmental characteristics. Machine learning techniques, such as clustering, may help with data driven diagnosis and treatment decisions.

At the heart of this data driven healthcare framework is AI/ML models that must be able to accurately process a large amount of data. In this section, we call AI/ML medical products as Software as a Medical Device (SaMD). As with any other medical device, the manufacturer needs to seek approval from regulatory bodies such as FDA and MDR, which requires proof of safety and effectiveness of the SaMD. The self-learning characteristics of such devices make it non-trivial to show safety and effectiveness, and regulatory agencies are still grappling to define methods and procedures for SaMD submissions.

As a baseline, AI/ML SaMD must describe their training sets, their size and diversity so as to perform well when dealing with a diverse patient population. As far as effectiveness is concerned, SaMD may intend to provide assistance to a physician only, for instance, in diagnosis within a specific specialty (see [20] for use cases). In that respect, efficacy may be measured by how fast the physician came to the correct diagnosis when aided by the SaMD, as opposed to doing it alone. This is in sharp contrast to non-learning medical devices, where efficacy is oftentimes proved by comparison with equivalent devices in the market. Finally, in terms of safety, medical devices with learning capabilities must include safeguards in case the learning process goes awry. For instance, a smart ICU pump should support maximum drug delivery rates, so as not to endanger patient's life.

As clinical trials are part of regulatory approval process, specific procedures to accommodate learning mechanisms should be devised for SaMDs [21].

VI. Conclusion and Future Work

In this paper, we have addressed safety, security, and privacy aspects of modern medical systems and their components, with emphasis to medical IoT devices, such as biosensors. We have exposed challenges of modern medical IoT devices security and privacy, vis a vis ensuring user safety. Furthermore, we have argued for a balanced approach in analyzing device risks and risk controls. We have also underlining the challenges in proving the safety and effectiveness of medical devices that make use of AI/ML mechanisms.

Specific topics of research for medical devices include: reliable source of time; authentication and secure communication; secure firmware updates; AI guardrail framework. We are currently investigating some security aspects of medical devices with impact on user privacy and safety.

Acknowledgments

Work supported by JSPS KAKENHI Grant #24K03045.

References

[1] F. M. C. Nanni et al., "Taxonomy and Statistics of Cyber and Physical Vulnerabilities in Medical Devices," 9th International Conference on Smart and Sustainable Technologies - SpliTech, Digital Object Identifier 10.23919/SpliTech61897.2024.10612327, June 2024.

[2] T. Granlund et al., "On Medical Device Cybersecurity Compliance in EU", IEEE/ACM 3rd International Workshop on Software Engineering for Healthcare (SEH), pp. 20-23, 2021.

[3] T. Yaqoob and H. Abbas, "Integrated Security, Safety, and Privacy Risk Assessment Framework for Medical Devices," IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 6, pp. 1752-1761, June 2020.

[4] A. Ahmed et al., "Harnessing Big Data Analytics for Healthcare: A Comprehensive Review of Frameworks, Implications, Applications, and Impacts," IEEE Access, Digital Object Identifier 10.1109/ACCESS.2023.3323574, Oct.2023.

[5] Android Developers, "Google Play Integrity", URL: https://developer.android.com/google/play/integrity/overview, last accessed Oct. 19, 2025 (UTC).

[6] A. Barua et al., "Security and Privacy Threats for Bluetooth Low Energy in IoT and Wearable Devices: A Comprehensive Survey," IEEE ComSoc, Digital Object Identifier 10.1109/OJCOMS.2022.3149732, Feb.2022.

[7] ISO 13485: 2003, Medical devices – Quality management systems.

[8] ISO 14971: 2019, Medical devices – Application of risk management to medical devices.

[9] IEC 80002-1: 2009, Medical device software — Part 1: Guidance on the application of ISO 14971 to medical device software,https://www.iso.org/standard/54146.html, last accessed Nov. 29, 2025.

[10] NIST SP 800-30 Rev. 1 Guide for conducting risk assessment, https://csrc.nist.gov/pubs/sp/800/30/r1/final,last accessed Nov. 29, 2025.

[11] IEC/TS 81001-2-2:2025 Health software and health IT systems safety,effectiveness, and security Part 2-2: Guidance for the implementation, disclosure and communication of security needs, risks and controls https://www.iso.org/standard/85765.html, last accessed Nov. 28, 2025.

[12] IEC 62304: 2006, Medical device software — Software life cycle processes, https://www.iso.org/obp/ui/iso:std:iec:62304:ed-1:v1:en,last accessed Nov. 16, 2025.

[13] IEC 82304: 2016, Health software — Part 1: General requirements for product safety, https://www.iso.org/obp/ui/en/iso:std:iec:82304:-1:ed-1:v1:en, last accessed Nov. 16, 2025.

[14] H. Song et al., "Wireless Non‐Invasive Monitoring of Cholesterol Using a Smart Contact Lens," https://onlinelibrary.wiley.com/doi/10.1002/advs.202203597, last accessed Jan. 18, 2026.

[15] U.S. Department of Health and Human Services, https://www.hhs.gov/hipaa/index.html, last accessed Nov. 16, 2025.

[16] General Data Protection Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016, Directive 95/46. Official Journal of the European Union (OJ), 2016, vol. 59, no. 1-88, p. 294.

[17] K. Sel et al., "Electrical Characterization of Graphene-based e-Tattoos for Bio-Impedance Physiological Sensing," IEEE Biomedical Circuits and Systems Conference (BioCAS), Print on Demand, Oct., 2019.

[18] S.A.Siddiqui et al., "Pain-Free Blood Glucose Monitoring Using Wearable Sensors: Recent Advancements and Future Prospects," IEEE Reviews in Biomedical Engineering, Vol. 11, pp. 21-35, Apr., 2018.

[19] M. A. Motin et al., "Compact Pulse Oximeter Designed for Blood Oxygen Saturation and Heart Rate Monitoring," 3rd Int. Confrence on Electrical & Electronic Engineering (ICEEE), pp. 125-128, Dec. 2021.

[20] S. Ahmed, J. Y. Raja, M. Y. A. Raja, "eHealthcare: IoT & AI Enhance the Scope and Effectiveness of Diagnostics and Treatment Modalities", IEEE 21st International Conference on Smart Communities: Improving Quality of Life using AI, Robotics and IoT (HONET), pp. 156-161, 2024.

[21] S-R Yang, J-T Chien, and C-Y Lee, "Advancement in Clinical Evaluation and Regulatory Frameworks for AI-Driven Software as a Medical Device," IEEE Open Journal of Engineering in Medicine and Biology, vol. 6, pp. 147-151, 2025.

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

Studying the Practicality of Theoretically Proven Online Routing/Scheduling Algorithms for Autonomous Logistics Systems

Su Dong, Parimala Thulasiraman, Pak Ching Li
Department of Computer Science
University of Manitoba
Winnipeg, Canada
e-mail: dongs@myumanitoba.ca, parimala.thulasiraman@umanitoba.ca, ben.li@cs.umanitoba.ca

***Abstract*—Online routing is a complex problem in autonomous logistics systems such as online delivery applications. The problem data is not known in advance. Therefore, finding efficient solutions to schedule and route for online problems is difficult. Practitioners rely on historical data to propose data-driven machine learning solutions, while theoreticians propose strong, theoretically bound, non-data-driven methods using competitive analysis. The question this paper attempts to answer is whether we can implement theoretical algorithms for logistics systems in an uncertain environment. In this paper, we do this by studying the Online Travelling Salesperson Problem (OLTSP). We examine the gap between the theoretical performance guarantees of the proposed algorithms for OLTSP and their observed performance in practice. We perform various experiments and show that competitive algorithms perform significantly better in practice than their worst-case guarantees suggest.**

Keywords-Online Travelling Salesperson Problem; Intelligent Transportation System; Competitive Algorithms.

I. INTRODUCTION

Recent advances in automation have made connected autonomous vehicles and aerial drones a promising solution for modern logistics systems. One of the key components of logistics systems is an Intelligent Transportation System (ITS). An ITS comprises two main components: a road network and a vehicular network or Internet of Vehicles (IoV). An IoV consists of vehicles interconnected as a distributed network. An IoV performs tasks such as delivering and collecting goods (e.g., online food delivery) within a city without human intervention. Such urban logistics is more complex and time-sensitive. Designing efficient online real-time scheduling strategies for IoV applications is a difficult problem.

In a typical logistics scenario, a service provider operates autonomous vehicles or drones to perform delivery and pickup tasks within a fixed service region. The region consists of predefined locations such as warehouses, customer sites, and pickup points. These locations are connected by routes with known costs that represent travel time or distance. Service requests are released over time. Each request specifies a location and a release time. A request cannot be served before its release time. At any moment, vehicles or drones only know the requests that have already been released. Information about future requests is unavailable. As a result, routing decisions must be made based on incomplete information and may need to be updated when new requests appear. The objective is to serve all requests while keeping the total travel cost as low as possible. This setting can be naturally modeled by the OLTSP that captures the key challenges of online logistics systems, where requests arrive over time and routing decisions must be made without knowledge of future demands.

Ausiello et al. [1] proposed the OLTSP. In this problem, a single salesperson is assumed. In our work, we interpret the salesperson as a single autonomous vehicle. The OLTSP, in general, can be extended or adapted to multi-vehicle systems through decomposition or assignment strategies.

There are two versions of OLTSP, Homing OLTSP (H-OLTSP) and Nomadic OLTSP (N-OLTSP) [1]. Although in both versions the goal is to minimize the total completion time to serve all the presented requests, in H-OLTSP, the server must return to the origin after completing all requests. In this paper, we focus on the H-OLTSP on general metric spaces. In this model, the server starts from a designated depot and must return to the same depot after completing all requests. This setting is practical for real-world applications.

