Поиск отклонений от типичных сценариев использования веб-приложений

А. А. Хашаев[1], И. Ю. Терёхина[2], Д. Ю. Гамаюнов[3]

В работе исследуется задача поиска отклонений от известных сценариев типичного использования веб-приложений. Рассмотрена модификация существующей формальной постановки задачи, для которой предложена рекуррентная процедура проверки соответствия трассы шаблону. Разработан алгоритм обнаружения аномалий в трассе, эффективный относительно длины трассы и размера шаблона.

Ключевые слова: поиск аномалий, ациклический ориентированный граф, анализ последовательностей событий

1. Введение

Поиск отклонений от типичных сценариев использования веб-приложения является актуальной задачей информационной безопасности, которая может быть решена как с помощью детерминированных алгоритмов, так и с помощью алгоритмов машинного обучения [1, 2, 3, 4, 5]. В данной работе предлагается модификация формальной задачи, предложенной, например, в работе [6], для эффективного обнаружения аномалий в реальном времени.

Сценарии использования некоторого веб-приложения пользователем представляют собой многоступенчатые процессы, каждый из которых состоит из последовательности пользовательских действий, связанных между собой отношением частичного порядка по времени их выполнения. Некоторые действия являются пререквизитами к выполнению других действий, и в корректных сценариях использования веб-приложения ряд пользовательских действий может исполняться лишь в определенном порядке.

[1] *Хашаев Артур Акрамович* — директор по разработке, SolidWall.io, e-mail: arthur.khashaev@solidwall.io.

Khashaev Artur Akramovich — SolidWall.io.

[2] *Терёхина Ирина Юрьевна* — программист, SolidWall.io, e-mail: irina.terekhina@solidwall.io.

Teryokhina Irina Yurevna — SolidWall.io.

[3] *Гамаюнов Денис Юрьевич* — к.ф.-м.н., доц. кафедры информационной безопасности ф-та ВМК, Московский государственный университет им. М. В. Ломоносова, e-mail: gamajun@seclab.cs.msu.ru.

Gamayunov Denis Yurevich — Lomonosov Moscow State University.

Предполагается, что построена (обучена) некоторая модель, позволяющая получить множество сценариев типичного использования веб-приложения. Необходимо на этапе защиты веб-приложения в реальном времени проверять соответствие наблюдаемой последовательности пользовательских действий выявленным ранее наборам закономерностей.

2. Основные понятия и формулировка результата

Пусть известен набор возможных *действий* из некоторого конечного алфавита $\mathcal{A}$, которые может совершать пользователь при взаимодействии с веб-приложением. Последовательность действий $w \in \mathcal{A}^n$ будем называть *трассой*.

Определение 1. *Шаблоном называется размеченный конечный ациклический ориентированный граф без петель $\alpha = \langle V, \leq, g \rangle$, где V — множество вершин графа, $\leq$ — отношение частичного порядка между вершинами из V, $g\colon V \to \mathcal{A}$ — функция разметки вершин графа действиями.*

Отношение частичного порядка $\leq$ шаблона α отражает временные отношения между парами выполняемых пользователем действий — упорядоченность пары вершин означает, что соответствующие действия должны быть упорядочены и в трассе.

Необходимо для наблюдаемой трассы $w \in \mathcal{A}^n$ понять, содержит ли данная трасса поведенческие аномалии. Типичные сценарии использования заданы некоторым конечным множеством шаблонов.

Определение 2. *Трасса $w \in \mathcal{A}^n$ моделирует шаблон $\pi = \langle V, \leq, g \rangle$, $w \models \pi$, если существует функция соответствия $f\colon V \to [1; |w|]$ такая, что:*

1) $\forall\, v \in V\colon\ w_{f(v)} = g(v)$ — *каждой из вершин шаблона сопоставлено некоторое действие из трассы;*

2) $\forall\, u, v \in V\colon\ u \leq v \implies f(u) \leq f(v)$ — *свойство монотонности.*

Задача, которая ставится в данной работе: пусть задан шаблон π и трасса $w \in \mathcal{A}^n$, необходимо проверить $w \models \pi$.

Решение задачи поиска аномалий, предложенное в работе [6], имеет экспоненциальную временную сложность. Отличием данной задачи от описанной является отсутствие требования инъективности функции f в определении моделирования шаблона трассой. Отказ от требования инъективности функции f может увеличивать количество шаблонов, которые будут моделироваться некоторой трассой. Однако данный факт не является критичным для рассматриваемой предметной области, так как

предполагается, что множество шаблонов, описывающих типичные сценарии пользователя является минимальным, полным и неизбыточным.

Обозначим $cut_\pi(v)$ — сужение исходного шаблона $\pi = \langle V, \leq, g\rangle$ на множество вершин, из которых существует путь в вершину $v \in V$.

