Применение технологии Process Mining для обработки редких событий при фиксации эксплойта

К.А. Лифанова, К.С. Зайцев

***Аннотация*. Стремление к применению информационных технологий во всех сферах человеческой деятельности в последнее время требует новых подходов к управлению процессами. Настоящая статья посвящена решению задачи применения технологии Process Mining для выявления нелегитимных воздействий на различные процессы по информации журналов событий для обеспечения информационной безопасности. Для этого разработан алгоритм генерации искусственных журналов событий на основе цепи Маркова первого порядка, который успешно взаимодействует с алгоритмами Process Mining для поиска эксплойтов. Проведено исследование влияния размера и качественных характеристик промежуточной модели в виде системы переходов на размер и качественные характеристики целевой модели в виде сети Петри при конвертации с помощью алгоритма регионов. Полученные результаты позволяют повысить эффективность обработки больших наборов данных алгоритмами Process Mining с целью построения моделей процессов и, как следствие, улучшить применимость алгоритмов, чувствительных к размеру входных данных, при работе с большими журналами событий реальных информационных систем.**

***Ключевые слова* – анализ процессов, Process Mining, эксплойт, цепи Маркова, сеть Петри, журнал событий, вероятностная модель.**

I. ВВЕДЕНИЕ

Интеллектуальный анализ процессов (Process Mining) является достаточно молодой, но перспективной областью знаний, так как позволяет устранить разрыв между реальными событийными данными бизнес-процесса и управлением операционными процессами, тем самым провоцируя постоянное улучшение и оптимизацию процессной модели бизнес-подразделений.

Статья получена 27 августа 2021. Лифанова Ксения Алексеевна, Национальный Исследовательский Ядерный Университет МИФИ, магистрант, xlifanova@gmail.com
Зайцев Константин Сергеевич, Национальный Исследовательский Ядерный Университет МИФИ, профессор, KSZajtsev@mephi.ru

Также эта технология позволяет аналитикам получать более актуальную информацию о фактическом функционировании процессов для принятия более точных и обоснованных тактических решений [1]. Субъектами этой технологии являются бизнес-аналитики, владельцы процессов и исследователи данных, а объектами — бизнес-процессы. Процессы представляются различными абстракциями в виде статистических, графических или других видов моделей, а также конкретными записями событий, выполненных экземплярами процессов, в логах (журналах событий).

Одна из ключевых проблем, с которыми сталкивается технология Process Mining, — большое количество данных для анализа в логах, генерируемых программным обеспечением информационных систем (ПОИС). Эта проблема известна как проблема больших данных и, как известно, включает четыре аспекта: размер (volume), скорость генерирования (velocity), разнообразие (variety) и достоверность (veracity). Сегодня уже имеются и продолжается разработка эффективных алгоритмов синтеза и анализа моделей процессов. Они основаны на разных концепциях, различаются сложностью по времени/памяти, а также точностью и достоверностью получаемых с их помощью результатов [2].

Существует огромное количество областей, в которых активно внедряется технология Process Mining. В настоящей работе мы исследуем возможность ее применения для поиска аномальных ситуаций в журналах событий, вызванных действиями злоумышленников.

II. МОДЕЛИ И МЕТОДЫ

Классификация эксплойтов

По способу получения доступа к уязвленной системе, эксплойты делятся на локальные и удаленные. Для того чтобы запустить локальный эксплойт у злоумышленника должен быть доступ к целевому хосту. Как правило этот эксплойт используется для повышения прав пользователя. Не трудно заметить, что удаленный эксплойт

запускается через сеть без использования прямого доступа к системе.

Эксплойты используют уязвимости автоматизированных систем (АС). По этой причине логично будет привести классификацию по используемой уязвимости:

- эксплойты направленные на переполнение буфера;
- эксплойты направленные на SQL инъекции;
- эксплойты направленные на межсайтовый скриптинг (XSS) и межсайтовую подделку запроса (CSRF) атаки и т.д.

Описывать все возможные классификации по уязвимостям не имеет смысла, так как их огромное количество. На официальном сайте CVE [4] при необходимости можно ознакомиться с данной информацией.

