УДК 004.056, 004.934, 004.912

МОДЕЛЬ СБОРА И ОБРАБОТКИ НЕСТРУКТУРИРОВАННОЙ ИНФОРМАЦИИ ОБ УЯЗВИМОСТЯХ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ

Нкодиа Д.-К. (Университет ИТМО) **Менщиков А. А.** (Университет ИТМО)
Научный руководитель – к.т.н., доцент Менщиков А.А.
(Университет ИТМО)

В работе рассмотрены методы сбора и обработки неструктурированной информации об уязвимостях программного обеспечения. Предложен формат представления данных об уязвимых продуктах в машинно-читаемом виде.

В различных открытых источниках постоянно появляется неструктурированная информация о новых обнаруженных уязвимостях программного обеспечения (далее – ПО) в виде текстов на естественном языке. Возможность своевременного получения данных об уязвимостях в машинно-читаемом виде ускорит передачу и обработку критически важной информации и сократит возможные расходы компании. На данный момент большинство решений по анализу информации об уязвимостях опираются на наличие специализированных идентификаторов уязвимостей, на назначение которых уходит дополнительное время, и не определяют уязвимое программное обеспечение.

На первом этапе работы в качестве источников информации предлагается рассматривать различные форумы и отчеты, связанные с информационной безопасностью. В дальнейшем для получения первичных данных планируется анализировать соответствующую информацию в DarkNet и DeepNet.

Для того, чтобы определить, актуальна ли анализируемая информация и с каким типом уязвимости она связана, применяются различные подходы: методы искусственного интеллекта, в том числе NLP (Natural Language Processing — обработка естественного языка), и статистический анализ. Особенность предложенной модели обработки информации заключается в том, что наличие и отсутствие идентификатора уязвимости в тексте не будет влиять на результаты классификации данных. Такая модель должна значительно понизить значение ошибки второго рода.

Результатом анализа информации источника являются структурированные данные в формате XML (eXtensible Markup Language — расширяемый язык разметки), содержащие название и версию уязвимого ПО, описание уязвимости и ссылку на источник. Предложенный формат основывается на проанализированных стандартах OVAL (Open Vulnerability and Assessment Language - открытый язык описания и оценки уязвимостей), SBOM (software bill of materials – стандарт описания компонентов в программном обеспечении), CPE (Common Platform Enumeration - структурированная схема именования для систем информационных технологий, программного обеспечения и пакетов) и PURL (package URL – формат идентификации и поиска программных пакетов).

Предложенное решение по сбору и обработке неструктурированных данных об уязвимостях ПО может быть использовано в качестве составной части системы анализа безопасности используемых продуктов и библиотек для повышения защищенности компаний с возможностью оповещения о найденных угрозах нулевого дня.

Нкодиа Д.-К. (автор) _____

Менщиков А.А. (научный руководитель) _____