УДК 004.056.5

Р. М. Алыгулиев, Я. Н. Имамвердиев, Ф. Д. Абдуллаева
e-mail: r.aliguliyev@gmail.com, yadigar@iit.science.az, a_farqana@mail.ru

Институт информационных технологий Национальной академии наук Азербайджана, Баку, Азербайджан

ОБНАРУЖЕНИЕ АНОМАЛИЙ В ОБЛАЧНЫХ BIG DATA ДАННЫХ

В статье предлагается метод обнаружения аномалий на уровне гипервизора инфраструктуры облачных вычислений. Для повышения точности обнаружения аномалий в предложенном подходе используется гибридный алгоритм, полученный комбинированием алгоритма плотностной кластеризации и алгоритмов классификации дерева решений (J48) и проективной адаптивной теории резонанса (PART-Projective Adaptive Resonance Theory).

В облачной среде могут быть реализованы атаки различного типа. С помощью одного метода обнаружить эти атаки практически невозможно. В методах обнаружения аномалий быстрота и точность считаются важными требованиями. С этой целью в данной статье предлагается метод, в котором используются методы, основанные на кластеризации и классификации. В области обнаружения аномалий имеется довольно большое количество подходов, основанных на методах классификации и кластеризации [1].

Результаты экспериментов, проведенных в процессе исследований, показывают, что комбинированное использование метода плотностной кластеризации, дерева решений (J48) и алгоритма PART позволяет получить лучшие результаты (далее в тексте предложенный метод называется VMM−J48+PART (VMM − Virtual Machine Monitor)).

Экспериментальная оценка предложенного подхода была проведена в программном пакете WEKA на базе данных NSL-KDD.

Naive Bayes, Random Forest, Decision Tree, SVM являются наиболее эффективными алгоритмами, используемыми для обнаружения аномалий. Как утверждает исследование, данные алгоритмы показывают хорошие результаты только при обнаружении

нормальных поведений и DoS атак. Однако применение этих методов для обнаружения U2R, R2L атак не дает желаемые результаты. Как видно из таблицы, алгоритмы SVM, Naive Bayes, J48 практически не обнаруживают R2L и U2R атак. Но предлагаемый гибридный алгоритм VMM−J48+PART может обнаружить все виды атак с высокой точностью. Итак, перечисленные методы показали низкий процент обнаружения U2R и R2L атак, в то время как алгоритм VMM-J48+PART выявляет данные атаки с точностью 86% и 80%, соответственно.

Сравнение методов обнаружения аномалий (в процентах)

Тип атаки	SVM	Naive Bayes	Random Forest	J48	VMM−J48+PART
Normal	95.78	86.87	97.30	97.00	81.75
DoS	82.37	71.13	82.86	77.39	70.92
U2R	12.00	25.50	1.50	4.50	86.86
R2L	10.60	10.46	6.61	6.54	80.48
Prob	67.25	81.58	64.64	65.47	94.14

Для повышения точности обнаружения аномалий в алгоритме VMM-J48+PART в первую очередь производится нормализация, затем плотностная кластеризация данных. В полученных кластерах строится дерево решений с применением комбинированных алгоритмов классификации J48 и PART. Предлагаемый метод обнаружения аномалий работает в двух фазах (рис. 1).

