VIETNAM NATIONAL UNIVERSITY, HANOI
VNU UNIVERSITY OF ENGINEERING AND TECHNOLOGY

SOCIALIST REPUBLIC OF VIETNAM
Independence – Freedom – Happiness

INFORMATION ON DOCTORAL THESIS

1. Full name : LE VIET HA ..... 2. Sex: Male .....

3. Date of birth: 03-11-1985 ..... 4. Place of birth: Hanoi .....

5. Admission decision number: 1146/QĐ-ĐT dated 21/10/2019 .....

6. Changes in academic process: Extend the study period for a total of 02 years according to Decision No. 980A/QD dated 7/11/2022 and Decision No. 878A/QD dated 10/5/2024.

7. Official thesis title: *Enhancing webshell detection with deep learning-powered methods* (*Nghiên cứu một số phương pháp học sâu trong phát hiện đoạn mã độc*)

8. Major: Information Systems..... 9. Code: 9480104 .....

10. Supervisors:

- Associate Professor, Doctor Nguyen Ngoc Hoa, VNU University of Engineering and Technology.
- Doctor Phung Van On, Hanoi Financial And Banking University.

11. Summary of the **new findings** of the thesis:

+ Proposing an DL-powered source code scanning framework for webshell detection that combines signature-based techniques with deep learning algorithms. This framework provides guidance for developing specific models for accurate and efficient webshell detection in a variety of programming languages. For each type of interpreted and compiled programming language, we chose PHP and ASP.NET as the most popular languages of each type to build a webshell detection model based on ASAF. We conducted experiments and compared the above model with other studies to prove the effectiveness of ASAF.

+ Propose a deep learning model to thoroughly analyze the HTTP traffic to the web application server in order to quickly detect webshell queries. To solve the problem of data imbalance for training sets, we also propose an algorithm to improve the quality of training sets employed in the deep learning model. To demonstrate its effectiveness, we experimented with and compared the model to other studies on the same CSE-CIC-IDS2018 dataset. The deep learning model can work with the intrusion detection and

prevention system to add attack source IPs to a blacklist and proactively block URI queries to webshell on the web server before they happen.

12. Practical applicability, if any:

+ The framework based on web application source code analysis by combining pattern matching techniques with deep learning algorithms is a guide to develop specific webshell detection models suitable for different programming languages that are capable of being deployed in practice.

+ The deep learning model thoroughly analyze the HTTP traffic to the web can work with the NetIDPS to detect webshell and proactively add attack source IPs to a blacklist and proactively block URI queries to webshell on the web server.

13. Further research directions, if any:

+ Conducting a general survey of webshell datasets used in current research, thereby building a good data set that can be used as a standard for later research related to webshells.

+ Continue research and experimentation with the latest single DL/ML models and ensemble models to improve the ability to accurately detect advanced webshells.

+ Deeper research into the operating mechanism and characteristics of Webshell will allow the construction of toolkits to automate the Yara rule creation process.

+ Expanding research on webshells written in other languages, such as JSP, Ruby, Python, etc., towards building a general model that can effectively detect all types of webshells without depending on the programming language.

14. Thesis-related publications:

+ Nguyễn, Hoá & Le, Viet-Ha & Phung, Van-On & Du, Phuong-Hanh. (2019). Toward a Deep Learning Approach for Detecting PHP Webshell. SoICT 2019: Proceedings of the Tenth International Symposium on Information and Communication Technology. 514-521. 10.1145/3368926.3369733;

+ Le, Viet Ha and Phung, Van On and Nguyen, Ngoc Hoa (2020) Information Security Risk Management by a Holistic Approach: a Case Study for Vietnamese e-Government. IJCSNS International Journal of Computer Science and Network Security, 20 (6). pp. 72-82. ISSN 1738-7906;

+ Le, Viet Ha and Nguyen, Ngoc Tu and Nguyen, Ngoc Hoa and Le, Linh (2021) An Efficient Hybrid Webshell Detection Method for Webserver of Marine Transportation Systems. IEEE Transactions on Intelligent Transportation Systems. ISSN 1524-9050;

+ Le, Viet Ha and Du, Phuong Hanh and Nguyen, Ngoc Cuong and Nguyen, Ngoc Hoa and Hoang, Viet Long (2021) A Proactive Method of the Webshell Detection and Prevention based on Deep Traffic Analysis. International Journal of Web and Grid Services. ISSN 1741-1114 (In Press);

+ Ha V. Le, Hoang V. Vo, Tu N. Nguyen, Hoa N. Nguyen, and Hung T. Du (2022) Towards a Webshell Detection Approach Using Rule-Based and Deep HTTP Traffic Analysis. Computational Collective Intelligence: 14th International Conference, ICCCI 2022. vol 13501 pp. 571–584.

+ Ổn, P. V., Hà, L. V., & Hóa, N. N. (2022). Giải pháp đánh giá và quản lý rủi ro an toàn thông tin trong Chính phủ điện tử. Tạp Chí Khoa học - Công nghệ Trong lĩnh vực An toàn thông Tin, 1(13), 35-48.

+ Sáng chế "Phương pháp phát hiện đoạn mã độc trong mã nguồn ứng dụng web sử dụng ngôn ngữ ASP.NET", được Cục Sở hữu trí tuệ cấp mã số 1-0036538-000 ngày 27/06/2023

Date: …….	Date: ……..
Signature: ……	Signature: ……
Full name: Nguyen Ngoc Hoa	Full name: Le Viet Ha