(19) Евразийское патентное ведомство

(11) 047341 (13) B1

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОМУ ПАТЕНТУ

(45) Дата публикации и выдачи патента
2024.07.08

(21) Номер заявки
202393451

(22) Дата подачи заявки
2023.12.26

(51) Int. Cl. ***G06F 17/40*** (2006.01)
H04L 41/142 (2022.01)
H04L 43/04 (2022.01)
G06N 20/10 (2019.01)

(54) СПОСОБ И СИСТЕМА ВЫЯВЛЕНИЯ АНОМАЛЬНОГО ВЗАИМОДЕЙСТВИЯ УЗЛОВ ИНФОРМАЦИОННО-ВЫЧИСЛИТЕЛЬНОЙ СЕТИ

(31) **2023100804**
(32) **2023.01.16**
(33) **RU**
(43) **2024.07.04**

(71)(73) Заявитель и патентовладелец:
ПУБЛИЧНОЕ АКЦИОНЕРНОЕ ОБЩЕСТВО "СБЕРБАНК РОССИИ" (ПАО СБЕРБАНК) (RU)

(72) Изобретатель:
Вышегородцев Кирилл Евгеньевич, Нагорнов Иван Григорьевич, Кузьмин Александр Михайлович, Смирнов Дмитрий Владимирович (RU)

(74) Представитель:
Герасин Б.В. (RU)

(56) RU-C1-2699577
RU-C1-2769084
US-A1-20180196694
US-A1-20160219066
US-A1-20200244673

(57) Изобретение относится к области компьютерной техники, в частности к решениям для мониторинга работы узлов информационно-вычислительной сети (ИВС) и определения аномального взаимодействия между узлами. Изобретение представляет способ выявления аномального взаимодействия узлов информационно-вычислительной сети (ИВС) на основании анализа взаимодействия узлов с помощью формирования графа, на котором определяют значения кратчайших путей между узлами ИВС и группой узлов, взаимодействующих между собой при нормальной работе, что позволяет впоследствии выявлять узлы, которые находятся в разных группах, что свидетельствует об их аномальном взаимодействии.

047341 B1

047341 B1

Область техники

Изобретение относится к области компьютерной техники, в частности к решениям для мониторинга работы узлов информационно-вычислительной сети (ИВС) и определения аномального взаимодействия между узлами.

Уровень техники

Из уровня техники известен способ анализа узлов ИВС для оценки возможных аномальных явлений в работе ИВС (US 20200021607 A1, 16.01.2020). Решение раскрывает платформу безопасности для обнаружения аномалий и угроз в среде компьютерной сети. Платформа безопасности основана на "больших данных" (Big data) и использует машинное обучение для выполнения аналитики безопасности. Платформа безопасности выполняет анализ поведения пользователя /объекта (UEBA) для обнаружения аномалий и угроз, связанных с безопасностью, независимо от того, были ли такие аномалии/угрозы известны ранее. Выполняется анализ путей и режимов обнаружения аномалий и угроз в реальном времени, с последующей оценкой рисков сетевой безопасности узлов. Недостатками известного подхода является его недостаточная эффективность в части выявления аномальной работы узлов ИВС, в связи с тем, что не выполняется анализ активности узлов в части их взаимодействия между собой и использования данной информации для формирования моделей поведения узлов ИВС.

Сущность изобретения

Изобретение направлено на решение технической проблемы в части создания более эффективного подхода в анализе аномального взаимодействия между узлами в ИВС.

Техническим результатом является повышение эффективности и скорости выявления аномального взаимодействия между узлами ИВС, за счет постоянного мониторинга количества сообщений между узлами в рамках заданного кластера. Заявленный технический результат достигается за счет способа выявления аномального взаимодействия узлов информационно-вычислительной сети (ИВС), который содержит этапы, на которых:

a) получают данные обмена сообщениями между узлами ИВС, при этом данные содержат по меньшей мере информацию о количестве сообщений, передаваемых между упомянутыми узлами;

b) формируют граф на основании полученных данных, в котором вершинами являются идентификаторы узлов ИВС, а ребрами - факт обмена сообщениями между узлами, при этом каждое ребро имеет вес, характеризующий интенсивность обмена сообщениями между соответствующими узлами;

