https://doi.org/10.2352/EI.2026.38.3.MOBMU-324

Overview and State-of-the-Art in Printer Forensics

Nikola Nachevski [1,2], ***Rifqi Ardia Ramadhan*** [1,2], ***Panharith An*** [1,2], ***Rana Shafi*** [1,2], ***Reiner Creutzburg*** [2,3,4]

[1] Kadir Has University, Department of Administrative 34083 Cibali/Fatih, İstanbul, Turkey

[2] SRH University, Berlin School of Technology and Architecture, Sonnenallee 221, D-12059 Berlin, Germany

[3] German University of Digital Science; Marlene-Dietrich Allee 14, D-14772 Potsdam

[4] Technische Hochschule Brandenburg, Department of Informatics and Media, Magdeburger Str. 50, D-14770 Brandenburg, Germany

niknacevski@gmail.com, rifqiaramadhan@gmail.com, mr.anpanharith@gmail.com, rana.shafi@stu.khas.edu.tr, reiner.creutzburg@gmail.com, reiner.creutzburg@german-uds.de, reiner.creutzburg@srh.de, creutzburg@th-brandenburg.de

Keywords: Printer Forensics, Document Examination, Printing Technology, Digital Forensics, Source Attribution, Counterfeit Detection, Steganography

Abstract

Printer forensics is a specialized field within digital and document forensics that focuses on identifying the source printer of a printed document through intrinsic and extrinsic characteristics. As printers play a crucial role in both legitimate and malicious activities ranging from document authentication to the dissemination of forged or anonymous materials, the need for robust forensic techniques has become increasingly important. This paper provides a comprehensive overview of the current landscape in printer forensics, including the classification of methods used for source identification, such as mechanical defect analysis, texture pattern recognition, and embedded code detection. Both traditional image processing techniques and recent advancements leveraging machine learning and deep neural networks are examined. Additionally, we explore the challenges associated with dataset availability, print-scan noise, and cross-model generalization. By surveying existing methodologies and the public limitations of current approaches, we identify emerging trends and propose potential directions for future research in the field.

Introduction

In the digital age, printers serve as a vital bridge between electronic and physical information. While much attention in digital forensics has been devoted to networked devices and online activity, the forensic analysis of printed documents, commonly referred to as printer forensics, remains an essential and evolving area within the broader domain of document forensics. Printer forensics is the science of identifying the source printer of a given document or verifying the authenticity of printed materials through the examination of printer-specific artifacts, mechanical signatures, and embedded data [1, 2].

The importance of printer forensics stems from its application in criminal investigations, counterfeit detection, intellectual property protection, and leak tracing. Law enforcement agencies rely on these techniques to attribute threatening letters, fake IDs, or classified leaks to a particular printer model or, in some cases, to a specific device [3]. With advances in desktop publishing and high-resolution printing, it has become easier for malicious actors to forge or manipulate printed documents, necessitating forensic systems that can extract and analyze minute, often imperceptible differences left by individual printers.

The methods employed in printer forensics can be broadly categorized into two groups: intrinsic and extrinsic techniques. Intrinsic methods analyze native characteristics, such as banding patterns, toner distribution, and printer-specific mechanical defects, that are unintentionally embedded in the printed output [4]. In contrast, extrinsic techniques involve the deliberate embedding of identifiers into printed documents, such as machine identification codes (MICs), sometimes known as yellow dot codes, that certain printer manufacturers include for anti-counterfeiting purposes [5].

Recent developments have expanded the use of machine learning and pattern recognition techniques for printer identification. These methods extract statistical features from printed textures, such as those based on gray-level co-occurrence matrices (GLCM), and apply classifiers, such as support vector machines (SVMs) or deep neural networks, to achieve high accuracy in source attribution [6, 7]. For instance, a recent quantum-inspired KNN model by Rajenderan et al. (2025) [8] demonstrates how adaptive classification algorithms can enhance accuracy and robustness against document variability.

Given the increasing sophistication of printing technologies and the persistence of printed media in sensitive and legal contexts, printer forensics remains a critical tool for upholding authenticity, accountability, and security in the physical domain. This paper provides a comprehensive overview of printer forensics, explores state-of-the-art identification techniques, and discusses current challenges and research trends in the field.

