УДК 004.85

ДЕТЕКТИРОВАНИЕ ФИШИНГОВЫХ САЙТОВ МЕТОДАМИ МАШИННОГО ОБУЧЕНИЯ

Макаренко З. В. (Университет ИТМО)**, Деева И. Ю.** (Университет ИТМО)
Научный руководитель – к.т.н., доцент Менщиков А. А. (Университет ИТМО)

Аннотация. В настоящем исследовании был проведен анализ применимости методов машинного обучения для детектирования фишинговых сайтов и предложен собственный метод определения фишинговых сайтов с использованием машинного обучения.

Введение. Развитие интернет-технологий несет за собой и увеличение случаев интернет-мошенничества. Одной из актуальных угроз современного интернета является фишинг в различных его проявлениях. Год от года количество фишинговых атак лишь растет, а мошенники подделывают все сайты, включая государственные. Организации, ставшие жертвами фишинговых атак, несут финансовые, материальные и репутационные убытки. Рядовые пользователи, столкнувшись с мошенниками, могут передать им свои персональные данные, данные банковских карт, либо перевести денежные средства.

Однако, своевременное обнаружение фишинговых сайтов позволяет существенно снизить количество жертв. Согласно последним исследованиям в области обнаружения фишинговых атак, методы детектирования фишинговых сайтов с использованием машинного обучения показывают большую эффективность по сравнению с другими методами и при позволяют находить фишинговые страницы в день их создания. Развитие и усовершенствование существующих методов позволит снизить ущерб от фишинговых атак.

Основная часть. С использованием серверного программного обеспечения производится поиск потенциальных сайтов-клонов. Происходит парсинг сайта-оригинала и полученных в результате поиска доменов и страниц. Выделяются характеристики для решения задачи классификации сайта. Разрабатывается и обучается модель классификации для распознавания фишинговых сайтов. Модель классификации определяет, принадлежит ли сайт к категории «фишинг» или нет, и выводит результаты.

Разработка модели классификации произведена на базе фреймворка для автоматизированного моделирования и машинного обучения FEDOT, так как данный фреймворк позволяет комбинировать несколько моделей машинного обучения разной сложности.

Представленная в исследовании модель классификации тестировалась на датасетах фишинговых адресов из открытых источников.

Выводы. Чем больше фишинговых сайтов будут детектироваться на стадии существования «нулевого часа», тем меньше пользователей и организаций будут терять свои денежные средства, персональные и конфиденциальные данные. Однако, как бы ни были эффективные технические методы противодействия злоумышленникам, без внесения правок в законодательную базу преследование и наказание мошенников часто бывает затруднительным.

Макаренко З. В. (автор) _____

Деева И. Ю. (автор) _____

Менщиков А. А. (научный руководитель) _____