Автоматизация в промышленности 2022. №04
www.avtprom.ru

DOI: 10.25728/avtprom.2022.04.10

М.В. Буйневич (СПб УГПС МЧС), К.Е. Израилов (СПбГУТ, ФИЦ РАН),
В.В. Покусов (КАИБ), Н.Е. Романов (СПбГУТ)

Способ вариативной классификации уязвимостей в программном коде.
Часть 2. Автоматизация на базе машинного обучения

Решается задача автоматической вариативной классификации уязвимостей программного кода с применением машинного обучения. В первой части статьи рассмотрены начальные четыре шага такого способа. Во второй части описаны оставшиеся четыре шага, а именно: описание уязвимостей, их текстовые признаки, идентификация и классификация. Для реализации последнего наиболее сложного шага сделан обзор релевантных научных работ. Выделены подходы, лежащие в основе способа автоматической классификации, отмечены их достоинства и недостатки. Предложен способ, основанный на машинном обучении в части классификации определенных характеристик уязвимостей по их текстовому описанию. Разработан прототип, реализующий техники Word2Vec и TF-IDF, а также пять классификаторов. Проведен эксперимент с разработанным прототипом на базе описаний уязвимостей ФСТЭК России. Определены наиболее результативные комбинации техник и классификаторов.

Ключевые слова: программное обеспечение, уязвимости программного кода, вариативная классификация, машинное обучение, прототип, эксперимент.

Буйневич Михаил Викторович *– д-р техн. наук, проф., проф. Кафедры прикладной математики и информационных технологий Санкт-Петербургского университета государственной противопожарной службы МЧС России,*
Израилов Константин Евгеньевич *– канд. техн. наук, доцент кафедры защищенных систем связи Санкт-Петербургского государственного университета телекоммуникаций им. проф. М.А. Бонч-Бруевича, старший научный сотрудник лаборатории проблем компьютерной безопасности Санкт-Петербургского Федерального исследовательского центра РАН,*
Покусов Виктор Владимирович *– председатель Казахстанской ассоциации информационной безопасности, Алматы, Казахстан,*
Романов Никита Евгеньевич *– студент кафедры защищенных систем связи Санкт-Петербургского государственного университета телекоммуникаций им. проф. М.А. Бонч-Бруевича.*

Список литературы

1. Буйневич М.В., Израилов К.Е., Матвеев В.В., Покусов В.В. Способ вариативной классификации уязвимостей в программном коде. Часть 1. Стратификация и категориальное деление // Автоматизация в промышленности. 2021. № 11. С. 42-49.
2. Аверченков В.И., Жога В.Л. Автоматизация процедуры конструкторско-технологической классификации деталей с использованием самоорганизующейся нейронной сети // Известия Волгоградского государственного технического университета. 2006. № 4 (21). С. 94-97.
3. Мешалкин В.П., Дли М.И., Пучков А.Ю., Лобанева Е.И. Предварительная оценка прагматической ценности информации в задаче классификации на основе глубоких нейронных сетей // Прикладная информатика. 2021. Т. 16. № 3 (93). С. 9-20.
4. Алимов Н.А., Ерофеева В.А., Шалымов Д.С. Анализ возможностей методов классификации для автоматизации работы дефибриллятора // Стохастическая оптимизация в информатике. 2017. Т. 13. № 1. С. 3-30.
5. Шипилина Л.Б. Использование техники машинного обучения для автоматизации классификации знаний // Управление развитием крупномасштабных систем (MLSD'2013): тр. VII международной конференции. 2013. С. 346-350.

6. Кретов В.С., Аблов И.В., Котов Н.М. Подход к применению нейронных сетей к задачам классификации объектов // Тр. XXVII международной конференции «Проблемы управления безопасностью сложных систем» . 2019. С. 93-100.
7. Надеждин Е.Н., Шершакова Т.Л. Задача классификации уязвимостей программного обеспечения в корпоративной информационной сети // Шуйская сессия студентов, аспирантов, педагогов, молодых ученых: материалы X Международной научной конференции. 2017. С. 207.
8. Израилов К.Е., Виткова Л.А., Чечулин А.А. Компонент классификации уязвимостей интерфейсов взаимодействия с транспортной инфраструктурой умного города: свидетельство о регистрации программы для ЭВМ № 2020661231 от 18.09.2020.
9. Вульфин А.М., Никонов А.В., Габбасова Д.Н. и др. Программа анализа уязвимостей программного обеспечения на основе технологий интеллектуального анализа и обработки естественного языка: свидетельство о регистрации программы для ЭВМ № 2021615080 от 02.04.2021.
10. Дойникова Е.В., Федорченко А.В., Котенко И.В. Выявление слабых мест информационных систем для автоматического выбора защитных мер // Проблемы информационной безопасности. Компьютерные системы. 2018. № 3. С. 89-99.
11. Kotenko I., Izrailov K., Buinevich M. Static Analysis of Information Systems for IoT Cyber Security: A Survey of Machine Learning Approaches // Sensors. 2022. Vol. 22. Iss. 4. PP. 1335. DOI: 10.3390/s22041335
12. Kotenko I., Izrailov K., Buinevich M. Analytical Modeling for Identification of the Machine Code Architecture of Cyberphysical Devices in Smart Homes // Sensors. 2022. Vol. 22. Iss. 3. PP. 1017.
13. Yosifova V., Tasheva A., Trifonov R. Predicting Vulnerability Type in Common Vulnerabilities and Exposures (CVE) Database with Machine Learning Classifiers. The proceedings of 12th National Conference with International Participation (ELECTRONICA). 2021. PP. 1-6.
14. Chen H., Zhang D., Chen J., Lin W., Shi D., Zhao Z. An Automatic Vulnerability Classification System for IoT Softwares. The proceedings of IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom). 2020. PP. 1525-1529.
15. Савченко Т.Ю. Обработка естественного языка для использования в машинном обучении: частотная векторизация, TF-IDF, Word2Vec // Аллея науки. 2018. Т. 4. № 6 (22). С. 1000-1002.
16. Буйневич М.В., Израилов К.Е., Щербаков О.В. Модель машинного кода, специализированная для поиска уязвимостей // Вестник Воронежского института ГПС МЧС России. 2014. № 2 (11). С. 46-51.

Buynevich M.V., Izrailov K.E., Pokusov V.V., Romanov N.E. A method for variability classification of vulnerabilities in program code. Part 2. Machine learning-based automation

The problem of variative classification of vulnerabilities in program code is solved with the help of machine learning. The first part of the article discussed the first four steps of the procedure. The second part focuses on the remaining for steps, namely: the description of vulnerabilities, their textual evidences, identification and classification. For implementing the most difficult last step, the relevant scholarly articles were overviewed. The approaches underlying the automated classification technique were pointed out, their merits and drawbacks were emphasized. A procedure based on machine learning with respect to the classification of certain vulnerability characteristics based on their textual description is developed. A prototype implementing Word2Vec and TF-IDF techniques as well as five classifiers is developed. The prototype was tested against vulnerability descriptions from the Federal Service for Technical and Export Control (FSTEC of Russia). The most effective combinations of techniques and classifiers were detected.

Keywords: software, program code vulnerabilities, variability classification, machine learning, prototype, experiment.