c) определяют с помощью полученного графа значения кратчайших путей между всеми узлами ИВС;

d) формируют для каждого узла ИВС векторное представление на основании полученных значений кратчайших путей;

e) понижают размерность полученных векторных представлений;

f) выполняют кластеризацию узлов ИВС на основании векторных представлений, полученных на этапе e);

g) формируют первую модель взаимодействия узлов ИВС, в которой узлы определены в группы в соответствии с выполненной кластеризацией;

h) формируют вторую модель взаимодействия узлов ИВС с помощью итеративного повторения этапов a) - g);

i) выполняют сравнение первой и второй моделей взаимодействия узлов ИВС, в ходе которого выявляют узлы ИВС, находящиеся в различных группах; и

j) формируют сигнал, характеризующий идентификаторы узлов ИВС, демонстрирующих аномальное взаимодействие, в ходе выполненного сравнения на этапе i).

В одном из частных вариантов осуществления способа вес ребра представляет собой величину, обратную количеству сообщений между узлами ИВС в единицу времени. В другом частном варианте осуществления способа этап e) выполняется с помощью алгоритма машинного обучения.

В другом частном варианте осуществления способа векторное представление узлов характеризует накопленную ретроспективную информацию о взаимодействии узла. Заявленный результат также достигается с помощью системы выявления аномального взаимодействия узлов ИВС, при этом система содержит по меньшей мере один процессор и по меньшей мере одну память, которая хранит машиночитаемые инструкции, которые при их исполнении процессором осуществляют вышеуказанный способ.

Краткое описание чертежей

Фиг. 1 иллюстрирует блок-схему выполнения заявленного способа.

Фиг. 2 иллюстрирует пример графа, сформированного на основании информации об узлах ИВС.

Фиг. 3 иллюстрирует схемы вычислительной системы.

Осуществление изобретения

На фиг. 1 представлена блок-схема выполнения заявленного способа (100) определения аномального взаимодействия узлов ИВС. На первом этапе (101) выполняется сбор данных об узлах ИВС, в частности, такими данными могут выступать идентификаторы устройств внутри сети, IP-адреса, MAC-адреса и т.п. Сбор данных может осуществляться с помощью автоматизированных решений, например, модулей-сборщиков данных, или с помощью программных компонент, или приложений, осуществляющих мони-

торинг ИВС. Далее по факту собранных данных об узлах ИВС на этапе (102) формируется граф (200), отображающий модель взаимодействия узлов. На фиг. 2 приведен пример формируемого графа (200) между узлами (201) - (206). Идентификаторы узлов формируют вершины графа, в то время как ребра характеризуют факт обмена сообщениями между узлами. Каждое ребро имеет вес, который зависит интенсивности обмена сообщениями между узлами (201) - (206) за единицу времени. Как представлено в примере на фиг. 2 количество сообщений, передаваемых между узлами (201)-(206), имеет следующий вид: 201-202: 4; 201-203: 5; 201-204: 2; 202-201: 8; 202-206: 14; 203-201: 3; 203-205: 2; 204-201: 5; 204-206: 3; 204-205: 7; 205-204: 12; 205-203: 4; 205-206: 8; 206-202: 6; 206-204: 1; 206-205: 5.

Характеристики интенсивности можно представить, как величину, обратную количеству сообщений от узла к узлу. При этом, если сообщений между узлами нет, то можно произвести добавление связи между такими узлами в данный граф (200). За количество сообщений между такими узлами можно определить величину, равную 1/(sum(N)), где N - количество сообщений, передаваемых между узлами. Тогда количество сообщений между такими узлами будет равно: 1/(4+5+2+8+14+...)=1/89=0,01123. На основании полученных данных формируется матричное представление связного графа, представленное в табл. 1. Для описания вышеуказанного графа (200) можно представить его в виде матрицы, наследуемой от матрицы смежности. В таком представлении каждая (i, j) позиция будет соответствовать связи между i и j узлами ИВС. Значениями в матрице будут расстояния dij, которые характеризуют длину пути в графе от i-го до j-го узла.

