

STATEMENT

on a PhD thesis for obtaining an educational and scientific degree "doctor"

Dissertation Title: **ANALYSIS OF LEAKED CRITICAL DATA TO IDENTIFY POTENTIAL INFORMATION SECURITY THREATS**

Professional Field: **5.13. General Engineering**

PhD Program: **Computer Technologies in Engineering**

Author of the Dissertation: **MSc. Eng. Ivan Lyubomirov Drankov**

Chair of the Scientific Jury: **Assoc. Prof. Eng. Yordanka Anastasova, PhD**

The submitted dissertation comprises 131 pages. It is structured into an Introduction, four chapters, Conclusion, Dissertation Contributions, Publications, and References. The dissertation includes 80 figures and 9 tables.

1. Relevance of the Dissertation Topic from a Scientific and Applied Research Perspective

The problem of critical data leakage is of particular importance for the security of information systems due to the fact that they contain sensitive, confidential or structurally defining information and can be used for subsequent breaches, compromise or unauthorized access.

The purpose of the dissertation is to develop and implement a methodological approach for automated processing, unification and analysis of large volumes of heterogeneous data from leaked or compromised sources for the purpose of profiling, categorization and extraction of potential threats. To achieve this goal, five main tasks have been defined and implemented, the solution of which uses modern approaches and technologies.

The research presented in the dissertation aims to reveal patterns and dependencies that will support the early detection of risks and the development of mechanisms for the prevention of cyber threats.

2. Degree of knowledge of the issue and creative interpretation of the reference material

The author demonstrates an excellent understanding of the research field and provides an appropriate interpretation of the reviewed literature. This is clearly evident from the overview presented in the first chapter of the dissertation concerning existing systems used for leaked critical data analysis. The PhD student also proposes a definition of a critical data analysis system and presents methods for classifying different types of leaked data.

Three main categories of critical data analysis systems are identified: Passive systems; Notification systems; Active systems.

Based on the conducted literature review, both the objective and the research tasks of the dissertation are clearly defined.

The dissertation consists of 131 pages and includes 80 figures and 9 tables. It cites 56 bibliographic sources, including one in Bulgarian and 55 in English. All references are correctly cited and include valid online access information. Most of the references have been published in recent years, demonstrating the author's thorough understanding of current developments in the field of critical data analysis.

3. Correspondence of the chosen research methodology and the set goal and tasks of the PhD dissertation with the contributions achieved

The PhD student has selected a methodology for developing the ASaID system, consisting of five major subsystems: Databases; Database Management System; Data Analyzer; Notification System and Storage System. The implementation of ASaID is based on a modular three-tier client-server architecture comprising a client layer, a server layer, a data layer (shared-nothing database), and auxiliary computational modules.

A key aspect of leaked critical data analysis is the collection and preprocessing of raw data. In the implemented **ASaID Pyformat** solution, this process includes modules for: validation; verification; formatting;

filtering; compression and archiving. The author selected the CSV format because of its compact size and employed the ZFS file system, which supports several compression methods.

The analyses performed include: password analysis; password type identification and decryption; brute-force attack analysis; location identification; triangulation; last known location analysis; document analysis; image analysis and metadata analysis. The purpose of these analyses is to detect potential threats and attacks.

The developed ASAlD system relies on a technology stack encompassing all required technologies, programming languages, frameworks, and software tools. The proposed modular three-tier client-server architecture includes: client layer; clustered server; shared-nothing database and additional computational modules. The client layer has been partially implemented in the prototype and provides controls for operating ASAlD and visualizing the data. The dissertation presents pseudocode illustrating the communication principles between the client and server layers through an API. The data layer is implemented using **MongoDB Sharded Cluster** technology. A notification module has also been developed as a core component of ASAlD. It operates in the background and is implemented in Python within the presented system.

The optimal implementation would be a cloud-based solution; however, the associated costs exceed the requirements of the dissertation. Within the proposed ASAlD architecture, data are first identified and verified using different technologies. The Pyformat module, developed in Python and utilizing an XML configuration file, then performs processing.

Experiments were conducted using approximately **2 TB of real-world data**, converted into CSV format. Password pattern analysis was performed using approximately **4 million real user passwords**, examining password length as well as the most frequently used characters and keyboard keys.

Modules for location analysis, triangulation, and last-known-location identification based on a real case study are also presented, together with domain-based analysis designed to monitor and evaluate leaked data associated with a specific organization. Additional modules for image and metadata analysis are likewise presented.