Background

Early History of Printer Forensics

The origins of printer forensics trace back to the broader field of forensic document examination, which traditionally focuses on handwriting, typewriting, and analysis. With the commercial introduction of computer printers in the late 1970s and early 1980s, investigators began to recognize that printers, such as typewriters,

left distinctive, traceable marks on documents [9]. Early forensic work primarily used dot-matrix printers, in which the mechanical impact of pins on the ribbon produced identifiable wear patterns over time. By the mid-1990s, as inkjet and laser printers became dominant, forensic focuses shifted to analyzing non-contact printing processes and microscopic printing artifacts [10]

In the late 1990s and early 2000s, research began to formalize methods for source printer identification through image processing and statistical analysis. Mikkilinneni et al. [1] were among the first to demonstrate systematic extraction of texture features from printed documents for printer attribution. Parallel to academic developments, certain printer manufacturers began embedding Machine Identification Codes (MICs), also known as "yellow dot codes," into printed pages for anti-counterfeiting purposes, a practice first publicly disclosed by the Electronic Frontier Foundation (EFF) in 2005 [5]

Types of Printers and Their Forensic Characteristics

Understanding the different printing technologies is critical to appreciating their forensic signatures:

- **Dot-Matrix Printers** - Use an array of pins striking an ink ribbon, producing mechanical impressions. Forensic analysis often focuses on pin alignment, wear patterns, and ribbon characteristics [9].
- **Inkjet Printers** – Spray tiny droplets of ink directly onto paper. Key forensic features include droplet size distribution, nozzle defects, and microbanding caused by mechanical feed variations [11].
- **Laser Printers** – Employ an electro-photographic process using a laser beam to charge a drum that attracts toner particles. Forensic artifacts include toner distribution patterns, drum wear, and halftoning signatures [12].
- **Thermal Printers** – Use heat to transfer pigment from a ribbon or activate heat-sensitive paper. Forensics in this category often focuses on ribbon defects, heating element patterns, and chemical composition of thermal coatings [13].

Evolution of Forensic Needs

Initially, printer forensics was used for counterfeit detection and criminal investigations involving threatening letters, ransom notes, or fraudulent documentation. As desktop publishing matured in the 1990s, the ability to create high-quality, forged documents expanded beyond professional print shops to home users [14]. The rise of high-resolution laser and inkjet printers introduced challenges, as the artifacts became subtler and harder to detect with traditional optical examination.

By the 2010s, the digitalization of investigative workflows and the growth of cybercrime introduced scenarios in which printed documents were part of hybrid digital-physical evidence chains [6]. Today, forensics needs extend beyond identifying the printer model; they often require linking a specific physical device to a document, even after multiple generations of scanning and reprinting. Additionally, adversarial techniques, such as removing or spoofing MICs and intentionally altering printing patterns, have driven the development of machine learning, deep learning, and even quantum-inspired approaches for robust attribution [8]

Source Identification Methods in the Literature

Source printer identification aims to attribute a printed document to its originating printer by exploiting systematic, device-specific artifacts introduced during printing. Over the past two decades, a wide range of intrinsic and extrinsic techniques have been proposed in the literature, evolving from handcrafted feature extraction and statistical classification to modern deep learning and hybrid approaches. This section provides a structured overview of the principal categories of source identification methods, highlighting their underlying principles, representative techniques, and known limitations.

Character-Based Texture Analysis

One of the earliest and most influential classes of intrinsic printer identification techniques is based on character-level texture analysis. Mikkilinneni et al. [1] pioneered this approach by demonstrating that individual printers leave distinctive texture patterns in printed characters due to microscopic variations in toner deposition, ink spread, and mechanical motion. In this framework, frequently occurring characters—most notably the lowercase letter "e" in English documents—are segmented from scanned pages, and texture features are extracted from their grayscale representations. Commonly used descriptors include gray-level co-occurrence matrices (GLCM), run-length statistics, and edge-direction histograms. These features are subsequently classified using distance-based or statistical classifiers such as k-nearest neighbors (KNN) or support vector machines (SVM).

Character-based methods offer high discriminative power under controlled scanning conditions and require relatively small amounts of training data. However, their performance is sensitive to font variation, character segmentation accuracy, and print–scan distortions. Additionally, reliance on specific characters limits applicability to multilingual or non-textual documents.

