ОБНАРУЖЕНИЕ АКТИВИРОВАННОГО АППАРАТНОГО ТРОЯНА В ЦИФРОВЫХ УСТРОЙСТВАХ КРИПТОГРАФИИ

Воронов А.Ю.

Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь

Стемпицкий В.Р. – канд. тех. наук.

В работе изложены результаты научных исследований эффективности применения классификатора на базе нейронной сети для обнаружение аномального поведения траффика цифрового устройства криптографии, реализующего алгоритм AES 256. Показано сравнение результатов при попытке расширения классификатора для обнаружения более сложных случаев подмены ключа шифрования. Подчеркивается перспективность предлагаемой методики и даны предполагаемые способы решения проблем рассматриваемого метода.

В настоящее время рост количества абонентов в цифровой сети и объемов передаваемой ин формации между ними является одним из обязательных условий развития коммерческих и государственных экосистем. Внедрение концепции интернета вещей во все цифровые устройства требует создания новых, более функциональных устройств, способных принимать, обрабатывать и передавать большие объемы данных. Подобно программному обеспечению, аппаратное обеспечение также уязвимо к внедрению вредоносных схемотехнических решений, называемых аппаратными троянами или аппаратными закладками, которые могут представлять опасность в области конфиденциальности передаваемой информации и функционирования всей цифровой экосистемы в целом. Цель проводимого исследования – определение эффективности способа обнаружения внедренных троянов в интегральных микросхемах, реализующих алгоритмы криптографического шифрования, с помощью методов машинного обучения.

Основной проблемой исследуемого подхода является теоретическая невозможность предсказания результатов криптографического шифрования алгоритма AES 256. Данная проблема подтвердилась при попытке использования архитектуры автоэнкодера при обучении модели на исходном датасете. Так при тренировке модели на протяжении 50 эпох на 4 миллионах данных, модель на рассматриваемой архитектуре не смогла отличить нормальное поведение AES 256 от аномального при активированной аппаратной закладке [1].

Для решения проблемы предсказания результатов принято решение использовать подход классификации изображений. Для этого при обучении использовалась модель полносвязной нейронной сети. Модель обучалась на исходном наборе данных, состоящих из 4 миллионов данных. После обучения на исходном датасете классификатор определял нормальную работу AES 256 в 98 % случаях, отправку незашифрованной посылки в 99 %, отправку ключа в 87 % и подмену ключа на одинаковые биты в 96 % [1]. Данный результат показал возможность применения данного подхода в сложных устройствах криптографии, но не дал уверенности в его достаточности для определения утечки информации при помощи более сложных комбинаций подмененного ключа.

Было проведено дообучения модели с расширением классификатора, которое включало отдельные случае подмены ключа на одинаковые числовые символы. Проверка полученной модели показало возросшее количество ошибок классификации при нормальной работе устройства (63 % точности против 98 %), хотя модель со средней точностью 85 % могла распознавать случаи подмены ключа на одинаковые числовые символы [2]. Основная проблема заключалась в том, что модель равномерно распределяла совпадения между новыми классами, которые отвечали за подмену ключа на одинаковые цифры.

Данное исследование показывает возможность применение описанной методики для обнаружения активированной аппаратной закладки в AES 256. Для повышения точности обнаружения функционального изменения дальнейшие работы будут направлены на изменение применяемой архитектуры нейронной сети либо переход на кластер различных классификаторов с использованием алгоритма или нейронной сети для анализа плотности ошибок каждого элемента кластера.

Список использованных источников.

1. Воронов, А. Ю. Обнаружение аппаратных троянов в устройствах криптографии с использованием машинного обучения / А. Ю. Воронов, В. Р. Стемпицкий // Доклады БГУИР. 2025. Т. 23, № 6. С. 71–79. http://dx.doi.org/10.35596/1729-7648-2025-23-6-71-79

2. Voronov A. Yu. Activated hardware trojan detection in AES-256 communications equipment. European Research: Theory, Practice and Innovation. P. 186 - 192 Budapest, Hungary. March 25, 2026. DOI 10.34660/INF.2026.91.89.299.