УДК 004.85

# *АНАЛИЗ СЕТЕВОГО ПОТОКА ИНФРАСТРУКТУРЫ ПОСРЕДСТВОМ АЛГОРИТМОВ FUZZY LOGIC*

***Чумаков В.Е.***

*Магистрант 2 курса напр. «Интеллектуальные системы и технологии»,*

*ИСОиП (филиал) ДГТУ г. Шахты,*

*Россия, Шахты*

**Аннотация**

В данной статье рассматриваются алгоритмы использования нечеткой логики; анализ и преобразование больших данных посредством машинного обучения для нахождения наиболее важных и полносвязнных атрибутов. Приводится пример создания архитектуры системы. Иллюстрируется практическая работа с существующим датасетом. Описываются преимущества использования нечеткой логики по сравнению с другими системами обнаружения аномалий сетевого трафика.

**Ключевые слова:** нечеткая логика, Decision Tree Classifier, аномалии сетевого трафика, модель Mamdani.

# *ANALYSIS OF INFRASTRUCTURE NETWORK FLOW USING FUZZY LOGIC ALGORITHMS*

***Chumakov V. E.***

*2nd year master's student e.g. "Intelligent systems and technologies",*

*Isoip (branch) of DSTU Shakhty*

*Russia, Shakhty*

**Annotation**

This article discusses the use of algorithms of fuzzy logic. Analyze and transform big data through machine learning to find the most important and fully connected attributes. An example of creating a system architecture is given. Practical work with an existing dataset is illustrated. The advantages of using fuzzy logic in comparison with other systems for detecting network traffic anomalies are described.

**Keywords**: fuzzy logic, Decision Tree Classifier, network traffic anomalies, Mamdani model

**Введение.** В настоящее время все большую популярность в среде анализа и работы с наборами данных завоевывают алгоритмы машинного обучения. Машинное обучение позволяет с высокой производительностью работать над распознаванием текста, классификации изображений, усовершенствовать различные интернет поисковики и переводчики. Глубокое обучение является одним из методов машинного обучения, который превосходит остальные методы, за счет большей производительности и лучшей точности при работе с большими объемами данных. Однако приведенные алгоритмы работают с точными данными, что является недостатком в предсказании сетевых атак. Алгоритмы нечеткой логики позволяют избавиться от такого недостатка. В данной статье рассматривается использование нечеткой логики для предсказания атак, а также рассмотрен пример возможной архитектуры системы.

**Архитектура системы.** Построение архитектуры системы является неотъемлемой частью в разработке какого-либо программного продукта. В данной статье проиллюстрирована архитектура потенциального программного обеспечения, представленная на рисунке 1.


Рис.1 – Архитектура системы

Для большего понимания архитектуры рассмотрим каждый модуль.

**Модуль хранилище.** В нашем эксперименте используется датасет, состоящий из 80 столбцов и 225000 строк, обозначающие различные собранные данные с сетевого оборудования за неделю работы сетевой инфраструктуры [1]. Для ускорения работы системы и уменьшения возможного шума необходимо анализировать только TCP – пакеты [2]. Следовательно, на вход модуля «хранилище» подается необработанный датасет, из которого выбирается объем данных, представляющих только TCP – сессии.

**Модуль маркировки атрибутов.** В этом модуле задействуются алгоритмы машинного обучения. Для увеличения скорости обучения модели и дальнейшего обнаружения, модель должна вычленить из всего набора данных наиболее важные атрибуты, по которым она будет определять возможную атаку. Тем самым из 80 атрибутов, находящихся в датасете, обученная модель будет использовать только два или три атрибута. В нашем эксперименте использовался алгоритм Decision Tree Classifier [3]. Применив представленный алгоритм, получили четыре атрибута, которые модель считает наиболее важными для выявления атаки. Эти атрибуты представлены на рисунке 2.

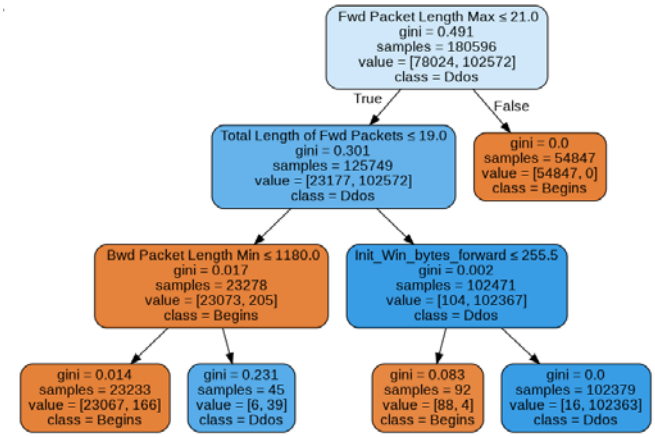


Рис.2 – Обнаруженные атрибуты

**Модуль Fuzzy_Logic.** Рассматриваемый модуль использует алгоритм нечеткой логики для повышения точности обнаружения аномалий в сети, путем работы с нечеткими данными. Представим полученные атрибуты в виде таблицы.

Таблица 1 – Обозначение обнаруженных атрибутов

| **Номер атрибута** | **Название атрибута** | **Обозначение** |
|---|---|---|
| 1 | Fwd Packet Length Max | Максимальное значение длины пакетов |
| 2 | Total Length of Fwd Packets | Общее количество пересылаемых пакетов |
| 3 | Bwd Packet Length Min | Минимальное количество обработанных пакетов |
| 4 | Init_Win_bytes_forward | Общее количество отправленных байтов |

Применив алгоритм Mamdani, построим модуль, который содержит четыре входа, обозначающие атрибуты, которые были выявлены в модуле маркировки, а также один выход, в котором используется три функции принадлежности

отвечающих за классификацию трафика, как атака, предупреждение и нормальное состояние(рис.3) [4].