Global Texture and Micro-Pattern Methods

To overcome the dependency on character segmentation, subsequent research explored global texture-based approaches that analyze larger regions of printed pages or entire documents. These methods aim to capture printer-specific micro-patterns arising from toner granularity, ink diffusion, and halftoning strategies. Features such as local binary patterns (LBP), wavelet coefficients, and frequency-domain representations have been widely adopted for this purpose [3, 11]. By aggregating texture statistics over larger areas, these methods improve robustness to font changes and partial document availability.

Global texture analysis is particularly effective for laser printers, where electro-photographic processes introduce repeatable irregularities in toner distribution. Nevertheless, these approaches remain vulnerable to variations introduced by different scanners, camera-based acquisition, and document rescaling, which can significantly alter the captured texture statistics.

Banding and Geometric Distortion Analysis

Another prominent category of intrinsic techniques focuses on geometric distortions and banding artifacts introduced by mechanical components such as paper feed rollers, photoconductor drums, and print heads. In laser printers, slight inconsistencies in drum rotation or laser scanning can produce periodic horizontal

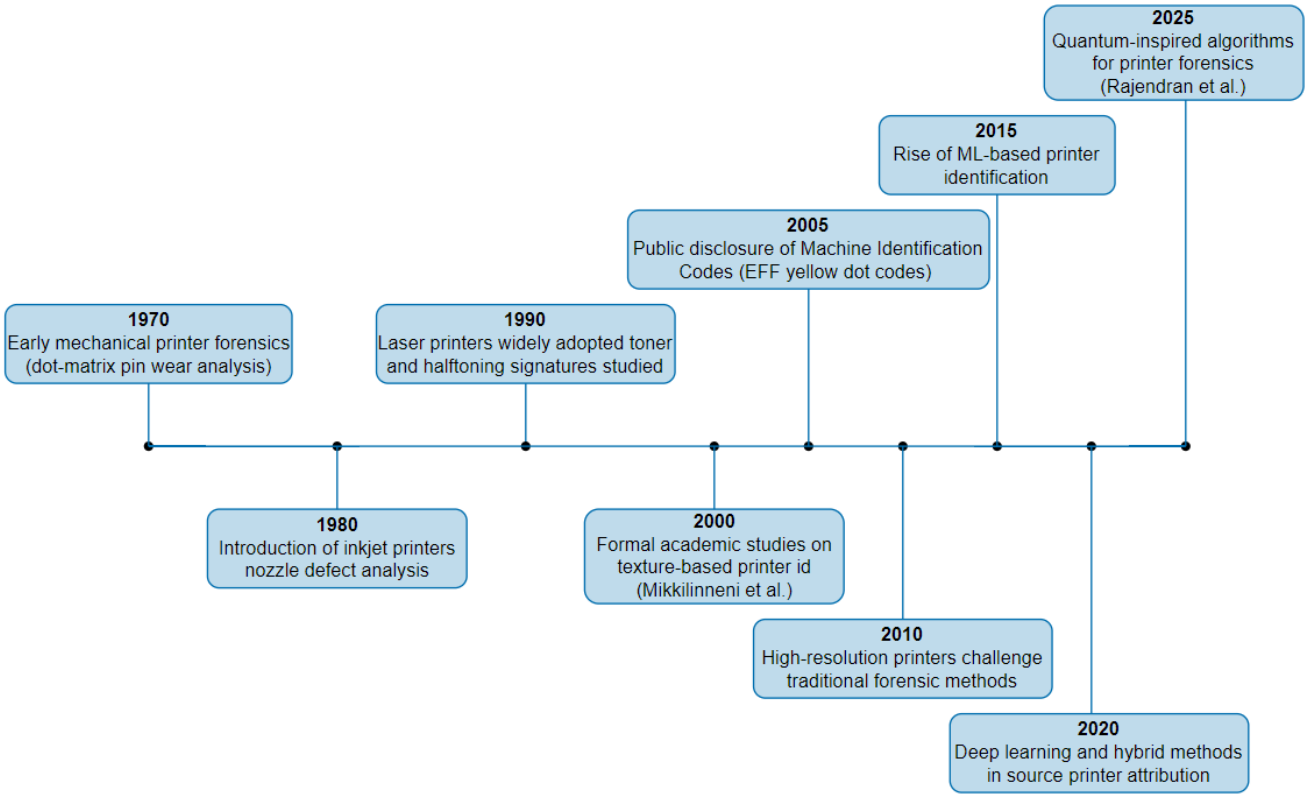


Figure 1: Evolution of Printer Forensics (1970–2025)

or vertical banding patterns that are stable over time and characteristic of individual devices [12]. These patterns are typically analyzed using frequency-domain techniques, including Fourier transforms and autocorrelation analysis.