Таблица 1

Матричное представление связного графа

	201	202	203	204	205	206
201	89	1/4	1/5	1/2	89	89
202	1/8	89	89	89	89	1/14
203	1/3	89	89	89	1/2	89
204	1/5	89	89	89	1/7	1/3
205	89	89	1/4	1/12	89	1/8
206	89	1/6	89	1/1	1/5	89

На этапе (103) выполняется определение кратчайших путей на основании сформированного графа и матричного представления графа связанности. Поиск кратчайших путей может выполняться с помощью любого алгоритма поиска кратчайших расстояний между вершинами графа, например, Беллмана, Флойда-Уоршелла, Дейкстры, Джонсона и т.п. Пример значений кратчайших путей для некоторых узлов представлен в табл. 2.

Таблица 2

Значения кратчайших путей между узлами ИВС

	201	202	203	204	205	206
201	1/4+1/8	1/4	1/5	1/2	1/2+1/7	1/4+1/14
202	1/8	1/14+1/6	1/8+1/5	1/14+1	1/14+1/5	1/14

Далее на этапе (104) для каждого узла (201) - (206) на основании полученных значений кратчайших путей формируется векторное представление, которое может иметь следующий вид:

для узла (201) вектор=(3/8; 1/4; 1/5; 1/2; 9/14; 9/28); для узла (202) вектор=(1/8; 5/11; 13/40; 15/14; 1/35; 1/14). Поскольку ширина матрицы не фиксирована, то для перехода к векторному представлению связей узлов с заданным размером на этапе (105) формируется пространство меньшей размерности. Для этого могут применяться такие методы, как: матричное разложение, SVD (сингулярное матричное разложение), PCA (метод главных компонент), IncrementalPCA, KernelPCA, SparsePCA (Анализ разреженных основных компонентов), MiniBatchSparsePCA, ICA (независимый компонентный анализ), NMF или NNMF (неотрицательная матричная факторизация), LDA (Скрытое распределение Дирихле), FactorAnalysis (Факторный анализ), K-means квантизация для размерностей (К-средних), SOM для размерностей (Самоорганизующаяся карта Кохонена), LVQ (квантование векторов обучения), t-SNE (T-distributed Stochastic Neighbor Embedding), UMAP (Uniform Manifold Approximation and Projection), Автоэнкодеры. Каждая характеристика в таком векторе будет характеризовать соответствие некоторой латентной связанности. Под латентной связанностью понимается скрытая (латентная), неявная связь узла с другими узлами по нескольким параметрам в совокупности. В этом случае каждый параметр вектора характеризует связь узла не только с одним выбранным узлом, а с группой узлов. При этом формирование таких групп производится на основе схожести узлов (например, функциональное назначение узла, серверы управления и т.п.).

Векторное представление для узлов (201) - (206) с пониженной размерностью может иметь следующий вид:

201 (15/12 7/8),

202 (7/5 43/17),
203 (13/12 9/8),
204 (31/11 23/12),
205 (7/3 5/2),
206 (3/2 3/2).

На основании полученных на этапе (105) векторных представлений пониженной размерности выполняется дальнейшая кластеризация узлов ИВС на этапе (106). Кластеризация выполняется автоматически с учетом взаимодействия узлов в заданном временном промежутке. Для решения поставленной задачи могут применяться саморегулируемые алгоритмы (неконтролируемые) без начального задания количества кластеров, с учетом природы нормального или иного распределения.

Может применяться модель гауссовой смеси, которая предполагает, что данные должны быть разделены на кластеры таким образом, чтобы каждая точка данных в данном кластере соответствовала определенному многовариантному распределению Гаусса, а распределения многомерного гаусса каждого кластера не зависели друг от друга. Чтобы кластеризовать данные в такой модели, необходимо рассчитать апостериорную вероятность точки данных, принадлежащей данному кластеру с учетом наблюдаемых данных. Примерным методом для этой цели является метод Байеса. Поскольку существует только необходимость найти наиболее вероятный кластер для данной точки, можно использовать методы аппроксимации, т.к. они уменьшают вычислительную работу. Одним из лучших приближенных методов является использование метода вариационного байесовского вывода. Вариационное байевское гауссовское смешение - это максимизации математического ожидания, которое максимизирует нижнюю границу параметров модели (включая априорные вероятности) вместо вероятности данных.