In recent years, data-driven methods have been key in addressing real-world routing and logistics problems, making machine learning and other heuristic approaches attractive in practice. However, these methods usually lack proven performance guarantees. They are also not theoretically proven. In automated logistics systems, service regions may change over time, and historical data collected from previous regions may no longer be accurate or useful. In such settings, data-driven approaches may perform poorly.

Online algorithms for H-OLTSP have been theoretically analyzed using competitive analysis, and are therefore often referred to as competitive algorithms, but have not been experimentally explored. *This method compares the performance of the online algorithm against an offline optimal algorithm* in which the input sequence is known in advance. The competitive analysis of an algorithm provides guarantees on solution quality under all possible input sequences in the worst-case. This property is important and reliable when there is insufficient data. As a result, competitive algorithms offer a robust alternative when data is limited, outdated, or unavailable.

Although the performance guarantees of competitive algorithms may appear pessimistic, they are derived from worst-case analysis rather than typical behavior. It is unclear how often such worst-case scenarios occur in practical applications. This raises an important question about the practical perfor-

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

mance of competitive algorithms.

Ausiello et al. [1] proposed online routing algorithms for scheduling requests: the *Plan At Home* (PAH) algorithm for the H-OLTSP on general metric spaces and the *Greedily Travelling between Requests* (GTR) algorithm for the N-OLTSP. Based on these competitive algorithms, our contribution is as follows:

- Provide a formal definition and detailed description of the GTR algorithm for H-OLTSP, following the adaptation stated by the authors [1].
- Experimentally evaluate scheduling strategies for routing requests in PAH and GTR.
- Study the gap between theoretical worst-case guarantees and observed empirical performance.

To the best of our knowledge, PAH and GTR are the only algorithms designed for H-OLTSP on general metric spaces, without additional assumptions or restrictions.

The remainder of this paper is organized as follows. Section II describes the problem model of the H-OLTSP. Section III presents the competitive algorithms studied in this paper. Section IV details the experimental data generation. Section V reports the experimental results. Section VI discusses these results. Section VII explains limitations. Section VIII concludes the paper and provides directions for future work.

II. Problem Model

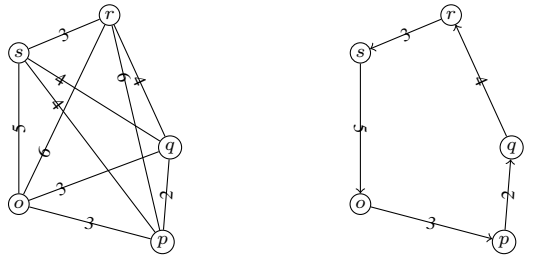
In the classical Travelling Salesperson Problem (TSP), given a set of locations, the goal is to find the shortest Hamiltonian cycle that visits each location exactly once. This is an offline problem where all the locations to be visited are given as input. Computing such an optimal tour is NP-hard on general metric spaces. As a result, computing an optimal solution for the offline version of H-OLTSP on general metric spaces is computationally intractable unless $P = NP$. The formal definition of H-OLTSP is given in Definition 1 [1]. It models a moving agent, called the server, that travels in a metric space to serve online requests.

Definition 1. *H-OLTSP*

The input to the H-OLTSP consists of the following:

- *A metric space (M, d), where M is a set of locations or points and $d : M \times M \to \mathbb{R}_{\geq 0}$ is a distance function satisfying the standard metric properties: non-negativity, identity, symmetry, and the triangle inequality.*
- *A designated starting point $o \in M$, called the* origin*, where the server is initially located at time zero.*
- *An online sequence of requests $Q = \{(t_1, p_1), (t_2, p_2), \ldots, (t_n, p_n)\}$, where each $p_i \in M$ is a requested location and each $t_i \in \mathbb{R}_{\geq 0}$ is the release time of the request. The sequence is ordered such that $t_i \leq t_{i+1}$ for all $1 \leq i < n$, and request (t_i, p_i) is revealed to the algorithm only at time t_i.*

The server moves through the metric space at unit speed, so traveling between any two locations x and $y \in M$ requires exactly $d(x, y)$ units of time. A request (t_i, p_i) is considered served if the server visits the location p_i at some time $t \geq t_i$.



(a) Metric Space (b) Offline Optimal Solution

Figure 1. An Example OLTSP Instance

The objective is to guide the server, in an online fashion, starting from the origin, to serve all requests in Q and return to the origin after all requests are completed. The goal is to minimize the ***completion time****, defined as the total time that elapses from the start (time zero) until all requests have been served and the server has returned to the origin. This includes both the time spent traveling and any time the server remains idle.*

III. Competitive Online Algorithms

To illustrate these algorithms, we consider an example of H-OLTSP: $[(M, d), o, Q]$. The request sequence Q is given by $\{(0, r), (2, p), (4, q), (10, s)\}$. The metric space (M, d) is defined by the graph shown in Figure 1 (a). In this graph, M represents the set of nodes, o is the origin node, and d is a distance function mapping pairs of nodes to their distances, with the corresponding values specified by the edge labels. An offline optimal solution for this instance is shown in Figure 1 (b).

A. PAH Algorithm

In the PAH algorithm, the server always schedules its route at the origin. Several special cases may arise during the algorithm's execution. These cases are described in Definition 2 [1].

Definition 2. *Main Logic of PAH*

- ***Case 1*** *The server is at the origin:*
  - *The server schedules or reschedules a route that starts at the origin, visits all remaining released but unserved request locations exactly once, and returns to the origin at the end. It immediately starts moving along this route.*
- ***Case 2*** *If the server is not at the origin and one or more new requests are released, the server decides whether or not to return to the origin to reschedule based on its location and the locations of new requests.*
  - ***Case 2.1*** *All newly released requests are located at positions whose distances from the origin are no greater than the distance from the origin to the server's current location at the moment they are released:*
    - *These new requests are temporarily ignored, and the server continues its current route. They are*

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

scheduled when the server next arrives at the origin (re-entering Case 1).

– ***Case 2.2*** *Otherwise, the server returns to the origin by a shortest route (re-enter Case 1).*

We illustrate the PAH algorithm on the example instance in Figure 1. At time 0, the server is at the origin and the request $(0, r)$ is released. Since the server is at the origin (**case 1**), it computes a tour $o \to r \to o$ and moves from o toward r.

At time 2, the request $(2, p)$ is released. The server is not at the origin, so **case 2** applies. Since $d(p, o) > R_d(server, o)$, **case 2.2** applies, and the server returns to the origin to recompute a tour where $R_d(server, o)$ is the distance of the shortest route from the server back to the origin.

The server arrives at the origin at $t = 4$, when the request $(4, q)$ is released. As the server is at the origin (**case 1**), it computes the tour $o \to p \to q \to r \to o$. The server moves from o to p from time 4 to 7, and from p to q from time 7 to 9. Thus, p is served at time 7 and q at time 9.

At time 10, the request $(10, s)$ is released. Since $d(s, o) > R_d(server, o)$, **case 2.2** applies and the server returns to the origin via server's location $\to q \to o$, arriving at $t = 14$. At time 14, the server computes the tour $o \to s \to r \to o$. It moves from o to s from time 14 to 19, from s to r from time 19 to 22, and from r back to o from time 22 to 28. Hence, the completion time is 28.

Scheduling Strategies: The PAH algorithm requires a scheduling algorithm to compute routes for serving requests, where each route is nothing but a TSP tour. In this paper, we consider two scheduling methods. The first method is dynamic programming [2], which can compute optimal TSP tours. The second is the Christofides heuristic [3]. Since a TSP tour is a cycle, the origin has two adjacent locations in the tour. When starting from the origin, the server can move toward either one. In this work, we let the server choose the one that is closer to the origin, as a simple and reasonable modeling choice consistent with routing decisions commonly used in autonomous logistics systems.

B. GTR Algorithm

The GTR algorithm guides the server to schedule or reschedule a route every time new requests are released. Ausiello et al. [1] proposed the GTR algorithm for N-OLTSP, and stated that the GTR can also be applied to the H-OLTSP by modifying the route so that it ends at the origin. However, the authors [1] did not provide a formal definition or pseudocode for the H-OLTSP variant. To make the algorithm well defined in our setting, we present an explicit formulation of the core logic of GTR for H-OLTSP, based on the descriptions and proofs given in [1]. While our formulation is intended to closely follow the original design philosophy, minor differences in interpretation or implementation details may arise, due to the absence of an explicit H-OLTSP specification in [1]. For completeness and reproducibility, the exact version of GTR used in our experiments is stated in Definition 3.

Definition 3. *Main logic of GTR for H-OLTSP*
At any time instant, when one or more new requests are released, the server immediately enters Case 2; if no new requests are released, Case 1 applies.

Case 1: There are no new requests.

- ***Case 1.1:*** *A route is currently scheduled and being followed (The route contains all released but unserved request locations):*
  – *The server continues moving along the route.*
- ***Case 1.2:*** *No route is currently being followed (i.e., Case 1.1 does not apply):*
  – *The server remains idle.*

Case 2: One or more new requests are released.

- ***Case 2.1:*** *The server is at a location (point).*
  – *The server computes a route starting from its current location that visits all remaining released but unserved request locations exactly once and returns to the origin at the end. It immediately begins moving along the route.*
- ***Case 2.2:*** *The server's current position lies on the shortest path (with respect to the underlying metric) between two locations, denoted by x and y.*
  – ***Case 2.2-a:*** *Each of x and y is either the origin or a request location that has been released at or before the current time.*
  ⇒ The server computes two candidate routes that begin at its current location, one that first visits x and one that first visits y, each then visiting all other released but unserved request locations exactly once and returning to the origin at the end. It immediately starts moving along the shorter route, breaking ties arbitrarily.
  – ***Case 2.2-b:*** *Exactly one of x or y is the origin or a request location that has been released at or before the current time.*
  ⇒ Identify the location $l_p \in \{x, y\}$ that is either the origin or a request location that has been released at or before the current time. The server computes a route that begins at its current location, first visits l_p, then visits all other released but unserved request locations exactly once and returns to the origin at the end. It immediately starts moving along this route.