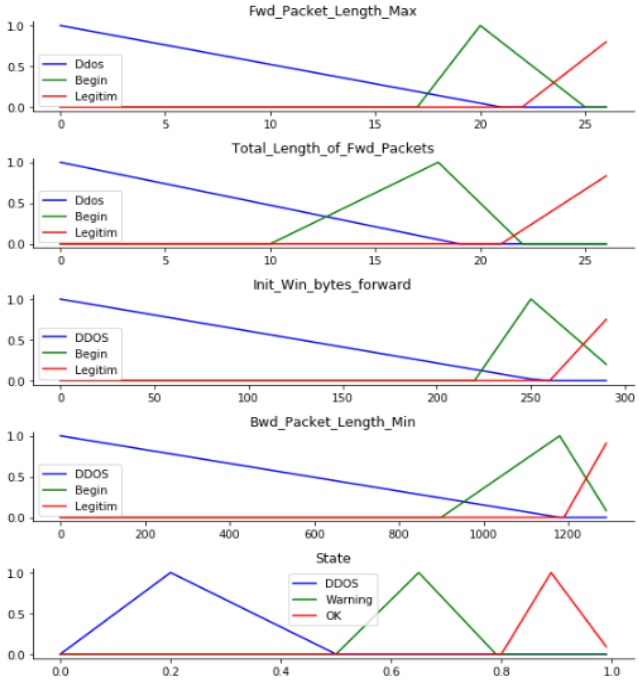


Рис.3 – Модуль Fuzzy_Logic

Для того, чтобы убедиться в работоспособности созданных нечетких правил и функций принадлежностей, проиллюстрируем графическое представление правил на рисунке 4.

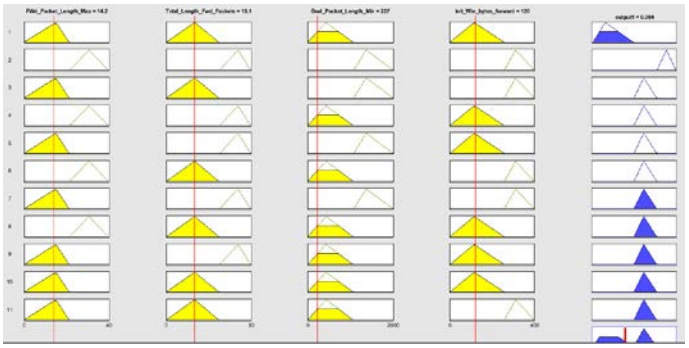

Рис.4 – Графическое представление правил

Для тестирования созданной модели зададим входные параметры наших атрибутов, равных:

```
Fwd_Packet_Length_Max =25
Total_Length_of_Fwd_Packets = 25
```

```
Init_Win_bytes_forward = 250
Bwd_Packet_Length_Min = 1300
```

Тем самым модуль нечеткой логики выделяет вероятность попадания в ту или иную область функции принадлежности на выходе, в соответствии с заданными параметрами:

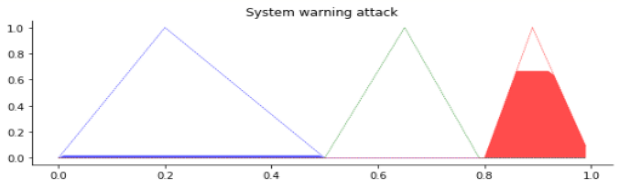


Рис.5 – Вероятность функции принадлежности на выходе

Применив созданные правила получим точный срез функции принадлежности на выходе, иллюстрирующим маркер, который попадает в класс легитимного сетевого трафика:

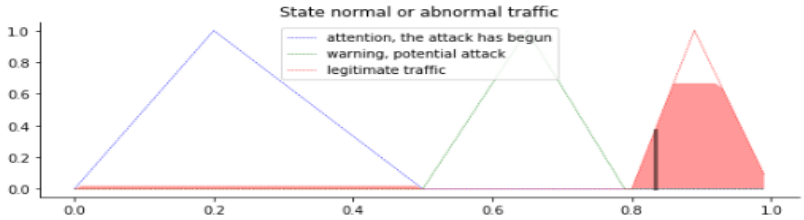


Рис.6 – Классификация сетевого трафика

**Вывод.** Статья иллюстрирует основные возможности применения нечеткой логики для обнаружения аномального сетевого трафика в сетевой инфраструктуре предприятия. Приводится практический пример использования алгоритмов машинного обучения и алгоритмов работы нечеткой логики на существующем датасете. Представленный эксперимент показывает преимущество нечеткой логики по сравнению с другими методами определения аномального трафика, за счет работы с нечеткими диапазонами входных параметров, позволяющих не только повышать уровень безопасности сетевой инфраструктуры, но и обнаруживать атаки нулевого дня за счет уменьшения шума во входных данных и использования нескольких функций принадлежности.

**Библиографический список:**

1. SAS. Evolution of machine learning. URL: https://www.sas.com/en_us/insights/analytics/machine-learning.html (дата обращения 29.09.2019)

2. Ильин О.А. Реконструкция TCP – сессий. URL: https://www.tamos.ru/htmlhelp/commwifi/reconstruct.htm (дата обращения 11.10.2019)

3. Geeks For Geeks. Decision tree regression using sklearn. URL: https://www.geeksforgeeks.org/python-decision-tree-regression-using-sklearn/ (дата обращения 18.11.2019)

4. Штовба С.Д. Нечеткая системная настройка вывода. URL: https://docs.exponenta.ru/fuzzy/fuzzy-inference-system-tuning.html (дата обращения 21.11-.2019)

*Оригинальность 98%*