Принцип, лежащий в основе вариационных методов, такой же, как и максимизация ожидания (то есть оба являются итерационными алгоритмами, которые чередуются между нахождением вероятностей для каждой точки, которая должна быть сгенерирована каждой смесью, и подгонкой смеси к этим назначенным точкам), но вариационные методы добавляют регуляризацию с интеграцией информации из предыдущих распределений. Это позволяет избежать особенностей, часто встречающихся в решениях максимизации ожидания, но вносит в модель некоторые тонкие искажения. Вывод часто происходит значительно медленнее, но обычно не настолько, чтобы сделать его использование нецелесообразным.

Из-за своей байесовской природы вариационный алгоритм требует больше гиперпараметров, чем максимизация математического ожидания, наиболее важным из которых является параметр концентрации. Задание низкого значения для концентрации заставит модель определить большую часть веса на несколько узлов ИВС, а веса остальных узлов в кластере будут очень близки к нулю. Высокие значения концентрации позволят большему количеству узлов быть в кластере.

Для моделирования концентрации применяется распределение Дирихле. Предпроцесс Дирихле - это априорное распределение вероятностей для кластеризации с бесконечным неограниченным числом разбиений. Вариационные методы позволяют включить эту априорную структуру в модели гауссовой смеси практически без потери времени расчёта по сравнению с моделью конечной гауссовой смеси. Большее значение концентрации будет заставлять формироваться более плотные кластеры. Предпроцесс Дирихле может использовать бесконечные и неограниченное число кластеров. Основа алгоритма заключается в итеративном процессе выборки объектов и основывается на подходе "ломка палки" (Stick-breaking). Начинается с полной выборки и на каждом шаге отделяется его часть. Каждый раз выполняется связывание узлов из выборки, которые попадают в кластеры. В конце выполняется связь узлов, не попадающих во все другие группы. В некоторых задачах это формирует определённый недостаток, который связан с тем, что все несвязанные точки (узлы графа) будут объединены в один общий, мусорный кластер. Однако этот кластер можно интерпретировать, как "кластер различных узлов", в котором узлы объединены по физическому смыслу в один кластер, не потому что имеют схожесть между собой по характеру соединений, а потому что имеют признак схожести между собой в части несхожести с узлами из других кластеров.

Пример разбиения на кластеры узлов (201) - (206) может иметь следующий вид:

Кластер I: (201, 203),
Кластер II: (202),
Кластер III: (205, 206),
Кластер IV: (204).

На этапе (107) по итогам выполненной кластеризации создается первая модель взаимодействия узлов ИВС, отображающая определение узлов в группе в части их взаимодействия между собой.

Далее на этапе (108) алгоритм осуществляет итеративное выполнение этапов (101)-(107) через заданный временной отрезок (например, день, неделя, и т.п.), по итогу чего формируется вторая модель взаимодействия узлов ИВС, которая может иметь следующий вид распределения кластеров:

Кластер I: (206, 205),
Кластер II: (201, 203, 202),
Кластер III: (204).

На этапе (109) выполняется сравнение полученных второй и первой моделей взаимодействия узлов

ИВС для выявления узлов, которые во второй модели поменяли свое размещение, что может говорить об аномальном взаимодействии с другими узлами ИВС. Из примера выше видно, что узел (202) попал в кластер с узлами (201), (203). Пример такого поведение можно рассмотреть в следующем случае. Пусть узлы (201) и (203) - это серверы с операционной системой Windows, узлы (205) и (206) - это серверы с операционной системой Ubuntu, а узлы (202) и (204) - это клиентские ЭВМ. В первой модели в один кластер попали узлы с операционной системой Windows, в другой кластер узлы с операционной системой Ubuntu. Клиентские ЭВМ не относятся ни к какому кластеру (в другом примере клиентские ЭВМ могут объединяться в один или несколько общих кластеров). При построении следующей модели клиентский ЭВМ (202) попал в кластер к серверным узлам. Это может свидетельствовать о том, что клиентский ЭВМ заражен вирусом и начал производить активную рассылку широковещательных или иных сообщений, что свойственно для серверов. С помощью заявленного способа осуществляется возможность оперативного реагирования на такого рода аномалию и формирования сигнала о появлении аномалии в ИВС для узла (202).