To illustrate Definition 3, we apply GTR to the example instance of Figure 1. At time 0, the server is at the origin o. The request $(0, r)$ is released, so the server computes the route $o \to r \to o$ and starts moving along it (**case 2.1**).

At time 2, the request $(2, p)$ is released while the server is between the nodes o and r (**case 2.2**). The server computes two candidate routes, $P(s) \to r \to p \to o$ and $P(s) \to o \to p \to r \to o$, and starts moving along the shorter route $P(s) \to r \to p \to o$, where $P(s)$ is the position of the server.

At time 4, the request $(4, q)$ is released. The server again compares two possible routes and follows the shorter one, which is $P(s) \to r \to q \to p \to o$. The server arrives at r at time 6 and arrives at q at time 10.

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

At time 10, the request $(10, s)$ is released. The server recomputes a route $q \to s \to p \to o$ and moves along this route (**case 2.1**). The server completes s at time 14, completes p at time 18, and arrives back at the origin at time 21. Hence, the completion time is 21.

Scheduling Strategies The GTR algorithm also requires a scheduling algorithm to compute routes. Since GTR schedules immediately, each computed route starts from the server's current location, serves all currently active requested locations, and returns to the origin at the end. In this paper, we consider two scheduling methods for GTR. The first is an exact scheduling algorithm based on dynamic programming [2], which can compute optimal routes. The second is the Minimum Spanning Tree (MST) heuristic [1].

IV. Experimental Setting

In experiments, our objective is to evaluate the practical behavior of competitive algorithms in a way that is representative and broadly applicable. Rather than targeting a specific application or dataset, we aim to capture typical characteristics of automated logistics systems while retaining full control over instance difficulty and structure.

To model the spatial component of the problem, we use benchmark instances from TSPLIB [4] to induce discrete metric spaces. Each TSPLIB instance defines a finite set of locations with pairwise distances, which can be naturally interpreted as a service region in an automated logistics setting, such as an area served by autonomous delivery vehicles or drones. In many practical applications, service requests are typically modeled as originating from a finite set of predefined locations within the service region, and any such location may issue requests. To reflect this assumption without introducing spatial bias, requested locations are selected uniformly at random from the set of locations in the induced metric space.

The temporal component of request generation is handled differently. In real automated logistics systems, requests are rarely issued uniformly over time. Instead, demand often exhibits temporal concentration and peak periods. To approximate this behavior in a controlled manner, we generate request release times using a normal distribution. By varying the standard deviation, we control how closely release times are clustered, allowing us to model different levels of temporal congestion and overlap among active requests.

A key design choice is the use of a fixed release time interval. If request times were sampled over an unbounded or excessively long horizon, the gaps between successive requests could become large. In such cases, the server might complete one request before the next is released. This would weaken the need for online decision-making and reduce performance differences among algorithms, potentially causing their behavior to converge. Fixing the time interval prevents this effect and ensures that increasing the number of requests leads to greater scheduling difficulty.

This choice also aligns with real-world intuition. A fixed interval can be interpreted as a realistic operational window, such as several hours within a day. In practice, different time windows may exhibit different demand patterns, which corresponds to sampling from distributions with different parameters over comparable time spans. Our approach captures this structure while keeping the experimental setting controlled and comparable across instances.

Algorithm 1 H-OLTSP Instance Generation

Require: (I_{TSP}, σ, N)
- A symmetric TSPLIB instance I_{TSP}.
- Standard deviation σ.
- Total number of requests N.

$M \leftarrow$ the set of all locations in I_{TSP}
$d \leftarrow$ the distance function induced by I_{TSP}
$o \leftarrow$ a location chosen uniformly at random from M
$Q \leftarrow [\,]$ ▷ empty list
$d_{\text{median}} \leftarrow$ the median distance over all distinct pairs of locations in M
$T_{\max} \leftarrow 10 \cdot d_{\text{median}}$
$\mu \leftarrow 5 \cdot d_{\text{median}}$
$T \leftarrow$ a list of N integers sampled from a truncated normal distribution $\mathcal{N}(\mu, \sigma^2)$ restricted to $[0, T_{\max}]$ (duplicates allowed)
$P \leftarrow$ a list of N locations sampled uniformly at random from M (duplicates allowed)
Sort the list T in increasing order
for $i \in \{1, \ldots, N\}$ **do**
    Append request $(T[i], P[i])$ to Q
end for
return $\{(M, d), o, Q\}$

The instance generation procedure is described in detail in Algorithm 1. Request release times are generated within a fixed interval $[0, T_{max}]$, where $T_{max} = 10 \times d_{median}$ and d_{median} denotes the median distance between locations in the metric space M. The median distance provides a representative travel scale that is less sensitive to extreme values and thus offers a stable reference for relating spatial distances to temporal release patterns.

In our experiments, we consider four normal distributions with the same mean μ, which is fixed at the center of the interval $[0, T_{max}]$, and different values of the standard deviation σ. Specifically, we use $\sigma = T_{max}$, $\frac{T_{max}}{3}$, $\frac{T_{max}}{6}$, and $\frac{T_{max}}{10}$ to model varying degrees of temporal concentration. The total number of requests is set to $N = 5, 7, 9, 11, 13, 15$, and 17. Due to space limitations, we use two symmetric TSPLIB instances, berlin52 and gr202, obtained from a publicly available Kaggle dataset [5], to induce the discrete metric spaces.

For each configuration of the input parameters (I_{TSP}, σ, N), five independent instances are generated for experimental evaluation. The experiments are implemented in Python 3.11 and executed on a machine equipped with an Intel i5-13500H processor. To mitigate variability arising from system and hardware conditions, the execution time of each algorithm on a given instance is reported as the average over five independent runs.

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

TABLE I. AVERAGE COMPLETION TIME.

berlin52							
N	5	7	9	11	13	15	17
PAH-DP	6201.45	6662.00	7400.25	7526.50	7837.05	8296.20	8638.90
PAH-Chris	6216.30	6713.95	7516.80	7719.50	8022.45	8505.85	8804.20
GTR-DP	5439.30	5863.20	6352.30	6346.10	6569.80	6773.50	7121.45
GTR-MST	5457.15	5908.40	6427.60	6498.80	6770.95	7172.95	7392.25
Offline-Optimal	4838.10	5119.80	5626.55	5686.75	5798.10	6069.05	6224.10
gr202							
N	5	7	9	11	13	15	17
PAH-DP	14192.90	16580.90	17164.55	18010.15	18600.35	20351.60	19455.75
PAH-Chris	14272.40	16755.00	17503.75	18531.80	19247.70	21081.65	19968.05
GTR-DP	12226.50	14325.25	14377.55	15331.10	15403.90	17548.65	16237.70
GTR-MST	12314.30	14525.90	14650.65	15823.35	15735.60	18478.30	16885.60
Offline-Optimal	10535.15	12293.15	12593.55	13421.50	13735.55	15302.20	14539.00

TABLE II. RATIO.

	Min.	Mean	Median	Max.	Std.	Competitive Ratio
PAH-DP	1.0585	1.3405	1.3296	1.7692	0.1327	2(tight)
PAH-Chris	1.0585	1.3676	1.3522	1.7740	0.1399	3
GTR-DP	1.0042	1.1357	1.1247	1.3841	0.0615	2.5 †
GTR-MST	1.0226	1.1649	1.1544	1.4655	0.0725	unknown

† The competitive ratio is stated by Ausiello et al. [1], but a formal proof is not provided.

V. EXPERIMENTAL RESULTS

We refer to the PAH algorithm using dynamic programming as PAH-DP, and to the PAH algorithm using the Christofides heuristic as PAH-Chris. We refer to the GTR algorithm using dynamic programming as GTR-DP, and to the GTR algorithm using the MST heuristic as GTR-MST.

Table I presents average completion times for varying numbers of requests N, with separate sub-tables for instances generated from berlin52 and gr202; all values are rounded for presentation. From Table I, we observe that, across different values of N, the average completion time of all algorithms is higher than the corresponding offline optimal average completion time, and remains below twice the offline optimal average completion time for all evaluated cases. In comparisons between dynamic programming and heuristic variants within the same algorithm family (PAH-DP vs. PAH-Chris and GTR-DP vs. GTR-MST), we observe that the dynamic programming variants generally achieve a lower average completion time for most values of N in the table. Among online decision making methods, GTR yields lower average completion times than PAH when comparing corresponding variants (GTR-DP vs. PAH-DP and GTR-MST vs. PAH-Chris).

Table II reports the ratio between the completion time of each algorithm and the offline optimal completion time; all reported values are rounded for presentation. Lower values indicate solutions closer to the optimal. For all algorithms with known competitive ratios, the observed ratios are well below their competitive ratio.

Table III presents average execution times for varying N, with four sub-tables corresponding to instances whose release times are generated from a normal distribution with different σ values; all values are rounded for presentation. Here, execution time refers to the simulated time required by an algorithm to complete a single H-OLTSP instance. This measure provides an approximate indication of the total computational effort incurred over the course of the instance. Although execution time does not directly measure per-decision latency, it can offer insight into an algorithm's overall responsiveness. In particular, a longer execution time suggests that scheduling and rescheduling decisions may require greater computational effort as requests arrive, while shorter execution times indicate more efficient decision-making in dynamic settings.

Overall, for algorithms that rely on dynamic programming, instances with smaller σ and larger N tend to have higher average execution times in most cases. This is because more concentrated release times typically lead to a larger number of simultaneously active requests, which increases scheduling difficulty. For polynomial time scheduling methods (Christofides and MST heuristics), this trend is less apparent in our experiments, partly due to the relatively small instance sizes considered. However, in online scheduling, execution time is influenced not only by scheduling complexity but also by the frequency of rescheduling. A notable special case is GTR-DP, whose average execution time increases sharply at $N = 17$ (third sub-table of Table III). In some instances, newly released requests repeatedly interrupt the current route, forcing frequent rescheduling, as GTR recomputes the entire routes whenever one or more requests are released. This behavior can have a particularly strong impact when an exact scheduler, such as dynamic programming, is used. Despite this, GTR-DP still exhibits a lower average execution time than PAH-DP across the evaluated cases.