Geometric distortion-based methods have also been applied to inkjet printers by modeling systematic spatial deviations caused by nozzle misalignment and carriage motion [15]. Such approaches are attractive from a forensic standpoint due to their strong physical interpretability. However, their effectiveness can degrade as printers age, undergo maintenance, or operate under varying environmental conditions, thereby altering the underlying mechanical behavior.

Embedded Code and Extrinsic Identification Techniques

In contrast to passive intrinsic methods, extrinsic techniques rely on deliberately embedded identifiers within printed documents. The most widely known example is the Machine Identification Code (MIC), often referred to as yellow dot codes, which some color laser printers embed as microscopic patterns that encode information such as the printer's serial number and timestamp [5]. These codes were originally designed for anti-counterfeiting and law enforcement purposes and can provide near-deterministic attribution when present and successfully decoded.

Despite their high reliability, extrinsic methods face significant practical and ethical limitations. MICs are manufacturer-dependent, absent in many printer models, and can be intentionally removed or obfuscated through image processing or reprinting. Moreover, their use raises privacy concerns, limiting their availability and applicability in forensic investigations.

Machine Learning and Deep Learning Approaches

Recent advances in machine learning have significantly reshaped research on source printer identification. Classical machine learning methods integrate handcrafted features—such as texture descriptors or banding profiles—with classifiers including SVMs, random forests, and ensemble models [6]. These systems offer improved scalability and classification accuracy compared to early statistical techniques but still depend heavily on feature engineering.

More recently, deep learning approaches, particularly convolutional neural networks (CNNs), have been introduced to automatically learn discriminative printer signatures directly from raw or minimally processed document images [7, 17]. CNN-based methods have demonstrated superior performance on controlled datasets and are resilient to moderate print–scan noise. Extensions incorporating attention mechanisms and transformer-based architectures further enhance the ability to capture long-range spatial dependencies in printed textures.

In parallel, quantum-inspired classifiers, such as the quantum-inspired KNN model proposed by Rajendran et al. [8], represent an emerging research direction. These methods leverage quantum computing concepts—such as state superposition and kernel transformations—to improve class separability in high-dimensional feature spaces. While promising, such approaches remain largely experimental and lack extensive validation across diverse printers and acquisition conditions.

Method Category	Representative Features / Signals	Advantages	Key Limitations
Character-Based Texture Analysis [1]	GLCM, run-length statistics, edge-direction histograms extracted from segmented characters	High discriminative power; low computational complexity; effective with limited training data	Sensitive to font variation, segmentation errors, and print–scan distortions
Global Texture and Micro-Pattern Methods [3, 11]	LBP, wavelet coefficients, frequency-domain texture descriptors	Robust to font changes; applicable to partial documents	Vulnerable to scanner and camera noise; limited scale robustness
Banding and Geometric Distortion Analysis [12, 15]	Periodic banding profiles, spatial distortion fields	Physically interpretable; device-specific signatures	Sensitive to printer aging and maintenance
Extrinsic Embedded Code Techniques [5]	MICs, yellow dot patterns	Near-deterministic attribution	Manufacturer-dependent; privacy concerns
Deep Learning-Based Methods [7, 17]	CNN-learned texture representations	State-of-the-art accuracy	Limited explainability; domain shift sensitivity
Quantum-Inspired Classification [8]	Quantum state mappings, kernel-based distances	Enhanced class separability	Experimental; limited forensic validation

Table 1: Comparison of major source printer identification approaches in the literature, summarizing representative signals, key strengths, and principal limitations relevant to forensic reliability and real-world deployment.

