ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE

REVIEWER'S OPINION OF FINAL THESIS

I. IDENTIFICATION DATA

Thesis name:	Decentralized Federated Learning for Network Security
Author's name:	Pavel Janata
Type of thesis :	master
Faculty/Institute:	Faculty of Electrical Engineering (FEE)
Department:	Computer Science Department
Thesis reviewer:	Carlos A. Catania
Reviewer's department:	Department of Computer Science, National University of Cuyo, Mendoza, Argentina

II. EVALUATION OF INDIVIDUAL CRITERIA

Assignment **challenging**

Evaluation of thesis difficulty of assignment.

Federated learning is a relatively new research field and its application to network security is still in its early stages of development. Developing Federated Learning requires dealing several issues regarding data distribution, communication efficiency, model converge and heterogeneity of the involved devices, among others. In the proposed solution has tackle several of these issues by integrating different computer science topics such as distributed systems, machine learning and software engineering.

Satisfaction of assignment **fulfilled**

Assess that handed thesis meets assignment. Present points of assignment that fell short or were extended. Try to assess importance, impact or cause of each shortcoming.

The work in the thesis has met the assignment. The central aspects of a federated learning system applied to network security have been developed. Machine Learning Algorithms (Supervised and Unsupervised) are tested under different network conditions. The model convergence, its performance and the communication efficiency of the solution were analyzed.

Method of conception **correct**

Assess that student has chosen correct approach or solution methods.

The student has followed the correct methodology. First, he has posed the challenges involved in development of federated learning systems, Then he has analyzed the state of the art and propose a solution applied to network security. The proposed solution was evaluated under the most possible realistic conditions. In particular, the statistical heterogeneity of the data and the imbalance in both the participants' computational resources and data volume. Finally, a set of models were carefully designed for evaluating its hypothesis following machine learning standard procedure.

Technical level **B - very good.**

Assess level of thesis specialty, use of knowledge gained by study and by expert literature, use of sources and data gained by experience.

The student has demonstrated valuable skills for dealing with a new and difficult problem and

provided a valid solution using a diverse set of tools. He has shown expertise in several areas such as software development, machine learning, and the field of distributed systems.

Formal and language level, scope of thesis **A – excellent.**

Assess correctness of usage of formal notation. Assess typographical and language arrangement of thesis.

In general, the thesis is informative and clear. The student expressed in a correct language the different aspects involved in the process of building a federated learning system using formal notation when required.

Selection of sources, citation correctness **A – excellent.**

Present your opinion to student's activity when obtaining and using study materials for thesis creation. Characterize selection of sources. Assess that student used all relevant sources. Verify that all used elements are correctly distinguished from own results and thoughts. Assess that citation ethics has not been breached and that all bibliographic citations are complete and in accordance with citation convention and standards.

The student has always made reference to third party articles and software applications used for meeting the thesis assignment. In general, all references used in the work followed the proper quality standards. A minor issues to improve is that in the first mention to the flower framework, where a proper citation is missing (page 29)

Additional commentary and evaluation

Present your opinion to achieved primary goals of thesis, e.g. level of theoretical results, level and functionality of technical or software conception, publication performance, experimental dexterity etc.

Please insert your commentary (voluntary evaluation).

III. OVERALL EVALUATION, QUESTIONS FOR DEFENSE, CLASSIFICATION SUGGESTION

Summarize thesis aspects that swayed your final evaluation. Please present apt questions which student should answer during defense.

In this thesis, the student has proposed a new method for distributed federated learning applied to the network security problem. The main contribution of the proposed solution consisted in the inclusion of the so-called `vaccines`, which are small portions of malware sent to the different clients and incorporated into their local datasets for federated training. The major benefits of vaccines is observed during the training of supervised algorithms. In addition, the student has analyzed the performance of the proposed solution under different network and data conditions.

The student was involved in all the different stages of developing a distributed federated system for network security. The student has placed particular emphasis on the construction of a dataset which has been provided to the community for free. Three different neural network were analyzed and carefully evaluated under different network conditions following the standard machine learning methodology.

Apt questions:

1) The proposed solution focused on federated learning using neural network algorithms. Is it possible to use other algorithms than neural networks in the current implementation of the system? If not, how difficult will be to adapt current solution to some other algorithm such as Random Forest.

2) Regarding the use of the malware vaccine. When is the vaccine malware sent to the clients? Every time a new model is built and the client does not have malware labels? Or vaccine malware is sent to all the clients no matter they already have malware labels or not?

3) Do you think the performance of the dual-head model was caused by the relative small size of the dataset? In other words, Do you think with more data the dual-head model would outperform the Classifier-only model?

4) Could you provide information about the time required for a Federate Learning model to converge? Just to get an ideal about the differences between centralized learning.

I evaluate handed thesis with classification grade A - excellent.

Date: **3.2.2023**

Signature: