УДК 004.056

ИССЛЕДОВАНИЕ СПОСОБОВ АВТОМАТИЗАЦИИ ТАРГЕТИРОВАННЫХ ФИШИНГ-АТАК И МЕТОДОВ ИХ ДЕТЕКТИРОВАНИЯ

Коркунова А. А. (федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»)

Научный руководитель – к.т.н., доцент ФБИТ Воробьева А. А.

(федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»)

В представленной работе рассмотрены и проанализированы основные способы автоматизации таргетированных фишинг-атак, а также методы их детектирования. Сформированы выводы о наиболее опасных подходах к автоматизации атак и эффективных методах их детектирования.

Введение.

В настоящее время можно наблюдать очень активное развитие возможностей ИИ и его интеграцию в различные сферы жизни. Безусловно, пользуются этим и злоумышленники – например, применяя ИИ для автоматизации атак, в том числе таргетированных фишинг-атак, (полностью или отдельных их этапов) и повышения таким образом их эффективности.

Основная часть.

Исследование заключается в выявлении способов автоматизации таргетированных фишинг-атак, в ряде случаев не детектируемых известными на данный момент методами. В работе проанализированы способы автоматизации различных этапов атаки: подготовка (сбор информации о людях и системе), проникновение (олицетворение доверенного лица, обход ограничений) и достижение целей (хищение данных, сокрытие следов). Выделено наиболее перспективное направление: применение технологии генерации естественного языка (NLG) – в результате многопрофильного обучения снижается детектируемость созданных ИИ текстов. Также рассмотрены 4 метода детектирования фишинг-атак (в вариациях), основанных на ИИ: глубокое обучение (часто используется, но неэффективна в ряде случаев), машинное обучение (используется чаще всего, зафиксирована высокая эффективность – до 99% TP), основанный на сценариях (применим только для конкретной среды) и гибридное обучение (может давать более высокую точность, чем RF).

Выводы.

Проведен анализ возможных способов автоматизации таргетированных фишинг-атак и анализ методов их детектирования. По результатам исследования выделен способ, не детектируемый в равной степени для текстов, созданных людьми и ИИ, а следовательно, представляющий более высокий уровень угрозы.

Список использованных источников:

1. Basit, A., Zafar, M., Liu, X. et al. A comprehensive survey of AI-enabled phishing attacks detection techniques // Электронный ресурс. – 2020.
2. Shih-Wei Guo, Tzu-Chi Chen, Hui-Juan Wang & Yao-Chung Fan, Fang-Yie Leu Generating Personalized Phishing Emails for Social Engineering Training Based on Neural Language Models // Электронный ресурс. – 2022.
3. Ahmad Najee-Ullah, Luis Landeros, Yaroslav Balytskyi & Sang-Yoon Chang Towards Detection of AI-Generated Texts and Misinformation // Электронный ресурс. – 2022.

Коркунова А. А. (автор)

Воробьева А. А. (научный руководитель)