Graph operations are implemented using NetworkX [6]. The Christofides and MST heuristics rely more on graph operations, which introduce a small data conversion overhead. This overhead is more noticeable when the number of requests N is small, since only a few requests are scheduled at each decision point and complexity differences between methods are less pronounced. As a result, heuristic methods may exhibit longer average execution times than dynamic programming for some small values of N in Table III.

Table III also shows that GTR-DP consistently achieves lower average execution times than PAH-DP, and that GTR-MST consistently achieves lower average execution times than PAH-Chris in the reported results. A key reason lies in how the two algorithms handle scheduling and rescheduling. GTR reschedules immediately when new requests are released, so each scheduling decision is based on the set of requests that are active at that moment. In contrast, PAH only schedules or reschedules routes when the server is at the origin. When new requests arrive, and PAH decides that rescheduling is needed, it must first return to the origin before computing a new route. During this return period, additional requests may be released. As a result, when PAH eventually performs rescheduling, it may need to consider a larger set of active requests than GTR. This can increase the effort required for each tour computation and tends to result in higher overall execution time.

VI. DISCUSSION

Based on the experimental results, we observe that the evaluated competitive algorithms perform significantly better

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

TABLE III. AVERAGE EXECUTION TIME(MS).

$\sigma = T_{max}$							
	5	7	9	11	13	15	17
PAH-DP	21.61	35.67	40.41	148.75	292.94	237.89	818.47
PAH-Chris	22.66	39.77	43.34	169.73	51.32	121.31	96.75
GTR-DP	9.01	12.15	12.97	39.99	28.25	42.59	68.92
GTR-MST	9.50	13.80	14.75	26.99	25.65	31.64	36.61
$\sigma = \frac{T_{max}}{3}$							
N	5	7	9	11	13	15	17
PAH-DP	12.08	36.85	30.40	91.01	208.88	748.49	4010.96
PAH-Chris	12.53	37.67	27.45	79.65	45.68	66.71	77.40
GTR-DP	8.42	12.47	16.94	21.92	36.40	525.76	324.34
GTR-MST	9.22	13.88	17.61	20.04	24.51	36.87	34.42
$\sigma = \frac{T_{max}}{6}$							
N	5	7	9	11	13	15	17
PAH-DP	31.89	46.06	91.19	77.09	439.09	627.62	4118.77
PAH-Chris	33.33	47.10	74.14	36.32	39.48	51.43	88.51
GTR-DP	8.97	11.40	22.33	32.54	150.78	143.70	3063.06
GTR-MST	9.60	13.36	20.63	22.15	27.72	33.43	42.31
$\sigma = \frac{T_{max}}{10}$							
N	5	7	9	11	13	15	17
PAH-DP	13.26	32.89	40.72	86.19	664.54	1821.45	6171.44
PAH-Chris	14.82	32.10	19.51	37.61	43.57	90.51	65.95
GTR-DP	9.05	14.18	35.36	62.78	433.57	1764.86	1913.66
GTR-MST	10.25	14.98	18.75	22.79	31.03	42.59	44.01

in practice than their competitive ratios suggest. In all tested settings, the gap between solutions of these algorithms and the offline optimal solutions remains small. This indicates that competitive analysis provides a conservative bound, and that these algorithms can achieve much stronger performance under realistic conditions.

Among the evaluated methods, GTR consistently demonstrates strong practical performance when evaluated through paired comparisons against the corresponding PAH methods (GTR-DP vs. PAH-DP and GTR-MST vs. PAH-Chris). Its strategy of immediately scheduling or rescheduling routes allows it to incorporate new requests quickly, which typically results in lower completion times in these paired comparisons. This behavior is particularly advantageous in delivery scenarios where a timely response to new orders is important. At the same time, our experimental results indicate that PAH performs better in practice than its competitive ratio suggests for the evaluated cases. However, its design requires the server to return to the origin before recomputing a route, which introduces additional travel costs in practical settings, especially when requests arrive frequently.

VII. LIMITATIONS

The experimental evaluation is subject to several limitations. First, the experiments are conducted on benchmark-based metric spaces induced from TSPLIB instances, while request sequences are synthetically generated. This design enables precise control over instance characteristics, such as request volume and temporal distribution, and facilitates systematic comparison across algorithms. However, it may not capture all aspects of demand patterns encountered in specific real-world logistics deployments, where request behavior can be influenced by application-specific factors.

Second, the implementation relies on external libraries for graph-related operations. These libraries are used to support correctness and reproducibility. At the same time, they introduce a small overhead due to data structure conversion between library representations and the internal data structures used by our algorithms. As a result, the reported execution times may include minor implementation-related overhead that is not intrinsic to the algorithms themselves.

Finally, the number of requests is limited to $N \leq 17$, since computing offline optimal solutions is computationally expensive. This enables comparison with the offline optimal objective value but restricts evaluation to small instances.

VIII. CONCLUSION AND FUTURE WORK

In this paper, we evaluated competitive algorithms in practical settings to assess their performance in autonomous logistics systems. The experimental results show that competitive algorithms perform well in practice, despite their conservative theoretical performance guarantees. In particular, GTR demonstrates promising practical applicability for autonomous logistics systems such as autonomous vehicles and drone delivery in IoV applications. Its ability to make effective online decisions without relying on historical data makes it well-suited for deployment in dynamic and uncertain environments.

Several directions remain for future work. One important direction is to evaluate these competitive algorithms on real-world data from automated logistics systems, such as electric vehicle or drone delivery, enabling direct comparison with data-driven approaches. Another direction is to extend the problem model with additional practical constraints arising in autonomous logistics systems.

REFERENCES

[1] G. Ausiello, E. Feuerstein, S. Leonardi, L. Stougie, and M. Talamo, "Algorithms for the on-line travelling salesman," *Algorithmica*, vol. 29, no. 4, pp. 560–581, 2001. DOI: 10.1007/s004530010071. Accessed: Jan. 21, 2026. [Online]. Available: https://link-springer-com.uml.idm.oclc.org/article/10.1007/s004530010071.

[2] R. Bellman, "Dynamic programming treatment of the travelling salesman problem," *Journal of the ACM (JACM)*, vol. 9, no. 1, pp. 61–63, Jan. 1962. DOI: 10.1145/321105.321111.

[3] N. Christofides, "Worst-case analysis of a new heuristic for the travelling salesman problem," *Operational Research Forum*, vol. 3, no. 1, p. 20, Mar. 2022. DOI: 10.1007/s43069-021-00101-z.

[4] G. Reinelt, "Tsplib—a traveling salesman problem library," *ORSA Journal on Computing*, vol. 3, no. 4, pp. 376–384, 1991. DOI: 10.1287/ijoc.3.4.376. Accessed: Jan. 21, 2026. [Online]. Available: https://doi.org/10.1287/ijoc.3.4.376.

[5] "Tsplib symmetric dataset," Accessed: Jan. 21, 2026. [Online]. Available: https://www.kaggle.com/datasets/hiimhoanglam/tsplib-symmetric.

[6] A. A. Hagberg, D. A. Schult, and P. J. Swart, "Exploring network structure, dynamics, and function using networkx," in *Proceedings of the 7th Python in Science Conference (SciPy2008)*, G. Varoquaux, T. Vaught, and J. Millman, Eds., Pasadena, CA, USA, 2008, pp. 11–15.

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

Private Information in Public Data - A Forensic Analysis of Unaffected User Image Metadata Provided by Online Platforms

Felix Fischer, Elisabeth Ziesch, Dirk Labudde
Faculty Applied Computer Sciences and Biosciences
Hochschule Mittweida University of Applied Sciences
Mittweida, Germany
e-mail: {fische11 | eziesch | labudde}@hs-mittweida.de

***Abstract*—Metadata provides additional information to an image about its origin or use case. This paper examines how online platforms treat metadata in images. Pictures in the graphic formats of Joint Photographic Experts Group (JPEG), Portable Network Graphics (PNG), Tagged Image File Format (TIFF) and WebP are prepared with numerous metadata using Exchangeable Image File Format (EXIF), Extensible Metadata Platform (XMP) and International Press Telecommunications Council (IPTC). The entire dataset were uploaded and downloaded to different platforms and then evaluated. Forensically relevant metadata, such as copyright, Global Positioning System (GPS) information, camera details and dates, as well as timestamps were examined more closely. This showed that the platforms differed greatly from one another in how they deal with metadata in images. It has not been possible to recognize a uniform handling of metadata on the inspected platforms. The removal of metadata on platforms aligns with the expectations of user privacy.**

Keywords-social media; Privacy; Metadata Publishing; Messaging.

I. Introduction

Every day, millions of users interact with a massive influx of online images through views, ratings and shares. However, visual perception only scratches the surface. Beyond the visual content, pictures could contain hidden private data. Without careful management, publishing these images risks accidentally leaking personal details, such as your identity or exact coordinates. The metadata is shared together with the images without the user being aware of it. With regard to dating platforms, users do not expect to disclose sensitive data. Harassment by stalkers could be an unwanted consequence for users. This can be avoided by thoughtful handling of images by the platforms.

However, this is not the case with photo-sharing platforms, where it is important for users to publish camera-specific information and settings along with the image. Users also do not expect an altering of image data, while using a storage platform. Therefore, a download of a stored image should result in unchanged information. The image should not be distinguishable from the original.

Another perspective arises from the forensic interest in digital image evidence. Information that may be included in the image material provides important data and clues for forensic investigations. Metadata could reveal the location, the photographer, used equipment, used programs, light intensity of the environment and times. This may lead to new subjects or crime scenes.