По факту выполнения этапа (109) на этапе (110) выполняется формирование сигнала с помощью системы контроля, оповещающего о факте аномального взаимодействия для определенного узла ИВС. Система контроля формирует сигнал, передаваемый, как правило, на устройство ответственного сотрудника службы кибербезопасности. Дополнительно может применяться изолирование выявленного узла, в части его отключения в сети ИВС.

Предлагаемый подход позволяет постоянно формировать ретроспективный "портрет" нормальной работы узлов ИВС. Данный "портрет" будет представлять собой накопление (интеграцию) множества предыдущих состояний хоста и отражает его изменение в истории. Подобная характеристика является гибкой к настройке, может концентрировать внимание на последних состояниях или наоборот, быть более консервативной. Интегральный исторический портрет позволит производить сравнение текущего, вновь получаемого состояния хоста с его предыдущими состояниями, и выявлять резкие (аномальные) изменения в работе. Данное детектирование аномальной работы даст возможность быстро выявлять нехарактерные изменения в хосте и проводить соответствующие расследования.

На фиг. 3 представлен общий вид вычислительной системы, реализованной на базе вычислительного устройства (300) и обеспечивающей выполнение заявленного способа (100). В общем случае, вычислительное устройство (300) содержит объединенные общей шиной информационного обмена один или несколько процессоров (301), средства памяти, такие как ОЗУ (302) и ПЗУ (303), интерфейсы ввода/вывода (304), устройства ввода/вывода (305), и устройство для сетевого взаимодействия (306).

Процессор (301) (или несколько процессоров, многоядерный процессор) может выбираться из ассортимента устройств, широкоприменяемых в текущее время, например, компаний Intel™, AMD™, Apple™, Samsung Exynos™, MediaTEK™, Qualcomm Snapdragon™ и т.п. Под процессором также необходимо учитывать графический процессор, например, GPU NVIDIA или ATI, который также является пригодным для полного или частичного выполнения способа (100). При этом средством памяти может выступать доступный объем памяти графической карты или графического процессора. ОЗУ (302) представляет собой оперативную память и предназначено для хранения исполняемых процессором (301) машиночитаемых инструкций для выполнения необходимых операций по логической обработке данных. ОЗУ (302), как правило, содержит исполняемые инструкции операционной системы и соответствующих программных компонент (приложения, программные модули и т.п.).

ПЗУ (303) представляет собой одно или более устройств постоянного хранения данных, например, жесткий диск (HDD), твердотельный накопитель данных (SSD), флэш-память (EEPROM, NAND и т.п.), оптические носители информации (CD-R/RW, DVD-R/RW, BlueRay Disc, MD) и др.

Для организации работы компонентов устройства (300) и организации работы внешних подключаемых устройств применяются различные виды интерфейсов В/В (304). Выбор соответствующих интерфейсов зависит от конкретного исполнения вычислительного устройства, которые могут представлять собой, не ограничиваясь: PCI, AGP, PS/2, IrDa, FireWire, LPT, COM, SATA, IDE, Lightning, USB (2.0, 3.0, 3.1, micro, mini, type C), TRS/Audio jack (2.5, 3.5, 6.35), HDMI, DVI, VGA, Display Port, RJ45, RS232 и т.п. Для обеспечения взаимодействия пользователя с вычислительным устройством (300) применяются различные средства (305) В/В информации, например, клавиатура, дисплей (монитор), сенсорный дисплей, тач-пад, джойстик, манипулятор мышь, световое перо, стилус, сенсорная панель, трекбол, динамики, микрофон, средства дополненной реальности, оптические сенсоры, планшет, световые индикаторы, проектор, камера, средства биометрической идентификации (сканер сетчатки глаза, сканер отпечатков пальцев, модуль распознавания голоса) и т.п.

Средство сетевого взаимодействия (306) обеспечивает передачу данных устройством (300) посредством внутренней или внешней вычислительной сети, например, Интранет, Интернет, ЛВС и т.п. В качестве одного или более средств (306) может использоваться, но не ограничиваться: Ethernet карта, GSM модем, GPRS модем, LTE модем, 5G модем, модуль спутниковой связи, NFC модуль, Bluetooth и/или BLE модуль, Wi-Fi модуль и др.