Overall, learning-based techniques achieve state-of-the-art accuracy but introduce new forensic challenges, particularly related to explainability, reproducibility, and generalization to unseen printers. These limitations motivate ongoing research into interpretable models and domain-adaptive learning strategies, as discussed in subsequent sections.

Table 1 summarizes the principal source printer identification approaches discussed in the literature, highlighting their representative features, strengths, and key forensic limitations.

Challenges and Limitations

Despite significant progress in source printer identification, printer forensics continues to face fundamental challenges that limit its robustness, scalability, and forensic reliability in real-world deployments. A primary limitation of intrinsic printer identification techniques lies in their sensitivity to intra-device variability. Factors such as printer aging, maintenance operations, consumable replacements (e.g., toner cartridges or ink), and environmental conditions can alter the microscopic characteristics of printed output over time, leading to reduced consistency across multiple print instances from the same device [15, 16]. This non-stationarity of printer signatures complicates long-term attribution and weakens confidence in forensic conclusions.

Another major challenge arises from the print–scan and print–capture process. Most forensic analyses rely on scanned or camera-captured versions of printed documents, introducing additional sources of distortion, including resampling artifacts, optical blur, illumination variation, and sensor noise. These acquisition-induced effects can significantly obscure or distort printer-specific features, particularly fine-grained texture patterns and banding artifacts, thereby degrading identification performance [17]. Variability across scanners and smartphone cameras further exacerbates this problem, making cross-device generalization difficult.

Generalization to unseen printers and heterogeneous operating conditions remains an unresolved limitation across both traditional and learning-based approaches. Many proposed methods achieve high accuracy under controlled laboratory settings but exhibit notable performance degradation when applied to printers, paper types, or acquisition devices not represented in the training data. This domain shift poses a critical challenge to forensic applicability, as test conditions are often unknown or adversarial.

From an operational and legal standpoint, explainability and reproducibility constitute additional limitations. Classical handcrafted-feature methods offer some physical interpretability but often lack robustness, whereas modern deep learning approaches achieve superior accuracy but lack transparency. The black-box nature of many neural models raises concerns regarding evidentiary admissibility, as forensic conclusions must be explainable, repeatable, and defensible under legal scrutiny [16]. Finally, the scarcity of standardized, publicly available benchmark datasets limits reproducibility, impedes fair comparison of methods, and slows cumulative progress in the field.

Current Trends and Open Problems

In response to the limitations of classical forensic techniques, recent research in printer forensics has increasingly shifted toward data-driven and learning-based methodologies. Deep neural networks, particularly CNNs, have become a dominant trend due to their ability to automatically learn discriminative printer signatures from scanned and camera-captured documents without explicit feature engineering [17]. These models have demonstrated improved robustness to moderate noise and variability compared to handcrafted feature-based approaches, especially under controlled acquisition conditions.

Building upon CNN architectures, recent studies have explored attention mechanisms and transformer-based models to capture long-range spatial dependencies and subtle global patterns in printed textures. Such hybrid CNN–attention frameworks aim to improve discrimination among visually similar printers and to address limitations in local feature extraction. In paral-

lel, quantum-inspired classifiers and optimization techniques have emerged as a novel and exploratory research direction. By leveraging concepts such as quantum state representations and kernel-based mappings, these methods seek to enhance class separability in high-dimensional feature spaces [8]. While promising, quantum-inspired approaches remain largely experimental and lack comprehensive validation across diverse printers, document types, and acquisition scenarios.

Despite these advances, several open problems persist. Chief among them is the challenge of developing models that generalize reliably to unseen printers, heterogeneous acquisition devices, and unconstrained environmental conditions. Domain adaptation, transfer learning, and cross-modal learning remain underexplored in printer forensics. Furthermore, the trade-off between classification accuracy and forensic interpretability remains unresolved. There is a growing need for explainable and transparent identification systems that combine the performance of deep learning with the evidentiary requirements of forensic science.

Another open problem concerns preparedness for emerging and non-traditional printing technologies, including advanced inkjet mechanisms, multifunction devices, and hybrid digital–physical document workflows. Existing forensic models are largely tailored to conventional laser and inkjet printers, leaving a gap in attribution capabilities for newer and evolving technologies. Consequently, while neural networks and quantum-inspired classifiers represent significant current trends in printer forensics, the development of interpretable, generalizable, and legally robust methodologies remains an open research problem requiring sustained investigation.