В качестве примера решения задачи выявления эксплойта при анализе журналов событий, в настоящей работе исследована критическая уязвимость CVE-2020-0796 [4]. К основной проблеме уязвимости относится переполнение целочисленного элемента, он в свою очередь отвечает за контроль размера выделяемой приложением памяти.

Постановка задачи

Основным практическим приложением технологии «извлечение и анализ процессов» (process mining) является программное обеспечение (ПО) для аналитиков данных. ПО может носить исследовательский (академический) характер, быть ориентировано на индустриальное применение, либо совмещать в себе две эти функции. В зависимости от назначения, к ПО могут предъявляться различные требования: по функциональному наполнению, производительности, модульности, наличию интеграции с существующими инструментами бизнес-аналитики и т.д.

Одна и та же задача может решаться различными инструментами по-разному. Например, наличие возможности создания графических схем рабочего процесса (workflow) позволяет наглядно представить последовательность операций, выполняемых для синтеза/анализа/преобразования модели процесса. С другой стороны, для выполнения сложных аналитических сценариев с вовлечением большого числа компонентов и настроек удобным является наличие языков написания сценариев (скриптов). Большое значение имеет архитектура ПО и среда исполнения. Так, использование высокоуровневых языков программирования для разработки ПО, таких как Java и Python, позволяет получить рабочий прототип в сжатые сроки. При этом разработанные инструменты могут быть высокого качества, однако они будут страдать от ограничений, накладываемые использованием виртуальной машины. В первую очередь это ресурсные ограничения: решение одной и той же задачи такими инструментами будет в общем случае занимать больше времени и потреблять больше памяти, чем при использовании инструментов, реализованных с помощью нативных языков программирования (например, C++) [5].

Разработка программных инструментов, реализующих алгоритмы Process Mining наиболее эффективным образом, является одним из важнейших путей повышения применимости этих алгоритмов и, как следствие, методов аналитики, предлагаемых этой технологией.

Отдельной задачей является профилирование разрабатываемых инструментов для эффективной работы с большими массивами данных — большими журналами событий, производимыми реальными высоконагруженными информационными системами. Существенную поддержку в этом направлении оказывает адаптация существующих инструментов для работы большими данными для применения их в области Process Mining. В первую очередь это системы управления базами данных [2].

Требования к программным модулям прототипа системы подготовки данных для выбора модели Process Mining

Process Mining является исследовательской технологией, предлагающей методы и инструменты для анализа различных процессов. В настоящее время разработано большое число разных техник анализа процессов, связанных с такими направлениями, как добыча данных (DM), машинное обучение (ML), менеджмент бизнес-процессов (BPM) и др. Общим для всех этих техник является использование журналов событий в качестве источника данных, с которыми они работают, и модели процесса, которая может являться как результирующей, так и входной компонентой таких техник.

Журнал событий является точкой входа для любой задачи Process Mining. В зависимости от конкретной задачи журнал событий может использоваться с большей или меньшей интенсивностью. Обычно рассматриваются две разновидности журнала событий [5]:

1) Искусственные журналы событий, как правило, используются при разработке новых или изучения поведения существующих алгоритмов. Они генерируются специальным образом (вручную или с помощью специальных инструментов), и содержат данные, обладающие определенной спецификой.

2) Реальные журналы событий получаются в результате работы реальных информационных систем, выполняющих поддержку того или иного

процесса, либо накапливающие данные, которые после препроцессинга могут быть использованы для анализа. Как правило, реальные журналы событий содержат большое количество данных и имеют большой размер. Это служит причиной появления ряда проблем, которые следует учитывать при разработке программной подсистемы, работающей с журналом событий.

В большинстве случаев журнал событий представляет собой текстовый файл, содержащий подготовленную информацию о событиях. Примерами текстовых форматов журналов событий являются: MSXML, XES, CSV и др. Большинство разработанных к настоящему моменту инструментов для process mining ориентированы на журналы событий, представленные именно в таком виде. Тем не менее, некоторые инструменты используют внутреннее представление для оптимизации работы с такими журналами. Например, такие системы принимают файл лога на вход, производят импорт данных из него и сохраняют эти данные в оперативной памяти. Одним из основных побочных эффектов такого подхода является дублирование данных и ограничение на их размер, определяемое размером оперативной памяти, доступной для процесса.