This study intends to examine the automatic removal or retention of metadata in images on different platforms. For this purpose, metadata is attached to the images, which are then uploaded and downloaded across multiple platforms. The resulting evaluation shows whether online platforms retain, remove or alter metadata in published images and whether users' privacy is compromised. Additionally, the received data can be seen as an upper limit for forensic usage.

The remainder of the paper is organized as follows. Section II explains the formats of image metadata. Section III presents use cases in a forensic context. Section IV describes the methods used in this paper. Section V lists the platforms selected for examination. Section VI presents the findings, and finally, Section VII discusses possible reasons for deviations in our findings.

II. Formats of metadata in images

Metadata is additional attached data to data. It provides secondary information to the stored data to help organizing or understanding the content. They are essential for understanding the data, its storage, preservation and finding data for future use [1].

The Exchangeable Image File Format (EXIF) is a standard format for image metadata developed by Japan Electronic and Information Technology Industries Association (JEITA) [2]. The metadata of an image file contains a variety of information, such as the date and time it was created, all camera settings, and information about the shooting environment, camera manufacturer and model, ISO sensitivity, and horizontal and vertical resolution of the image. Where appropriate, geographical information may also be included. This data is created by the camera or program and then stored in the header of the digital image file. The main purpose of EXIF metadata is to enable precise search, retrieval and viewing of images. This simplifies management and organization, and allows the tracking of the original settings and operations on an image [3]. For example, EXIF metadata can be found in Joint Photographic Experts Group (JPEG), Portable Network Graphics (PNG), Tagged Image File Format (TIFF) and WebP graphic formats [4].

The Extensible Metadata Platform (XMP) format is a standard Adobe format that was introduced in 2001 along with Adobe Acrobat version 5.0 [5]. Based on the Extensible Markup Language (XML) programming language, the metadata format can also be embedded in JPEG, PNG, TIFF, and

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

WebP image file types [6]. It is format-independent and extensible [2]. Primarily, it is used for image processing and gives many possibilities to describe images extensively and flexibly [7]. Such information can include, for example, copyright information, title and descriptions of the image, as well as geographical information, date and time of creation of the image [1].

The International Press Telecommunications Council (IPTC) is a metadata format, which is gradually being replaced by the metadata format XMP [8]. This format is mainly used in journalistic areas and for museum or cultural heritage collections to store information about the author and the content of the image [2]. This includes contact information, title, description, location, creation date and time. This format can be saved in graphic formats, such as JPEG, PNG and TIFF [9].

III. Forensic value of metadata in images

An image itself can be very useful in a crime investigation. Images containing metadata can provide additional context or information about an author, place or creation process. These additional data could be a date or time. Times can enable an ordering in time of images and therefore confirm or deny alibi. Depending on the sun's position in the sky, the casted shadows change as well during the day. By combining time with shadow directions and length, a geographic location will be possible along a longitude [10].

Metadata about the camera equipment can lead to used hardware to create the image. This indicator could also lead to the author of an image. By analyzing the focus length, a distance to the image objects or measurements inside the image can be taken [11]. The exact location of the used camera during the creation of the image will be possible.

Metadata can also contain information about the shooting environment. Lighting of the surrounding effects shutter-speed and ISO sensitivity. Using this, time or weather could be guessed. In combination with weather reports, a location could be roughly determined.

Global Positioning System (GPS) enables a precise localization worldwide. If the GPS location is included in the metadata, a precise positioning of the camera during the creation of the photo can be deduced. This method was used to locate John McAfee in Guatemala after he fled Belize, where a warrant of arrest for murder was issued [12]–[14].

The copyright field in the metadata informs about the applied rights to the image. They can also contain contact information or names. Therefore, these could be used to determine the author of an image. Additionally, a title or description can be provided. The used language can give hints for the location or nationality of the author.

Sometimes programs override fields in the metadata. This could be used to exclude computers as an origin. If a culprit has not installed the used program, he is likely not the author of the image. Furthermore, can the listing of a program in the metadata be a hint for image manipulation.

IV. Methods

Writing metadata to images refers to adding additional information, such as author, date of capture, camera details, and GPS information. Different programms can be used to add and read metadata from images [15]. We use the website IMGonline.com.ua to fill the metadata fields with information in a self created image, showing the second building of our university. The included EXIF editor offers many different functions for managing metadata in images. EXIF, XMP, and IPTC metadata in JPEG images can be viewed, edited, copied, or completely removed in tabular form. For EXIF metadata, there are 59 different metadata fields, such as Software, Model, Last Modified Date, Author, ISO, Flash, Color Space, Exposure Mode, Zoom Factor, Focal Length, Contrast, Saturation, Sharpness, Lens Model and Copyright. There are also several fields to edit for IPTC and XMP metadata. Figure 1 illustrates the full process chain. In the first step, we use IMGonline.com.ua to add as much metadata to the image as possible.

However, these functions can only be applied to JPEG images. In order to convert the metadata in the images to other graphic formats, these must be converted. The conversion of the JPEG image into the graphic formats TIFF and WebP can be done without loss of metadata via websites that are freely accessible on the Internet. When converting to the graphic format PNG, however, all metadata is lost, which is why the image editing software GNU Image Manipulation Programm (GIMP) was used for conversion [16]. Consequently, the JPEG and TIFF graphic formats contain metadata in EXIF, XMP and IPTC formats, PNG metadata in EXIF, XMP and partially IPTC formats, and the WebP graphic format contains metadata in EXIF and XMP formats.

The dataset is then uploaded to the online platforms listed in Section V. Problems may occur, such as errors, when logging in or verifying the account with the platform, generally uploading the record, and problems with the respective graphic formats or file sizes.

We continue by downloading the images just uploaded. Many of the platforms do not offer the possibility to download published images via a download function. To get around this, the "Save As" function of the browser can be used.

After the dataset has been uploaded and downloaded on different platforms, the images are evaluated. Here, the images of the dataset are compared with the images after the upload and download. The different graphic formats are differentiated. When comparing, objective attention is paid to whether and how the metadata in the images has changed.

The software Manja Digital Asset Management is used for the evaluation [17]. It allows to manage, exchange and archive different file types, such as images, audios and videos. With regard to the metadata, Manja Digital Asset Management offers the possibility to automatically recognize and display important image information. In order to compare the images with each other, two images can be selected with the help of this software. The metadata of the two images displayed in tabular form can

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

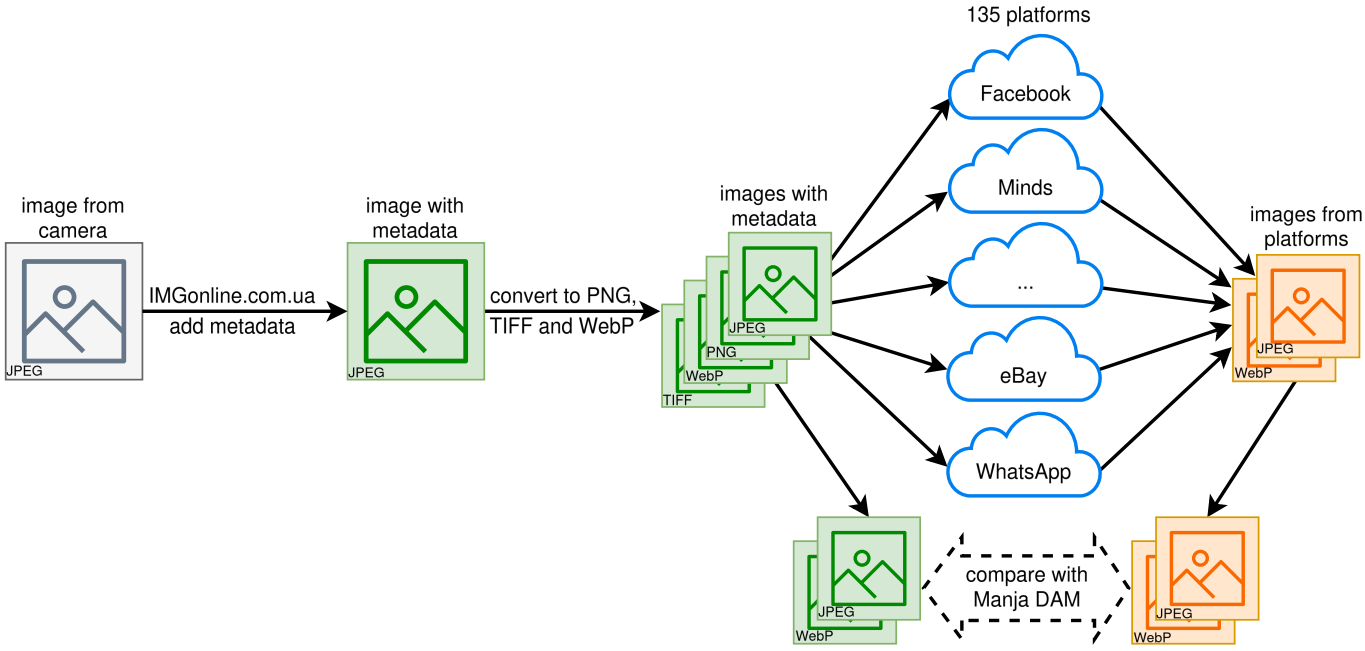


Figure 1. Images are enriched with metadata and compared with the downloaded versions of these images from different platforms. This process is done with the image formats JPEG, PNG, WebP and TIFF. The comparison is realized with the online tool Manja Digital Asset Management.

then be compared - identical information is displayed in black, while non-identical information is highlighted in red.

V. Selected Platforms

The selected platforms for the up and download were divided into different categories: social media, community platforms, dating platforms, photo sharing platforms, marketplaces, messengers, cloud storage and cooking platforms. The selection of these categories allows a wide range of online platforms. These platforms do not represent all platforms that exist in this category. It is not possible to take into account all available platforms due to their scope. Below, the online platforms are listed in their respective categories.

Social Media: Amino, ASKfm, Basistar, Befilo, CloutHub, Demoxia, Facebook Lite, Facebook, FC2, Flickr, Gab, Galleria, GORF, hi5, Instagram, LinkedIn, Mastodon, MeWe, Mixi, Pinterest, Plurk, Reddit, Snapchat, SpinChat, Taringa, TikTok, Tumblr, Twitch, Vero, Wer kennt wen, Whisper, X (formerly known as Twitter), Xing, Yareny, Yooco, YouTube

Community Platforms: 23snaps, AboutMe, Academia, Brainly, Care2, CaringBridge, Crockes, DailyStrength, Diaspora, Elpha, Evernote, girls ask guys, GoodReads, Influenster, LiveJournal, Medium, Minds, NextDoor, Opportunity, Patient Like Me, Polywork, Quanswer, Quora, RallyPoint, ResearchGate, Shutterfly, Slack, Snapfish, Stack Overflow, Travellerspoint, Wattpad, Yelp

Dating Platforms: Badoo, BlackPlanet, Bumble, Canoodle, EliteSingles, LOVOO, Match, MeetMe, OkCupid, Tagged, Zoosk

Foto Sharing Platforms: 500xp, ArtStation, Behance, Crevado, DeviantArt, Dribble, EyeEm, FotoCommunity, GuruShots, Houzz, ImageShack, Imgur, Pexels, PicsArt, Pixaby, PostImage, Raverly, SmugMug, TinyPic, Unsplash, VSCO, YouPic

Marketplaces: Craigslist, eBay, Etsy, Facebook Marketplace, Kleinanzeigen (formerly known as eBay Kleinanzeigen), Shpock, Vinted

Messengers: Discord, Element, Fluffy Chat, Google Chat, HalloApp, ICQ, Line, Messenger (Meta), Messenger von Google, Signal, Skype, Telegram, Viber, VK, Webex by Cisco, WeChat, WhatsApp, Zoom

Cloud Storage: Amazon Photos, Dropbox, Google Drive, Google Photos, Microsoft OneDrive

Cooking Platforms: Chefkoch, Cookpad, Tastemade, Yummly

VI. Results

In total, the dataset was applied to 135 platforms. Of these, 82 removed all metadata from the uploaded images. We list them in Section VI-A. In addition, 19 platforms retained all available metadata associated with the uploaded images. These platforms are listed in Section VI-B. The remaining 34 platforms did not uniformly handle the metadata in the images. Section VI-C explains the handling in more detail. Due to the diversity of the outcomes, we analyze the data from different information perspectives. In Section VI-D we present our findings for the most used platforms in detail. Our findings for platforms that process very sensitive data are presented in Section VI-E. Finally, in Section VI-F we summarize the results.

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

A. Complete removal of metadata

The analysis shows that many platforms completely remove the metadata in the images after uploading and downloading. Therefore, the images after downloading are free of metadata regardless of the graphic format and form. The image formats used are JPEG, PNG, TIFF and WebP in the form of images, profile pictures and banners. In addition, it was found that both the size of the files and the resolution were reduced.

The following platforms remove the metadata completely: Amino, 23snaps, AboutMe, ArtStation, ASKfm, Badoo, Basistar, Befilo, BlackPlanet, Brainly, Bumble, Canoodle, Care2, Chefkoch, CloutHub, Cookpad, Craigslist, Crevado, Crockes, DailyStrength, Demoxia, Diaspora, eBay, EliteSingles, Etsy, Facebook Lite, FotoCommunity, Gab, GORF, GuruShots, HalloApp, hi5, Houzz, Imgur, Influenster, Instagram, Kleinanzeigen (formerly known as eBay Kleinanzeigen), Line, LinkedIn, LOVOO, Mastodon, Match, MeetMe, Messenger von Google, Minds, Mixi, NextDoor, OkCupid, Opportunity, Pexels, Pixaby, Plurk, Polywork, Quora, RallyPoint, Reddit, Shpock, Signal, Snapchat, SpinChat, Stack Overflow, Tagged, Tastemade, Telegram, TikTok, Travellerspoint, Twitch, Unsplash, Vero, Vinted, VK, Wattpad, WhatsApp, Whisper, X (formerly known as Twitter), Xing, Yareny, Yelp, YouPic, YouTube, Yummly, Zoosk.

B. Full retention of metadata

The analysis showed that some platforms retain the metadata of the image completely after uploading and downloading. Images in JPEG, PNG, TIFF, and WebP formats were used as images or profile images. Metadata, such as EXIF, XMP and IPTC have been fully preserved. With the exception of two messengers (*Element* and *FluffyChat*), the GPS information remained in the image. The file size and resolution of the images have not changed significantly. With the exception of four cases where the file size or resolution was reduced, the values remained unchanged.

The following platforms retain the metadata completely: FC2, Academia, Amazon Photos, DeviantArt, Dropbox, Element, Evernote, Flickr, Fluffy Chat, Google Chat, Google Drive, Google Photos, Microsoft OneDrive, Quanswer, SmugMug, Snapfish, TinyPic, Webex by Cisco, Zoom.

C. Partial preservation of metadata

The evaluation yielded distinctive results regarding the partial removal of the metadata. Due to their complexity, different scenarios have been created in which important metadata fields are presented and their occurrence on the respective platforms is explained. The main focus for these results is metadata.

Consideration regarding date and time: Information, such as date and time, can be included in images in different forms as metadata. Images can include the date of creation, date of modification, and date of last modification of the metadata including time. This information can be found mainly in the standard metadata as well as in EXIF, XMP and IPTC.

Providing date and time as metadata can be meaningful. In this way, images can be sorted chronologically and the exact time of capture can be determined. It may also play an important role in criminal cases from a forensic point of view, as it may serve as evidence of certain events or offences. It provides information about whether or not a person was actually present on a certain day and at a certain time. This information may also help in the case of contradictory witness statements, in order to verify the accuracy of the statements. In general, the information serves to support the credibility of digital image evidence.

The following platforms provide date and time information: Behance, CaringBridge, Discord, Elpha, EyeEm, Galleria, girls ask guys, GoodReads, ICQ, ImageShack, LiveJournal, MeWe, Patient Like Me, Pinterest, PostImage, Raverly, ResearchGate, Shutterfly, Skype, Taringa, Viber, Wer kennt wen, Yooco.

Image formats are not treated equally.

Consideration with regard to copyright: The information about the author can be found in the EXIF, XMP and IPTC metadata. There, the name of the author or creator of the image is noted. With the help of this information, the legal distribution and use of the image can be clearly regulated. Images with copyright notice are protected by copyright and allow a clear identification of the author. These images may also help to track the distribution of an image on the Internet.

The following platforms preserve copyright information: Behance, CaringBridge, Discord, Dribble, Elpha, EyeEm, Facebook, Facebook Marketplace, Galleria, girls ask guys, GoodReads, ICQ, ImageShack, LiveJournal, Medium, Messenger (Meta), MeWe, Patient Like Me, Pinterest, PostImage, Raverly, ResearchGate, Shutterfly, Skype, Taringa, Tumblr, Viber, Wer kennt wen, Yooco.

Image formats are not treated equally.

Consideration in relation to the color space: The color space is crucial information for the printing of images. It determines how the colors are displayed in the image and has a significant influence on the print quality and color accuracy. For example, there are the Red Green Blue (RGB) and Cyan Magenta Yellow Key (CMYK) color spaces. In terms of forensic meaning, the color space provides potential information on whether images have been manipulated or not by comparing them with the original color spaces.

The following platforms keep color space information available: 500px, Behance, CaringBridge, Discord, Elpha, Galleria, girls ask guys, GoodReads, ICQ, ImageShack, LiveJournal, MeWe, Patient Like Me, PicsArt, PostImage, ResearchGate, Shutterfly, Skype, Taringa, Tumblr, Viber, VSCO, Wer kennt wen, Yooco.

Image formats are not treated equally.

Consideration of location information: Information about the location where the image was taken may be included in the metadata of an image. This data makes it possible to determine the exact location, as longitude, latitude and altitude can be specified. The location information can be automatically embedded in the image or added manually. This information is extremely valuable from a forensic point of view, as it can provide information about the location of different images or the reconstruction of movements. The metadata can be used to

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

confirm that the image was actually taken at the appropriate location and thus determine the exact location of the event. It can also prove or refute inconsistencies in the statements of persons prosecuted with criminal law about their whereabouts. They can also serve as an alibi by being able to confirm the location of a person at a certain time.

The following platforms provide location information: Behance, CaringBridge, Discord, Elpha, Galleria, girls ask guys, ICQ, ImageShack, LiveJournal, MeWe, Patient Like Me, PostImage, ResearchGate, Shutterfly, Skype, Taringa, Wer kennt wen, Yooco.

Image formats are not treated equally.

Consideration in relation to the camera model: The camera model with which the image was taken is stored in the EXIF metadata of the image. There is information about the brand and model of the digital camera, such as "Canon EOS M200" or "Panasonic Lumix DCGX880". *Canon* is the brand and *EOS M200* is the model. The camera model can be of great importance for forensic purposes, in particular to determine the origin and integrity of captured images. The suspect claims to have certain camera equipment, but the metadata of the images shows a different camera model. This may indicate a false statement or manipulation, or may confirm that certain images were actually taken with the specified camera. This can confirm the credibility of the source.

The following platforms retain camera model information: Behance, CaringBridge, Discord, Elpha, Galleria, girls ask guys, GoodReads, ICQ, ImageShack, LiveJournal, MeWe, Patient, PostImage, Raverly, ResearchGate, Shutterfly, Skype, Taringa, Viber, Wer kennt wen, Yooco.

Image formats are not treated equally.