Conclusion

Printer forensics has evolved into a critical component of document and digital forensics, enabling the attribution and authentication of printed materials in legal, investigative, and security-sensitive contexts. This paper presents a comprehensive overview of the field, systematically examining the progression from early character-based texture analysis and mechanically interpretable intrinsic signatures to contemporary learning-based and quantum-inspired identification methods. By organizing existing approaches according to their underlying forensic principles, this survey highlights both the strengths and inherent limitations of current source printer identification techniques.

Despite notable advances in classification accuracy, the analysis reveals that many existing methods remain constrained by sensitivity to intra-device variability, print–scan and print–capture distortions, and limited generalization to unseen printers and acquisition conditions. While deep neural networks and hybrid attention-based architectures have demonstrated superior performance under controlled settings, their black-box nature poses significant challenges for forensic transparency, explainability, and legal admissibility. Similarly, emerging quantum-inspired classifiers offer promising theoretical advantages but currently lack sufficient empirical validation and operational maturity for forensic deployment.

The survey further underscores the absence of standardized benchmark datasets and evaluation protocols as a persistent obstacle to reproducibility and fair comparison across studies. Addressing this gap is crucial for transitioning printer forensics from laboratory-focused experimentation to a robust, real-world application. In addition, the rapid evolution of printing technologies and hybrid digital–physical workflows necessitates adaptive forensic models capable of handling heterogeneous devices, acquisition modalities, and adversarial manipulation.

Future research in printer forensics must therefore prioritize the development of interpretable, domain-robust identification frameworks that balance accuracy and evidentiary reliability. Promising directions include explainable machine learning models tailored to forensic requirements, cross-domain and transfer learning strategies for improved generalization, and collaborative efforts toward open, large-scale benchmark datasets. By aligning methodological innovation with legal and operational constraints, printer forensics can continue to mature as a reliable and defensible discipline within modern forensic science.

Acknowledgments

The European Union partially supported this work through ERASMUS MUNDUS, Project CyberMACS (Project No. 101082683, `https://cybermacs.eu`).

References

[1] A. K. Mikkilineni, A. Chawla, and V. Monga, "Printer identification based on texture features of printed characters," Proc. IEEE ICASSP, 2005.

[2] T. Gloe, E. Franz, and R. Böhme, "Forensics for flatbed scanners," Proc. ACM Symposium on Applied Computing, 2010, pp. 1975–1981.

[3] C. Abhayaratne and K. Seneviratne, "Printer identification based on local texture patterns," Journal of Electronic Imaging, vol. 28, no. 1, pp. 013017, 2019.

[4] M. Bulacu, L. Schomaker, and L. Vuurpijl, "Writer identification using edge-based directional features," Pattern Recognition, vol. 41, no. 12, pp. 536–546, 2009.

[5] Electronic Frontier Foundation, "Printer Tracking Dots," 2015. [Online]. Available: `https://www.eff.org/issues/printers`

[6] H. Ryu, H. Kim, and H. K. Lee, "Source printer identification using scalable and interpretable CNNs," Proc. IEEE WIFS, 2017.

[7] C. Jiao, Y. Duan, and X. Su, "Printer identification based on CNN with self-attention mechanisms," Pattern Recognition Letters, vol. 168, pp. 59–66, 2023.

[8] A. Rajendran, A. Kumar, M. Singh, and A. Jaiswal, "Quantum-inspired K-nearest neighbors classifier for enhanced printer source identification in forensic document analysis," Scientific Reports, vol. 15, no. 1, 2025.

[9] R. A. Huber and A. M. Headrick, "Handwriting Identification: Facts and Fundamentals. CRC Press," 1999.

[10] L. Spitz and H. R. Rushing,"The Science of Forensic Document Examination. CRC Press," 2006.

[11] H. Cao and A. Kot,""Printer classification for forensic purposes,"" IEEE Transactions on Image Processing, vol. 16, no. 8, pp. 2076–2085, 2007.

[12] R. Garg, S. Garg, and S. Chaudhary,""Laser printer forensic identification using banding patterns,"" Forensic Science International, vol. 268, pp. 119–128, 2016.