В 2016 году IEEE был принят новый стандарт [6] обмена данными о событиях между информационными системами во многих областях приложений. Стандарт объявляет использование файлов XML-формата в качестве основы для представления журналов событий, структуру таких файлов в виде XML-схемы и возможность добавления дополнительных атрибутов, специфичных для предметной области process mining, с помощью стандартных расширений.

Журналы событий на основе XES широко используются в области анализа и извлечения процессов в последнее время. Технически, журнал в формате XES представляет собой текстовый XML-файл, структурированный следующим образом. Самым внешним XML-элементом является сам журнал (log), и сюда можно прикрепить произвольное количество атрибутов на уровне журнала. Следующий уровень представлен набором трасс, а каждая трасса (элемент trace) состоит из последовательности событий (event). И трассам, и событиям, и журналу могут быть назначены дополнительные атрибуты [7].

Фактически, каждый журнал XES содержит единственную встроенную перспективу. Переназначение заданного в журнале атрибута, представляющего активность, на другой атрибут, в целом возможно. Однако для перекомпоновки трасс в соответствии с новым атрибутом экземпляра процесса или трассы требуется полное перестроение журнала. Та же проблема возникает, когда необходимо осуществлять фильтрацию событий.

Таким образом, журналы на основе XES подходят в качестве стандарта для обмена данными о событиях, но довольно неудобны для практического использования. Существующие инструменты используют специальные подходы для работы с XES-журналами, например, полное отображение журнала в память (в инструменте ProM используется механизм MapDB), который имеет ограничения по использованию.

Большинство информационных систем (ИС), работающих с данными большого размера, используют специальные технологии для их эффективного хранения и обработки. Одним из наиболее распространенных подходов является использование технологий реляционных баз данных (РБД), поддерживаемых различным системами управления РБД (РСУБД). Такие базы данных могут содержать информацию, позволяющую выполнять анализ связанных с ней процессов различными путями. Это делает их незаменимым источником данных для Process Mining [8].

Программные инструменты Process Mining

Рассмотрим основные инструменты технологии Process Mining, наиболее часто упоминающиеся и получившие развитие в последнее время, связанные с темой настоящей работы.

Эти инструменты можно условно разбить на следующие три группы:

1) инструменты, представляющие графический пользовательский интерфейс (GUI) для моделирования и анализа. Сюда относятся ProM, RapidProM, DPMine/P, Disco, Celonis, Minit;

2) инструменты, реализованные в виде утилит командной строки: Petrify, Rbminer, genet;

3) инструменты, реализованные в виде пакетов-расширений к скриптовым языкам Python — PMLab, PM4Py и R — bupaR [9].

Из представленных инструментов поддержку графического управления потоками задач (workflow) имеют только два инструмента — RapidProM и DPMine/P, оба являются надстройками над инструментом ProM. Реализация какой-либо модификации алгоритма регионов есть только у инструментов ProM, Petrify, Rbminer, genet.

III. РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТА

Общее описание эксперимента

В рамках текущей работы для решения проблемы подготовки данных для выбора модели Process Mining было решено реализовать программный код для подготовки данных для

дальнейшей работы. В ходе работы экспертным путем было решено использовать вероятностную модель. Задача вероятностной модели – дополнить журнал событий недостающим в нем поведением для получения скорректированной модели процесса [10]. Таким образом, вначале построим вероятностную модель на основе имеющегося журнала событий. Далее используем вероятностную модель для искусственной генерации новых экземпляров процесса. Так как вероятностная модель имеет гибкость, то на выходе можно получить экземпляры процесса, которые не были представлены в исходном журнале процесса. Наличие такого дополнительного поведения должно увеличить точность получаемой модели и улучшить ее совокупные характеристики.

Экспериментальное исследование разработанного метода подготовки данных на реальных журналах событий показало практическую применимость этого метода для журналов с разной структурой и разных размеров. Для всех рассмотренных случаев удалось обнаружить область изменения параметров алгоритма частотной редукции, при которых метрики качества результирующей системы переходов и, следовательно, метрики синтезируемой на ее основе сети Петри, были лучше по сравнению с простой редукцией [10]. Во всех случаях оказывалось возможным осуществить редукцию до требуемых размеров, что позволяет применять метод синтеза сети Петри на основе предварительно редуцированной системы переходов для журналов событий любого размера. Это позволяет существенно расширить применимость алгоритма регионов, лежащего в основе указанного метода синтеза.