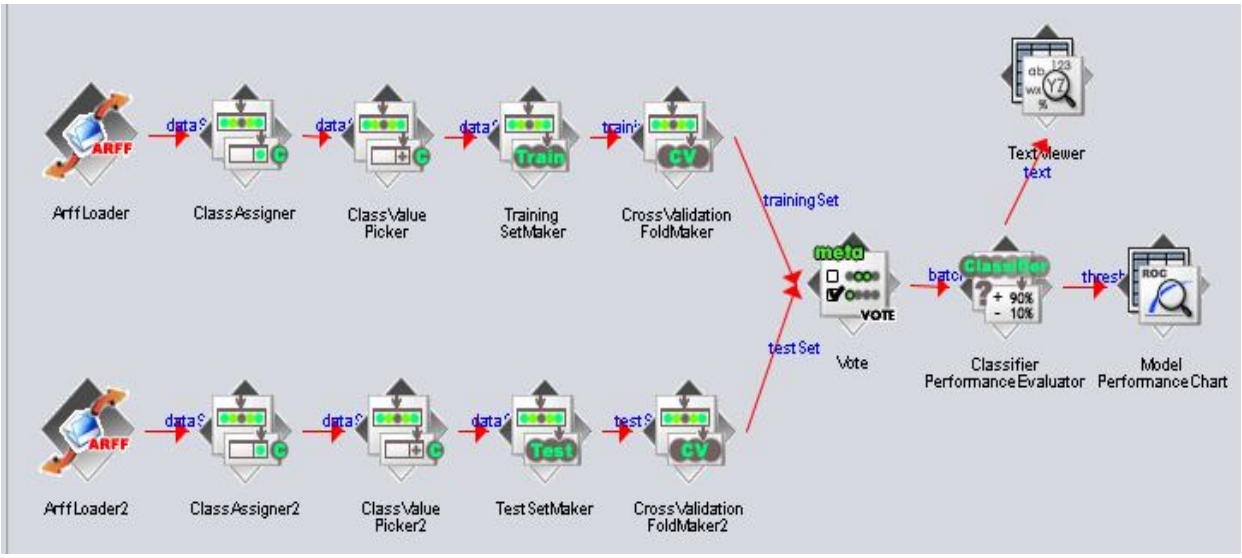

Рис. 1. Рабочий процесс метода VMM−J48+PART на среде Weka

На первом этапе с целью построения модели нормального профиля проводится обучение (Training), а на втором этапе проводится тестирование алгоритма для обнаружения аномалий.

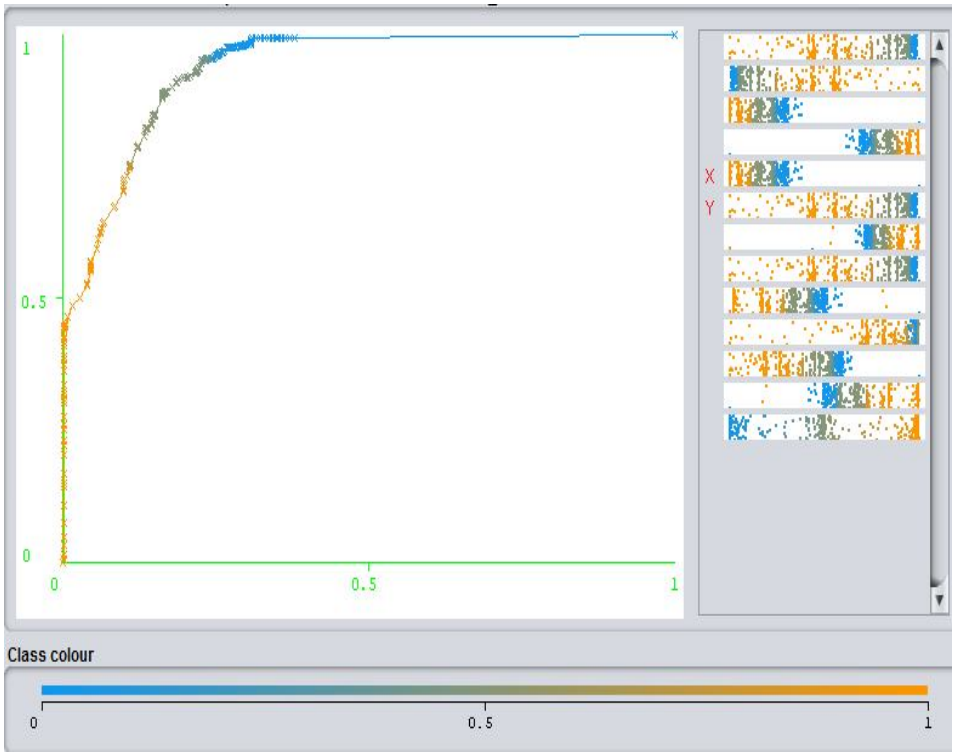

Рис. 2. ROC кривая

ROC (receiver operating characteristic) кривая, представленная на рис. 2, визуально отображает высокую точность обнаружения аномалий.

Данная работа выполнена при финансовой поддержке Фонда Развития Науки при Президенте Азербайджанской Республики – Грант № EİF-KETPL-2-2015-1(25)-56/05/1.

---

1. Pandeeswari N., Kumar G. Anomaly detection system in cloud environment using fuzzy clustering based ANN, Mobile networks and applications, 2016, vol. 21, No 3, p. 494–505.