The ASAlD system automatically analyzes available leaked or compromised databases, classifies the data into categories, associates each category with potential threats, and can simulate attack scenarios for risk assessment and implementation of appropriate security measures.

The final output is an automatically generated analytical report produced by the ASAlD system based on the analysis results.

The comparative evaluation between the developed ASAlD system and existing leaked critical data analysis systems demonstrates that ASAlD incorporates numerous analytical modules capable of processing different data types, automatically identifying relationships, and generating information about potential attacks and vulnerabilities.

The PhD thesis addresses the complex task of designing an integrated solution for investigating and analyzing leaked data, which is particularly important for improving information security - one of the major challenges of today's digital world.

In my opinion, the selected research methodology and implementation fully correspond to the objectives and research tasks formulated in the dissertation, presenting a comprehensive system design and demonstrating a partial implementation of ASAlD.

4. Evaluation of the author's abstract and publications related to the PhD thesis

The PhD thesis abstract complies with the requirements of the Academic Staff Development Act of the Republic of Bulgaria and its implementing regulations. Its structure and content accurately reflect the PhD thesis.

The research results of Eng. Ivan Drankov have been published in three scientific papers, two of which have already been published, while the third has been accepted for publication. All papers were presented at the annual international conference of the University of Mining and Geology. The publications, produced between **2022 and 2026**, correspond directly to the dissertation topic and reflect its achieved results and contributions. One publication is single-authored, while the remaining two are co-authored, with Eng. Drankov serving as the first author. The papers have been published in the **Annual of the University of Mining and**

Geology "St. Ivan Rilski" and are indexed in **Google Scholar**, **CrossRef**, and **ResearchGate**. At present, no citations of these publications have been identified.

The publication record satisfies the requirements for obtaining the educational and scientific degree of Doctor (PhD), and I believe that the publications have received sufficient visibility within the scientific community.

Considering both the number and quality of the publications, I conclude that the obtained results are entirely the author's own work.

Furthermore, I have found no evidence of unreliable data or plagiarism, as also confirmed by the report generated by **StrikePlagiarism**.

5. Assessment of the scientific results and contributions and the degree of personal participation of the PhD student in the developed topic

The author identifies the following contributions, which I fully accept.

Scientific and Applied Contributions:

1. Development of the **Pyformat** software product for processing large volumes of leaked critical data.
2. Comparative analysis of different analytical approaches and their operating principles for leaked critical data.
3. Formulation and analysis of the principal stages involved in processing large-scale leaked critical datasets.

Applied Contributions:

1. Development of a conceptual model of an **Active System for Analysis and Notification of Leaked Data (ASAIID)**.
2. Design and partial implementation of the ASAIID system.

6. Comments, Critical Remarks, and Recommendations

The dissertation follows a clear and logical structure. All references are cited correctly. The abbreviations used throughout the dissertation are thoroughly explained. The figures and tables are accurately referenced and effectively complement the text.

I recommend that the Msc. Eng. Ivan Drankov continue publishing the results of his future research in journals indexed in **Scopus** and **Web of Science**.

7. Conclusion with a clear positive or negative evaluation of the PhD dissertation.

I give a **positive evaluation** of the dissertation entitled "**Analysis of Leaked Critical Data to Identify Potential Information Security Threats**", authored by MSc. Eng. Ivan Lyubomirov Drankov. I believe that it was prepared precisely, accurately and at a high professional level, and the author has achieved the goals and objectives set in the dissertation. The results and contributions of the dissertation are reflected in scientific publications that have been presented to the scientific community at international conferences, published and indexed. The scientific, applied research and applied contributions are significant and comply with the requirements for acquiring the scientific and educational degree "doctor" under the Law on the Development of Academic Staff in the Republic of Bulgaria, the Regulations for its application and the Regulations on the Terms and Conditions for Acquiring Scientific Degrees at the University of Mining and Geology "St. Ivan Rilski", Sofia.

I propose to the esteemed Scientific Jury to award the educational and scientific degree "doctor" to the Msc. Eng. Ivan Lyubomirov Drankov in Professional Field 5.13 General Engineering, PhD Program "Computer Technologies in Engineering".

03 August 2026

Chair of the Scientific Jury:

Assoc. Prof. Eng. Yordanka Anastasova, PhD