Выбор моделей и исследование их на эффективность по разным журналам событий

Задачи извлечения процессов, проверка соответствия и улучшение процессов формируют основу извлечения и анализа моделей процессов. Они используют различные модели процессов, включающие сети Петри, системы переходов, нечеткие карты, каузальные сети, BPMN, UML и др. В контексте Process Mining системы переходов используются как самостоятельные модели, так и промежуточные, на основе которых выполняется построение других типов моделей. В последнем случае применяются подходы, основанные на теории регионов, рассматриваемые в работе [11].

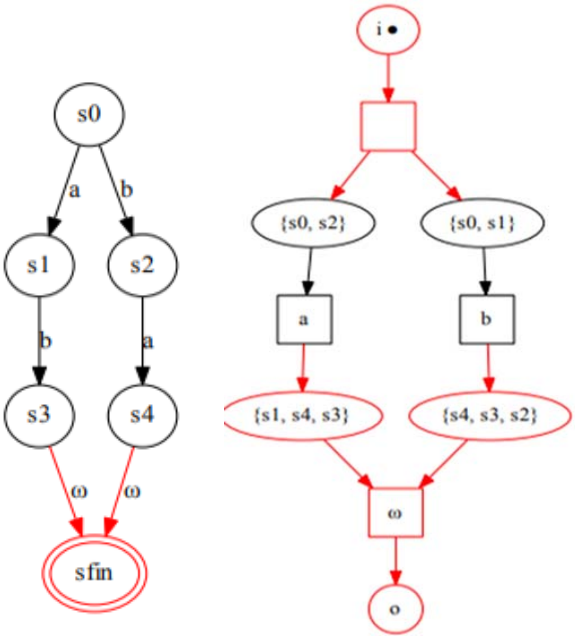
В настоящей работе наравне с метриками для деревьев процессов/сетей Петри мы определяем согласованные с ними метрики для моделей в виде систем переходов.

Степень соответствия определяет степень, до которой система переходов может воспроизводить трассы, записанные в журнале событий. Сложность определяет сложность модели. Она измеряется путем сравнения размера заданной системы переходов TS(L) с простейшей возможной системой переходов, которой является цветочная модель. Точность сравнивает системы переходов TS(L) с полной системой переходов, построенной для журнала событий L, которая рассматривается в качестве наиболее точной.

При моделировании некоторого бизнесс-процесса с помощью процессной модели можно говорить о том, что такая модель описывает жизненный цикл отдельных экземпляров этого процесса. Примерами таких процессов являются: заказ товара в интернет-магазине, размещение заявки на оказание услуги, размещение результатов тестирования в онлайн-образовании и др. Каждый такой процесс представляется индивидуальной моделью, и каждый случай возникновения такого процесса в индивидуальном контексте — со своими ресурсами факторами и своей временной линией — порождает новый экземпляр процесса. Каждый такой экземпляр процесса имеет четко обозначенные начало процесса и окончание процесса. Между началом и окончанием процесса составляющие его активности выполняются в соответствии с некоторыми заранее определенными правилами. Каждый процесс может инстанцироваться, то есть порождать очередной экземпляр процесса, множество раз. Отдельные экземпляры процессов могут выполняться как последовательно, так и с перекрытием вплоть до параллельного исполнения множества экземпляров.

В Process Mining отдельные экземпляры процессов, как правило, представляются в виде трассы в журнале событий. Трасса — упорядоченная последовательность активностей (activity) с сопоставленными им атрибутами. Одинаковые активности могут встречаться в одной трассе более, чем один раз. Появление определенной активности вместе с сопоставленными атрибутами называется событием (event). Упорядочение событий, как правило, осуществляется в соответствии с некоторыми сопоставленными им временными отметками, которые, в свою очередь, сами являются атрибутами событий. Наиболее часто это — время начала выполнения активности. На рисунке 1 представлена преобразованная система переходов (а) с единственным принимающим состоянием s_fin и соответствующая ей сеть потоков работ WF-сеть (б) с начальной позицией i и конечной позицией o, которая иллюстрирует синтез сетей потоков работ в алгоритме регионов.