Теорема 1. *Пусть $w \in \mathcal{A}^n$, $x \in \mathcal{A}$, $\pi = \langle V, \leq, g\rangle$, $v \in V$. Тогда отношение $\models$ допускает следующее индуктивное определение:*

$$wx \models cut_\pi(v) \iff \Big(w \models cut_\pi(v)\Big) \vee \Big(g(v) = x\Big) \wedge \left(\bigwedge_{\substack{u \in V:\\ u<v}} wx \models cut_\pi(u)\right)$$

Доказательство. Структурная индукция по π и w аналогично доказательству теоремы о вычислительной сложности проверки выполнимости LTL-формулы в конечной трассе в работе [7]. □

Используя теорему 1, можно предложить следующий алгоритм поиска аномалий:

- Построение топологической сортировки $T(\pi) = \langle v_1, \ldots, v_n\rangle$ шаблона π.
- Если вычислено $w \models cut_\pi(v)$ для всех $v \in V$ и приходит новый запрос $x \in \mathcal{A}$, то вычисление $wx \models cut_\pi(v)$ для $v \in V$ происходит в порядке топологической сортировки T согласно теореме 1.

Данный алгоритм имеет временную сложность $O\left(|\leq| \cdot |w|\right)$.

Таким образом, изменение формальной постановки позволяет свести задачу поиска аномалий в конечной трассе к задаче выполнимости LTL-формулы в конечной трассе. Тем самым становится возможным построение алгоритма обнаружения аномалий в трассе, который является полиномиальным относительно длины трассы и размера шаблона.

Список литературы

[1] Chattopadhyay, P., Wang, L., Tan, Y. P., “Scenario-based insider threat detection from cyber activities”, *IEEE Transactions on Computational Social Systems*, **5**:3 (2018), 660–675.

[2] Cauteruccio F. и др., “A framework for anomaly detection and classification in Multiple IoT scenarios”, *Future Generation Computer Systems*, **114** (2021), 322–335.

[3] Sharma B., Pokharel P., Joshi B, “User Behavior Analytics for Anomaly Detection Using LSTM Autoencoder-Insider Threat Detection”, Proceedings of the 11th International Conference on Advances in Information Technology, 2020, 1–9.

[4] Kamra A., Terzi E., Bertino E, “Detecting anomalous access patterns in relational databases”, *The VLDB Journal*, **17 5** (2008 Springer), 1063–1077.

[5] Kruegel C., Vigna G., “Anomaly detection of web-based attacks”, Proceedings of the 10th ACM conference on Computer and communications security, 2003, 251–261.

[6] Leemans M., van der Aalst W. M. P., “Discovery of frequent episodes in event logs”, International symposium on data-driven process discovery and analysis, 2014 Springer, 1–31.

[7] Fionda V., Greco G., “The complexity of LTL on finite traces: Hard and easy fragments”, Proceedings of the AAAI Conference on Artificial Intelligence, **30 1** (2016).

On deviations from typical web application usage

Khashaev A.A., Teryokhina I.Yu., Gamayunov D.Yu.

The paper investigates the problem of finding deviations from the known scenarios of typical use of web applications. The modification of existing formal problem is considered, for which the recursive method for checking trace matching to a pattern is proposed. An algorithm for detecting anomalies in trace is developed, which is effective with respect to the trace length and the pattern size.

Keywords: anomaly detection, DAG, sequence analysis

References

[1] Chattopadhyay, P., Wang, L., Tan, Y. P., “Scenario-based insider threat detection from cyber activities”, *IEEE Transactions on Computational Social Systems*, **5**:3 (2018), 660–675.

[2] Cauteruccio F. et al., “A framework for anomaly detection and classification in Multiple IoT scenarios”, *Future Generation Computer Systems*, **114** (2021), 322–335.

[3] Sharma B., Pokharel P., Joshi B, “User Behavior Analytics for Anomaly Detection Using LSTM Autoencoder-Insider Threat Detection”, Proceedings of the 11th International Conference on Advances in Information Technology, 2020, 1–9.

[4] Kamra A., Terzi E., Bertino E, “Detecting anomalous access patterns in relational databases”, *The VLDB Journal*, **17 5** (2008 Springer), 1063–1077.

[5] Kruegel C., Vigna G., “Anomaly detection of web-based attacks”, Proceedings of the 10th ACM conference on Computer and communications security, 2003, 251–261.

[6] Leemans M., van der Aalst W. M. P., “Discovery of frequent episodes in event logs”, International symposium on data-driven process discovery and analysis, 2014 Springer, 1–31.

[7] Fionda V., Greco G., “The complexity of LTL on finite traces: Hard and easy fragments”, Proceedings of the AAAI Conference on Artificial Intelligence, **30 1** (2016).