[13] J. M. Kelly,""Thermal printing: Forensic applications,"" Journal of Forensic Sciences, vol. 59, no. 4, pp. 1015–1022, 2014.

[14] K. Khanna and B. P. Singh, "Document counterfeiting and forensic examination of printed documents," Forensic Research & Criminology International Journal, vol. 6, no. 2, pp. 69–76, 2018.

[15] H. Jain et al., "Passive classification of source printers using geometric distortion signatures," arXiv, 2017.
[16] A review on laser printer classification and identification, Forensic Science International, 2025.
[17] Source printer identification using smartphone-captured images, Journal of Information Security and Applications, 2024.

Author Biography

Nikola Nachevski is a master's student in Applied Cybersecurity within the CyberMACS ERASMUS Mundus Joint Master's Degree program at SRH University Berlin and Kadir Has University. He holds a bachelor's degree in Software Engineering from Ss. Cyril and Methodius University in Skopje. His research interests include digital forensics, document and printer forensics, privacy in IoT systems, and the application of machine learning in cybersecurity.

Rifqi Ardia Ramadhan is a master's student in cybersecurity with interests in digital forensics and incident response. Previously, he had worked on security consulting and e-commerce companies as a security engineer before pursuing a, graduate degree. He also managed a cybersecurity community.

Rana Shafi is a Master's student in applied cybersecurity with interests in IoT security and defensive security. Her recent work focuses on privacy preservation in IoT sensor data.

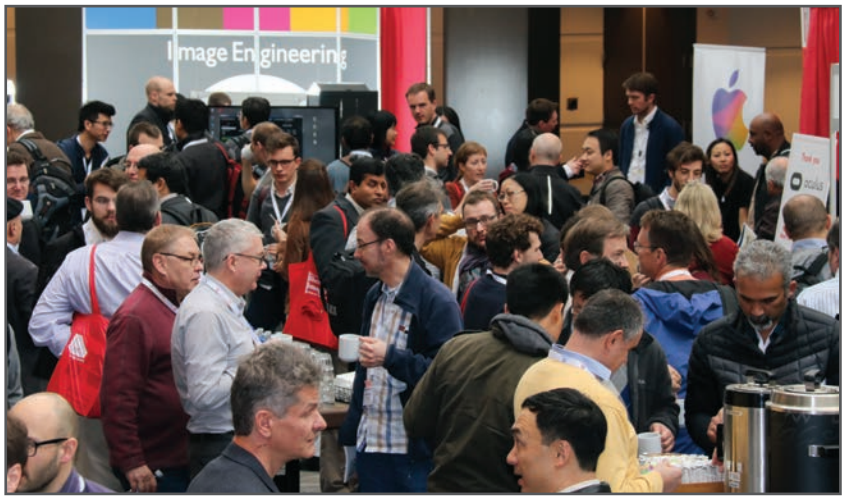
Panharith An earned his BEng in Telecommunication & Electronic Engineering in 2018 and spent five years as a full-stack software engineer in Digital Transformation across the business, banking, and public sectors in Cambodia. He is currently a scholar of the ERASMUS Mundus Joint Master's Degree in Applied Cybersecurity (CyberMACS) at Kadir Has University in Turkey and SRH Berlin University of Applied Sciences in Germany.

Reiner Creutzburg is a Retired Professor of Applied Computer Science at the Technische Hochschule Brandenburg in Brandenburg, Germany. Since 2019, he has been a Professor of IT Security at the SRH Berlin University of Applied Sciences, Berlin School of Technology. He has been a member of the IEEE and SPIE, and has served as the Chairman of the Multimedia on Mobile Device (MOBMU) Conference at the Electronic Imaging conferences since 2005. In 2019, he was elected a member of the Leibniz Society of Sciences to Berlin e.V. His research interests focus on Cybersecurity. Digital Forensics, Open Source Intelligence (OSINT), Multimedia Signal Processing, eLearning, Parallel Memory Architectures, and Modern Digital Media and Imaging Applications.

JOIN US AT THE NEXT EI!

electronic IMAGING

Imaging across applications . . . Where industry and academia meet!



• SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •
• INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •

www.electronicimaging.org

ADVANCING IMAGING SCIENCE 75 1947 IS&T 2022 imaging.org