Эта система содержит переходы, помеченные событиями a, b, x, где последнему соответствует петля (s2, x, s2).



а) б)

Рис. 1 - Преобразованная система переходов (а) с единственным принимающим состоянием s_fin и соответствующая ей WF-сеть (б) с начальной позицией i и конечной позицией o.

Если вопреки ограничению на наличие в системе переходов дуг подать TS1 на вход алгоритму, то после первого раунда отыскания минимальных пререгионов условие возбуждающего замыкания для события x не будет выполнено, так как множество найденных минимальных пререгионов для него будет пусто. Это приведет к выполнению расщепления событий a и b в попытке привести TS1 к виду возбужденно-замкнутой системы переходов (ECTS), однако на последующих раундах алгоритма пререгионы для события x по-прежнему не будут выделены, что приведет к бесконечному циклу.

Одним из ключевых этапов работы являлась проверка наших полученных моделей на соответствие «эталонному» журналу событий. Проверку было решено разбить на несколько этапов:

1. Использование модуля «Replay a Log On Petri Net For Conformance Analysis» для получения значения метрики «fitness» (пригодность). На вход подается модель бизнес-процесса и обрабатываемый журнал событий, на основе которого и будет проведена проверка бизнес-процесса.

2. Результат использования модуля «Replay a Log On Petri Net For Conformance Analysis» подается на вход модулю «Measure Precision/Generalization» для подсчета метрик «generalization» (обобщенности) и «precision» (точности).

Фрагмент результата подсчета метрик приведен в таблице 1.

Таблица 1 - Результаты подсчета метрик.

Размер выборки	Сеть Петри (выборка из исходного журнала)	Сесть Петри (искусственный журнал)
«fitness»		
50	0,999	0,967
142	0,999	0,971
200	0,999	0,961
…	…	…
«generalization»		
50	0,999	0,999
142	0,996	0,999
200	0,988	0,999
…	…	…
«precision»		
50	0,629	0,871
142	0,592	0,870
200	0,678	0,953
…	…	…

Таким образом, модель, построенная на основе сгенерированного журнала событий, имеет значительно большую точность, чем модель, построенная по «неполному» журналу.

Классический двухэтапный метод синтеза сети Петри для входного журнала событий ограничен в применении размером журнала. Реальные журналы событий могут иметь достаточно большой размер, в результате чего размер префиксного дерева, сконструированного для такого журнала, будет неприемлемо большим для синтеза алгоритмом регионов.

Сокращение размере системы переходов может быть осуществлена с помощью редукции— простой, или такой, которая принимает во внимание некоторые свойства процесса, записанного в журнале. Последнее позволяет достигать большего сокращения размера модели

при сохранении ее точности на приемлемом уровне.

Нами рассмотрен трехэтапный метод синтеза сети Петри по журналу событий, включающий построение префиксного дерева, редукции его в виде промежуточно системы переходов до размеров, пригодных к использованию с алгоритмом регионов, который выполняет синтез сети Петри на последнем этапе метода.

Качественные свойства синтезированной сети Петри — ее точность (метрика precision) и сложность (метрика simplicity) полностью соответствуют аналогичным метрикам системы переходов, на основе которой сеть Петри синтезируется. Это позволяет выполнять задание необходимых свойств конечной модели путем адаптивного изменения параметров алгоритма редукции без выполнения дорогостоящего (длительного) синтеза сети Петри.

Модифицированный алгоритм регионов корректно обрабатывает случаи, когда петли непосредственно появляются в системе переходов, когда петли представляются в виде «развертки», а также детектирует и корректно обрабатывает случаи, когда выделенный регион для некоторого события не является корректным пререгионом, если событие не моделируется петлей или ее «разверткой».

IV. ЗАКЛЮЧЕНИЕ

В статье, выполненной в рамках выпускной квалификационной работы магистра, проводилось исследование технологии Process Mining для обработки редких событий при фиксации эксплойта. Особое внимание было уделено разработке алгоритма генерации искусственных журналов событий на основе цепи Маркова первого порядка. Этот алгоритм взаимодействует с алгоритмами Process Mining для поиска эксплойтов. Также авторов интересовало исследование влияния характеристик промежуточной модели, представленной в виде системы переходов, на характеристики целевой модели, представленной в виде сети Петри, при конвертации с использованием алгоритма регионов.

