

LITHUANIAN COMPUTER SOCIETY

VILNIUS UNIVERSITY, INSTITUTE OF DATA SCIENCE AND DIGITAL TECHNOLOGIES

LITHUANIAN ACADEMY OF SCIENCES



**16th Conference on**

# **DATA ANALYSIS METHODS for Software Systems**

---

**November 27–29, 2025**

---

**Druskininkai, Lithuania, Hotel “Europa Royale”**

<https://www.mii.lt/DAMSS>

VILNIUS UNIVERSITY PRESS

Vilnius, 2025

**Co-Chairs:**

Dr. Saulius Maskeliūnas (Lithuanian Computer Society)

Prof. Gintautas Dzemyda (Vilnius University, Lithuanian Academy of Sciences)

**Programme Committee:**

Dr. Jolita Bernatavičienė (Lithuania)

Prof. Juris Borzovs (Latvia)

Prof. Janusz Kacprzyk (Poland)

Prof. Ignacy Kaliszewski (Poland)

Prof. Bożena Kostek (Poland)

Prof. Tomas Krilavičius (Lithuania)

Prof. Olga Kurasova (Lithuania)

Assoc. Prof. Tatiana Tchemisova (Portugal)

Assoc. Prof. Gintautas Tamulevičius (Lithuania)

Prof. Julius Žilinskas (Lithuania)

**Organizing Committee:**

Dr. Jolita Bernatavičienė

Prof. Olga Kurasova

Assoc. Prof. Viktor Medvedev

Laima Paliulionienė

Assoc. Prof. Martynas Sabaliauskas

Prof. Povilas Treigys

**Contacts:**

Dr. Jolita Bernatavičienė

*jolita.bernatavicienne@mif.vu.lt*

Prof. Olga Kurasova

*olga.kurasova@mif.vu.lt*

Tel. (+370 5) 2109 315

Copyright © 2025 Authors. Published by Vilnius University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

<https://doi.org/10.15388/DAMSS.16.2025>

ISBN 978-609-07-1200-9 (digital PDF)

© Vilnius University, 2025

# Improving Malware Detection by Analyzing Similarities of Multi-Category Benign Software

Juozapas Rokas Čypas, Juozas Dautartas,  
Olga Kurasova, Viktor Medvedev

Institute of Data Science and Digital Technologies  
Vilnius University

*juozapas.cypas@mif.vu.lt*

In today's digital world, the importance of cybersecurity is increasing rapidly. The evolution of technology and AI enables various threat actors to evolve malware as well as the methods of evading malware detection. In this study, we aim to analyze modern malware evasion methods presently used in the wild. It's important to identify such methods so that they can be analyzed and studied by security researchers for the purpose of improving the defense infrastructure of software systems. Traditionally, machine learning based analysis of Windows Portable Executable (PE) file static features uses datasets that have either two classes (benign or malicious), or multiple malware classes (e.g., worms, Trojans, ransomware, spyware), and one benign software class. Our proposed method and dataset (DOI:10.18279/MIDAS.265677) focus on the analysis of multiple categories of benign software (office tools, security, media, etc.) and just one class of malware. These categories are used to train a classifier that can distinguish benign software based on its static features. Our concept is that for a given malware sample, it is possible to identify a benign category corresponding to the lowest expected detection rate. This approach analyzes the similarity between each malware sample and multiple categories of benign software. For each instance of malware, we identify the closest benign category and then inject the most characteristic static features of that benign cluster into the malware sample to trick the classifier and maximize the evasion rate. Preliminary results indicate that, after injecting features from the category of benign files selected based on similarity, the initial classifier demonstrates a decrease in the detection rate of such concealed malware in comparison to the original malware file.

**Acknowledgments.** This project has received funding from the Research Council of Lithuania (LMTLT), agreement No S-MIP-24-116.