Consideration to the description of the image: The description of an image is included in the EXIF metadata. These descriptions should reflect the content of the image clearly and precisely, that is, explain the main elements and actions depicted in the image. This information can be embedded directly into the metadata of the image or in other cases added as a text field next to the image. The description of the image plays an important role in the accessibility of the Internet. It captures only the image content and can also serve as alternative text for visually impaired people. This ensures that they are able to fully perceive visual content. The image content is clearly communicated, allowing it to be perceived by a larger number of people.

The following platforms preserve information about the description of the image: Behance, CaringBridge, Discord, Elpha, Facebook, Facebook Marketplace, Galleria, girls ask guys, GoodReads, ICQ, ImageShack, LiveJournal, Messenger (Meta), MeWe, Patient Like Me, PostImage, Raverly, ResearchGate, Shutterfly, Skype, Slack, Taringa, Tumblr, Viber, Wer kennt wen, Yooco.

Image formats are not treated equally.

Protection of images from modification: When images are published on the Internet, some information is partially changed, such as file size, resolution, or metadata in the image itself. However, it is often advantageous to leave the images used unchanged and to maintain their original state. This is important, for example, for photo-sharing platforms where it is important for users to publish camera-specific information and settings along with the image. It is important to keep the metadata in the images.

Protection of the image content itself from modification is given by the following platforms: Academia, Amazon Photos, Behance, CaringBridge, DeviantArt, Discord, Dropbox, Element, Elpha, Evernote, FC2, Flickr, Fluffy Chat, Galleria, girls ask guys, Google Chat, Google Drive, Google Photos, ICQ, LiveJournal, MeWe, Microsoft OneDrive, Patient Like Me, PostImage, Quanswer, ResearchGate, Shutterfly, Skype, SmugMug, Snapfish, TinyPic, Webex by Cisco, Wer kennt wen, Yooco, Zoom.

Image formats are not treated equally.

Protection of personal data: Data protection and personal security require careful handling of personal information in images. To avoid disclosing sensitive location data or copyright information, platforms should remove metadata from published images. While such information is useful for forensic investigations, it does not protect personal rights. With regard to dating platforms, users do not expect to disclose sensitive data.

Platforms do not allow the option of retaining metadata or not. However, the removal may occur automatically if images are converted to other graphic formats without the user being informed.

Protection of personal data is provided by the following platforms: 23snaps, AboutMe, Amino, ArtStation, ASKfm, Badoo, Basistar, Befilo, Behance, BlackPlanet, Brainly, Bumble, Canoodle, Care2, CaringBridge, Chefkoch, CloutHub, Cookpad, Craigslist, Crevado, Crockes, DailyStrength, Demoxia, Diaspora, Dribble, eBay, EliteSingles, Etsy, Facebook Lite, FotoCommunity, Gab, GORF, GuruShots, HalloApp, hi5, Houzz, Imgur, Influenster, Instagram, Kleinanzeigen (formerly known as eBay Kleinanzeigen), Line, LinkedIn, LiveJournal, LOVOO, Mastodon, Match, MeetMe, Messenger (Meta), Messenger von Google, MeWe, Minds, Mixi, NextDoor, OkCupid, Opportunity, Patien Like Me, Pexels, PicsArt, Pinterest, Pixaby, Plurk, Polywork, PostImage, Quora, RallyPoint, Reddit, ResearchGate, Shpock, Shutterfly, Signal, Slack, Snapchat, SpinChat, Stack Overflow, Tagged, Tastemade, Telegram, TikTok, Travellerspoint, Tumblr, Twitch, Unsplash, Vero, Viber, Vinted, VK, Wattpad, WeChat, WhatsApp, Whisper, X (formerly known as Twitter), Xing, Yareny, Yelp, Yooco, YouPic, YouTube, Yummly, Zoosk.

Image formats are not treated equally.

No conversion of graphic formats: Many platforms convert uploaded images to consistent graphic formats for a variety of reasons. On the one hand, the conversion removes some sensitive metadata from the images in order to protect privacy. On the other hand, platforms require certain requirements for graphics formats. Therefore, the images are converted accordingly to make them compatible. This creates uniformity in the graphics formats used and the image quality can be

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

adapted to the requirements of the platforms. It also improves the loading times on websites.

The following platforms do not convert images to other graphic formats: 500xp, Academia, Amazon Photos, ASKfm, Basistar, CaringBridge, Crockes, DeviantArt, Diaspora, Discord, Dropbox, Element, Elpha, Evernote, FC2, Flickr, Fluffy Chat, Google Chat, Google Drive, Google Photos, hi5, ICQ, LiveJournal, Mastodon, MeetMe, MeWe, Microsoft OneDrive, Patient Like Me, Pixaby, RallyPoint, Raverly, Skype, Slack, SmugMug, Snapfish, Stack Overflow, TinyPic, Travellerspoint, Unsplash, VSCO, Webex by Cisco, Xing, Yooco, Zoom.

D. Forensic value of popular platforms

Continuing, the ten most popular platforms by active user count in our dataset will be examined in detail [18].

Facebook: It is possible to upload images in the data format JPEG, PNG, TIFF and WebP. Regardless of the used image format, downloaded images will only be available as JPEG. Metadata stored using the XMP standard will be removed completely. However, copyright, author and description will remain in images using the EXIF standard. While utilize the IPTC standard, images will keep information about copyright. The image contains additionally metadata of the e-mail field in IPTC if the uploaded image uses JPEG, PNG or TIFF as image format.

Instagram: Images can be uploaded using JPEG or PNG format. However, PNGs will be converted to JPEGs as well. Images do not contain any metadata after download. Therefore, no forensic value can be expected in the metadata of images provided by Instagram.

WhatsApp: WhatsApp accepts uploads of images as JPEG, PNG, WebP. Regardless of the uploaded format, JPEG will be chosen for download. A conversion happens for other formats to JPEG. The Messenger removes all metadata in the images. Therefore, no forensic value can be expected in the metadata of images of WhatsApp.

YouTube: On YouTube, the profile image was tested. Here is an upload possible as JPEG, PNG and WebP. Downloading the image, it will be converted to JPEG. The resulting image will not contain any metadata at all. Therefore, no forensic value can be expected in the metadata of profile images of YouTube.

TikTok: JPEG, PNG and WebP can be chosen as image format for uploads. But downloaded images will always use JPEG as image format. A conversion happens during the upload or download process. Images provided by the platform will have all metadata removed. Therefore, no forensic value can be expected in the metadata of images from TikTok. Additionally, alters TikTok the image by adding a gray bar at the top and a TikTok watermark.

WeChat: PNG, JPEG, and WebP have been uploaded to WeChat. Downloading results in a conversion to JPEG. Only when using the JPEG image format, the information about the resolution remains in the EXIF metadata. Everything else will be deleted. The resolution is redundant to inspecting the image data itself, so no additional value is provided. Therefore, no forensic value can be expected in the metadata of images provided by WeChat.

Telegram: Downloaded images on Telegram will always use the JPEG format regardless if JPEG, PNG or WebP are used as the original image format. All metadata will be removed. Therefore, no forensic value can be expected in the metadata of images provided by Telegram.

Messenger (Meta): The messenger from Meta supports uploads with JPEG, PNG, WebP and TIFF format. In downloading these images, a conversion to a different format can happen. TIFF will be converted to JPEG and WebP becomes GIF. Using the EXIF format, copyright and description are included in the metadata. While using IPTC copyright and the E-Mail are available. All fields in the XMP metadata will be deleted. Like in the conversion to a different image format, WebP handles metadata different as well. Images uploaded as WebP will have all metadata removed. Therefore, the available metadata depends on the used image format. Copyright information could lead to the creator of the image.

Snapchat: The upload of images to Snapchat was possible using the image formats JPEG, PNG and TIFF. Images provided by the platform will always use the image format JPEG. These images hold no metadata. Therefore, no forensic value can be expected in the metadata of images from Snapchat.

Reddit: The platform Reddit accepted uploads using JPEG, PNG and WebP as image formats. However, the used image format, images provided by Reddit will always use JPEG. A conversion is done when using PNG or WebP image format. The provided images contain no information in the metadata anymore. Therefore, no forensic value can be expected in the metadata of images provided by Reddit.

E. Platforms with sensitive private data

Analyzing the data, some platforms handle sensitive private data or data intended to share with a small target group. Here, dating platforms and cloud storage providers stand out. The results from the most popular platforms will be presented in detail.

LOVOO: LOVOO allows upload only as JPEG. Consequently, downloads use JPEG as the format as well. No format conversion takes place. LOVOO removes all metadata to protect users from stalkers. Therefore, no forensic value can be expected in the metadata of images from this platform.

Badoo: PNG and JPEG are used as image formats for uploads. However, PNGs will be converted to JPEG. Badoo alters the image content by adding a Badoo watermark. The dating platform Badoo removes all metadata for the protection of its users. Therefore, no forensic value can be expected in the metadata of images provided by Badoo.

Bumble: Upload using the JPEG or PNG image format is possible. Regarding of the used format, a downloaded image will use WebP. Strangely, an upload as WebP is not allowed. Consequently, a conversion is mandatory. The image has no metadata remaining. Users are protected regarding the metadata. Therefore, no forensic value can be expected in the metadata of images from Bumble.

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

BlackPlanet PNG and JPEG can be used as image formats for uploading to the dating platform BlackPlanet. Downloaded images will always use PNG as the image format. Received images will not contain any metadata. Additionally, the images downloaded using our method are just black. It is the only platform we could observe doing that. Therefore, no forensic value can be expected in images from BlackPlanet.

Amazon Photos: The image cloud storage of Amazon accepts all image formats used in this analysis. The downloaded images will be identical to the uploaded ones. Therefore, all images will contain the metadata available during the upload process. Consequently, images of this platform can yield high forensic value.

Dropbox: This cloud storage provider accepts all image formats used in this analysis. The downloaded images will be identical to the uploaded ones. Therefore, all images will contain the metadata available during the upload process. Consequently, images of this platform can yield high forensic value.