В итоге, применения разработанного алгоритма на двух разных журналах событий было выявлено, что:

1) точность модели достигается за счет сохранения большого числа «важных» поведений в части префиксного дерева;

2) с увеличением порогового параметра алгоритма редукции Threshold в сторону аномальных ситуаций в компоненту модели начинает попадать все большее количество частотных поведений, что достаточно быстро приводит к деградации модели с точки зрения точности.

3) точность модели не обязательно ухудшается пропорционально уменьшению ее сложности. Напротив, в некотором диапазоне изменения дополнительного параметра Vwsc точность повышается.

4) компактная модель журнала событий без заметно выраженных «особенностей» оказывается более точной (за счет уменьшения числа возможных «отклонений»), чем большая реальная модель, не равная журналу событий (префиксному дереву).

БЛАГОДАРНОСТИ

Авторы выражают благодарность Высшей инжиниринговой школе НИЯУ МИФИ за помощь в возможности опубликовать результаты выполненной работы.

БИБЛИОГРАФИЯ

[1] Шершаков С.А. Методы и инструменты повышения эффективности алгоритмов майнинга процессов. – автореферат диссертации кандидата наук, ВШЭ, 2020 [электронный ресурс] https://www.hse.ru/data/2020/07/06/1595281148/%D0%A8%D0%B5%D1%80%D1%88%D0%B0%D0%BA%D0%BE%D0%B2_%D1%80%D0%B5%D0%B7%D1%8E%D0%BC%D0%B5.pdf Дата обращения: 02.04 Date of request.2021

[2] R. Andrews, C.G.J. van Dun, M.T. Wynn, W. Kratsch, M.K.E. Röglinger, A.H.M. ter Hofstede, Quality-informed semi-automated event log generation for process mining, Decision Support Systems, Volume 132, 2020.

[3] Günther C. W., Van Der Aalst W. M. P. Fuzzy Mining: Adaptive Process Simplification Based on Multi-perspective Metrics // Proceedings of the 5th International Conference on Business Process Management. — Brisbane, Australia : Springer-Verlag, 2007. — С. 328—343. — (BPM'07). — URL: http://dl.acm.org/citation.cfm?id=1793114.1793145 . Дата обращения: 22.02.2021.

[4] CVE. [электронный ресурс] https://cve.mitre.org/ . Дата обращения: 02.07.2020.

[5] The Overview Of Anomaly Detection Methods in Data Streams. Режим доступа: http://ceur-ws.org/ Дата обращения: 01.03.2020.

[6] ГОСТ Р 53114-2008 Национальный стандарт Российской Федерации. Защита информации. Обеспечение информационной безопасности в организации. Основные термины и определения. [электронный ресурс] http://docs.cntd.ru/document/gost-r-53114-2008 . Дата обращения: 01.07.2020.

[7] Common Vulnerability Scoring System, V3 Development Update. Режим доступа: https://www.first.org/cvss . Дата обращения: 12.05.2020.

[8] Систематика уязвимостей и дефектов безопасности программных ресурсов. [электронный ресурс] http://www.npo-echelon.ru/doc/is_taxonomy.pdf . Дата обращения: 01.06.2020.

[9] Angluin D. Inference of Reversible Languages // J. ACM. — New York, NY, USA, 1982. — Июль. — Т. 29, № 3. —

C. 741—765. — URL: http://doi.acm.org/10.1145/322326.322334. Дата обращения: 22.03.2021.

[10] Process Mining in Healthcare: Data Challenges When Answering Frequently Posed Questions. / R. Mans [и др.] // ProHealth/KR4HC. Т. 7738 / под ред. R. Lenz [and otc.]. — Springer, 2012. — С. 140—153. — (Lecture Notes in Computer Science). — URL: http://dblp.uni-trier.de/db/conf/ bpm/kr4hc2012.html#MansAVM12. Дата обращения: 21.02.2021.

[11] Buijs J., Dongen B., Aalst W. On the Role of Fitness, Precision, Generalization and Simplicity in Process Discovery // OTM Federated Conferences, 20th International Conference on Cooperative Information Systems (CoopIS 2012). Т. 7565 / под ред. R. Meersman [и др.]. — Springer-Verlag, Berlin, 2012. — С. 305—322. — (Lecture Notes in Computer Science).

Process mining technologies for handling rare events when an exploit is committed

X.A. Lifanova, K.S. Zaytsev

***Abstract* - The desire to apply information technology in all spheres of human activity recently requires new approaches to process management. This article is devoted to solving the problem of using the Process Mining technology to identify illegitimate influences on various processes based on information from event logs to ensure information security. For this, an algorithm for generating artificial event logs based on a first-order Markov chain has been developed, which successfully interacts with Process Mining algorithms to search for exploits. The study of the influence of the size and qualitative characteristics of the intermediate model in the form of a system of transitions on the size and qualitative characteristics of the target model in the form of a Petri net when converting using the algorithm of regions has been carried out. The results obtained will improve the efficiency of processing large data sets by Process Mining algorithms in order to build process models and, as a result, improve the applicability of algorithms that are sensitive to the size of input data when working with large event logs of real information systems.**

***Keywords* - Process Mining, exploit, Markov chains, Petri net, event log, probabilistic model**

REFERENCES

[1] Shershakov S.A. Methods and tools for increasing the efficiency of mining algorithms. - Abstract of Ph.D. thesis, HSE, 2020 [electronic resource] https://www.hse.ru/data/2020/07/06/1595281148/%D0%A8%D0%B5%D1%80%D1%88%D0%B0%D0%BA%D0%BE%D0%B2_%D1%80%D0%B5%D0%B7%D1%8E%D0%BC%D0%B5.pdf (Date of request 02.04.2021).

[2] R. Andrews, C.G.J. van Dun, M.T. Wynn, W. Kratsch, M.K.E. Röglinger, A.H.M. ter Hofstede, Quality-informed semi-automated event log generation for process mining, Decision Support Systems, V.132, 2020.

[3] Günther C. W., Van Der Aalst W. M. P. Fuzzy Mining: Adaptive Process Simplification Based on Multi-perspective Metrics // Proceedings of the 5th International Conference on Business Process Management. — Brisbane, Australia : Springer-Verlag, 2007. p. 328-343. (BPM'07). - URL: http://dl.acm.org/citation.cfm?id=1793114.1793145. (Date of request 22.02.2021).

[4] CVE. [online resource] // https://cve.mitre.org/ . (Date of request 02.07.2020).

[5] The Overview Of Anomaly Detection Methods in Data Streams. [online resource] // http://ceur-ws.org/ (Date of request 01.03.2020).

[6] GOST R 53114-2008 National standard of the Russian Federation. Protection of information. Ensuring information security in the organization. Basic terms and definitions. [online resource] // http://docs.cntd.ru/document/gost-r-53114-2008 . (Date of request 01.07.2020).

[7] Common Vulnerability Scoring System, V3 Development Update. Режим доступа: https://www.first.org/cvss . (Date of request 12.05.2020).

[8] Systematics of vulnerabilities and security defects of software resources. [online resource] // http://www.npo-echelon.ru/doc/is_taxonomy.pdf. (Date of request 01.06.2020).

[9] Angluin D. Inference of Reversible Languages // J. ACM. - New York, NY, USA, 1982. - July.- V. 29, No 3. p. 741-765. - URL: http://doi.acm.org/10.1145/322326.322334. (Date of request 22.03.2021).

[10] Process Mining in Healthcare: Data Challenges When Answering Frequently Posed Questions. / R. Mans [and etc.] // ProHealth/KR4HC. T. 7738 / под ред. R. Lenz [and etc.]. — Springer, 2012. - p. 140-153. - (Lecture Notes in Computer Science). - URL: http://dblp.uni-trier.de/db/conf/bpm/kr4hc2012.html#MansAVM12. (Date of request 21.02.2021).

[11] Buijs J., Dongen B., Aalst W. On the Role of Fitness, Precision, Generalization and Simplicity in Process Discovery // OTM Federated Conferences, 20th International Conference on Cooperative Information Systems (CoopIS 2012). T. 7565 / ed. R. Meersman [and etc.]. - Springer-Verlag, Berlin, 2012. -p. 305