Google Drive: This cloud storage service from Google accepts all image formats used in this analysis. The downloaded images will be identical to the uploaded ones. Therefore, all images will contain the metadata available during the upload process. Consequently, images of this platform can yield high forensic value.

Google Photos: The image cloud storage from Google accepts all image formats used in this analysis. The downloaded images will be identical to the uploaded ones. Therefore, all images will contain the metadata available during the upload process. Consequently, images of this platform can yield high forensic value.

Microsoft OneDrive: This cloud storage service accepts all image formats used in this analysis. The downloaded images will be identical to the uploaded ones. Therefore, all images will contain the metadata available during the upload process. Consequently, images of this platform can yield high forensic value.

F. Summary

The process of uploading and downloading images in different graphic formats with a lot of metadata on different platforms is intended to illustrate how the respective platforms handle metadata. It was found that most platforms remove the metadata from the images, regardless of the graphic format. Out of a total of 135 platforms, this corresponds to a share of *60.74%*. Of these, about *25.07%* of the platforms retain some of the metadata in the images, depending on the graphic format. The smallest proportion, about *14.07%* of the platforms, retained the metadata in the images completely.

No similarities in graphic formats could be identified, regardless of whether the metadata was retained or removed. Also, when converting the images to the respective platforms during uploading and downloading, no similarities were found regarding the processing of metadata.

With regard to the platform categories and their handling with metadata, except three categories, no commonalities could be identified. In the cooking category, it was found that all metadata in the images is completely removed from all platforms. The same applies to dating platforms. In this regard, it is important that all metadata be removed. Users expect that sensitive private information will not be disclosed to other users. The opposite is cloud storage, with all platforms retaining the metadata completely. Here, users intend to download images in the same state as they were uploaded. Changing the metadata is undesirable.

Basically, the majority of platforms remove metadata from published images. This is a positive step towards protecting personal information. Insofar as the data is needed for digital forensics and law enforcement, the results show that the complete and partial removal of the metadata reduces available information. Accordingly, the information from the visual content of the images is more decisive and usable.

The forensic value is therefore minuscule for most platforms and highly dependent on the platform itself. A broad range of information can be expected on cloud storage platforms. The available metadata in this analysis represents an upper limit of available information. Especially copyright information that has been left untouched often will not be filled in by the daily use of non-photo-enthusiastic users. In this case, the field would also not yield any usable information.

VII. Conclusion and Future Work

In the following, potential sources of error that could have affected the results of this work or could lead to divergent results in comparative future work are identified.

Due to the increasing number of privacy laws and policies, the information provided by each platform varies. The laws in different countries also differ from one another. Therefore, a consistent presentation of the results over the long term cannot be guaranteed. For example, the handling of metadata in images may be specified in the terms of use or privacy policies of the respective platforms.

Furthermore, not all available public platforms were used for this work. Care was taken to make a selection from different areas. During the upload and download process, some platforms found that it is not possible to upload images. As a result, only 135 of the original selected 227 platforms could be used. Therefore, it is not possible to make general statements for the individual platform categories.

There is a need for optimization with regard to the graphic formats. The TIFF format was not compatible when used on different platforms and therefore often could not be used. Using a more appropriate graphic format would lead to more consistent and better results.

The process of uploading images to different platforms is often complex. On some platforms, images must be manually released or uploaded before being uploaded. Therefore, no statements can be made about those platforms, even if they allow the uploading and downloading of images.

Another point is that the process of uploading and downloading for web based platforms was done only by the same browser (in this case: Microsoft Edge). It is possible that the use of other

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

browsers could lead to different results. For mobile apps, a virtual device used the apps provided by the Google app store. The well-known messenger *WhatsApp* and the well-known social network *Snapchat* make it possible, among other things, to use the application not only as an app on a smartphone, but also as a web application on a computer. However, these web applications of the platforms were not taken into account in this study, which could also lead to altered results.

The dataset was uploaded and downloaded from Germany, Saxony. Therefore, other countries where the uploading and downloading might lead to different results have not been considered.

The images were downloaded immediately after the image were uploaded. Therefore, no statement can be made about how the metadata behaves when downloading older images. It is unclear whether the platforms used to handle metadata in images differently than they do today. Therefore, the results are only relevant for the current time and can not make statements about the previous handling of metadata by the platforms.

As far as messengers are concerned, the images in this study were shared exclusively in private chats. However, there is the option to extend the upload and download process to group chats. In addition, it remains unclear how the metadata in the images behaves when received in direct or group chats.

Many messengers and social networks offer their users the opportunity to share images both as posts and as stories or statuses. For example, *WhatsApp* and *Instagram* allow sharing images with other users in short sequences for 24 hours. This behavior was not taken into account in the present investigation.

Finally, the content of the image could be subject of a future investigation. TikTok for example adds a watermark to the image. Hidden watermarks or other modifications could be made by other platforms as well.

In conclusion, the majority of platforms remove all metadata from images shared publicly on social networks. No similarities could be identified in terms of graphic formats, different conversion processes or platform categories. The handling of metadata is very individual from platform to platform. The removal of metadata on platforms aligns with the expectations of user privacy. A huge violation could not be found.

ACKNOWLEDGMENT

This paper was funded by the European Union and the Free State of Saxony (Germany).

Kofinanziert von der Europäischen Union

Diese Maßnahme wird mitfinanziert durch Steuermittel auf der Grundlage des vom Sächsischen Landtag beschlossenen Haushaltes.

REFERENCES

[1] R. Grunzke et al., "The masi repository service — comprehensive, metadata-driven and multi-community research data management", *Future Generation Computer Systems*, vol. 94, pp. 879–894, 2019, ISSN: 0167-739X. DOI: https://doi.org/10.1016/j.future.2017.12.023.

[2] G. N. Hill and M. A. Whitty, "Embedding metadata in images at time of capture using physical quick response (qr) codes", *Information Processing & Management*, vol. 58, no. 3, p. 102 504, 2021, ISSN: 0306-4573. DOI: https://doi.org/10.1016/j.ipm.2021.102504.

[3] H.-C. Huang and W.-C. Fang, "Metadata-based image watermarking for copyright protection", *Simulation Modelling Practice and Theory*, vol. 18, no. 4, pp. 436–445, 2010, Modeling and Simulation Techniques for Future Generation Communication Networks, ISSN: 1569-190X. DOI: https://doi.org/10.1016/j.simpat.2009.09.004.

[4] P. Harvey, *Exif-tags*, 2026. Accessed: Feb. 2, 2026. [Online]. Available: https://exiftool.org/TagNames/EXIF.html.

[5] N. L. Romero et al., *Recovery of descriptive information in images from digital libraries by means of EXIF metadata*. Emerald Group Publishing Limited, 2008, p. 3. DOI: https://doi.org/10.1108/07378830810880388.

[6] P. Harvey, *Xmp-tags*, 2025. Accessed: Feb. 2, 2026. [Online]. Available: https://exiftool.org/TagNames/XMP.html.

[7] PDFlib GmbH, *Überblick über xmp (overview of xmp)*, 2023. Accessed: Feb. 2, 2026. [Online]. Available: https://www.pdflib.com/de/pdf-know-how/xmp/ueberblick-ueber-xmp/.

[8] T. Hoffmann-Walbeck et al., *Standards in der Medienproduktion (Standards in media production)*. Springer Vieweg Berlin, Heidelberg, 2013, pp. 82–85, ISBN: 978-3-642-15042-5. DOI: https://doi.org/10.1007/978-3-642-15043-2.

[9] P. Harvey, *Iptc-tags*, 2020. Accessed: Feb. 2, 2026. [Online]. Available: https://exiftool.org/TagNames/IPTC.html.

[10] P. Kakar and N. Sudha, "Authenticating image metadata elements using geolocation information and sun direction estimation", in *2012 IEEE International Conference on Multimedia and Expo*, 2012, pp. 236–241. DOI: 10.1109/ICME.2012.82.

[11] M. L. Heuschkel, F. Fischer, and D. Labudde, "Objektive bildqualitätsmessung in der digitalforensik : Mtf-basierte quantifizierung der bildqualität und erzielbarer messgenauigkeit für forensische bildanalysen (lenses image quality measurement in digital forensics : Mtf-based quantification of image quality and achievable measurement accuracy for forensic image analysis)", de, *NextGen Scientific Review*, no. 4, pp. 17–24, 2026, ISSN: 2940-0929. DOI: 10.48446/opus-16388.

[12] A. Liou, *John mcafee is trapped in paradise*, 2012. Accessed: Feb. 18, 2026. [Online]. Available: https://www.vice.com/en/article/aee3gj/john-mcafee-silicon-valleys-greatest-villain-is-trapped-in-paradise.

[13] Lyle, *Vice.com john mcafee exclusive reveals his location in iphone exif data*, 2017. Accessed: Feb. 18, 2026. [Online]. Available: https://www.mobileprivacy.org/2012/12/vice-com-publishes-exclusive-with-john-mcafee-reveals-location-in-iphone-metadata-exif/.

[14] A. Sawall, *John mcafee mit iphone-geolocation geortet (john mcafee located by iphone geolocation)*, 2012. Accessed: Feb. 18, 2026. [Online]. Available: https://www.golem.de/news/vice-john-mcafee-mit-iphone-geolocation-geortet-1212-96131.html.

[15] F. Fischer, E. Ziesch, and D. Labudde, "Erfassungssoftware für metadaten in bildern (extraction software for metadata in images)", de, *Polizeiinformatik*, no. 2024, pp. 209–218, 2024.

[16] GIMP, *Gnu image manipulation program*, https://www.gimp.org, 2026.

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

[17] R. Frunzke, *Manja digital asset management*, https : / / manjadigital.de/, 2026.

[18] D. Thuy, *Most popular social networks worldwide as of february 2025, by number of monthly active users*, 2025. Accessed: Feb. 2, 2026. [Online]. Available: https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/.

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org