Дополнительно могут применяться также средства спутниковой навигации в составе устройства

(300), например, GPS, ГЛОНАСС, BeiDou, Galileo. Представленные материалы заявки раскрывают предпочтительные примеры реализации технического решения и не должны трактоваться как ограничивающие иные, частные примеры его воплощения, не выходящие за пределы испрашиваемой правовой охраны, которые являются очевидными для специалистов соответствующей области техники.

ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Способ выявления аномального взаимодействия узлов информационно-вычислительной сети (ИВС), содержащий этапы, на которых:

a) получают данные обмена сообщениями между узлами ИВС, при этом данные содержат по меньшей мере информацию о количестве сообщений, передаваемых между упомянутыми узлами;

b) формируют граф на основании полученных данных, в котором вершинами являются идентификаторы узлов ИВС, а ребрами - факт обмена сообщениями между узлами, при этом каждое ребро имеет вес, характеризующий интенсивность обмена сообщениями между соответствующими узлами;

c) определяют с помощью полученного графа значения кратчайших путей между всеми узлами ИВС;

d) формируют для каждого узла ИВС векторное представление на основании полученных значений кратчайших путей;

e) понижают размерность полученных векторных представлений;

f) выполняют кластеризацию узлов ИВС на основании векторных представлений, полученных на этапе e);

g) формируют первую модель взаимодействия узлов ИВС, в которой узлы определены в группы в соответствии с выполненной кластеризацией;

h) формируют вторую модель взаимодействия узлов ИВС с помощью итеративного повторения этапов a) - g);

i) выполняют сравнение первой и второй моделей взаимодействия узлов ИВС, в ходе которого выявляют узлы ИВС, находящиеся в различных группах; и

j) формируют сигнал, характеризующий идентификаторы узлов ИВС, демонстрирующих аномальное взаимодействие, в ходе выполненного сравнения на этапе i).

2. Способ по п.1, в котором вес ребра представляет собой величину, обратную количеству сообщений между узлами ИВС в единицу времени.

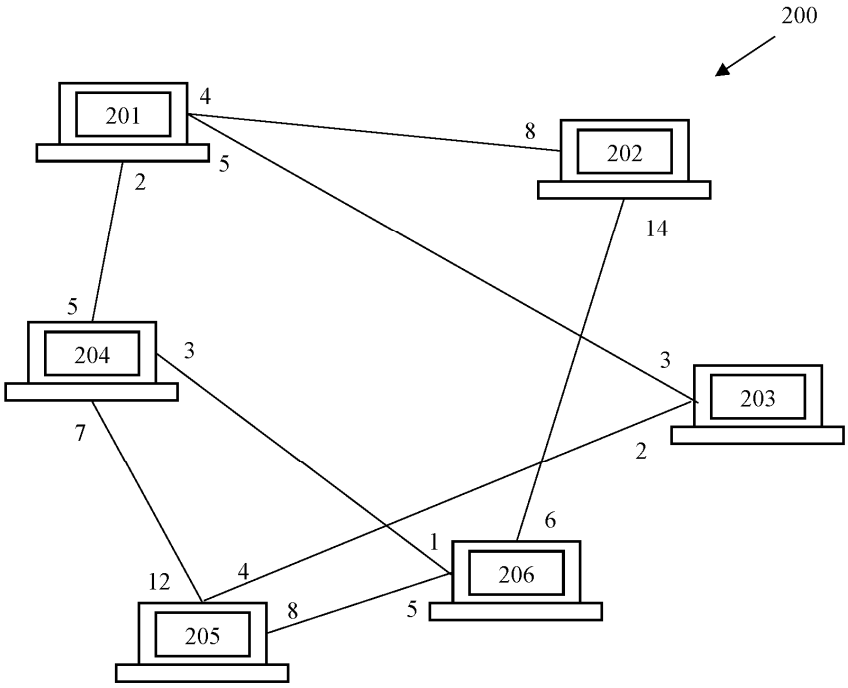
3. Способ по п.1, в котором этап e) выполняется с помощью алгоритма машинного обучения.

4. Способ по п.1, в котором векторное представление узлов характеризует накопленную ретроспективную информацию о взаимодействии узла.

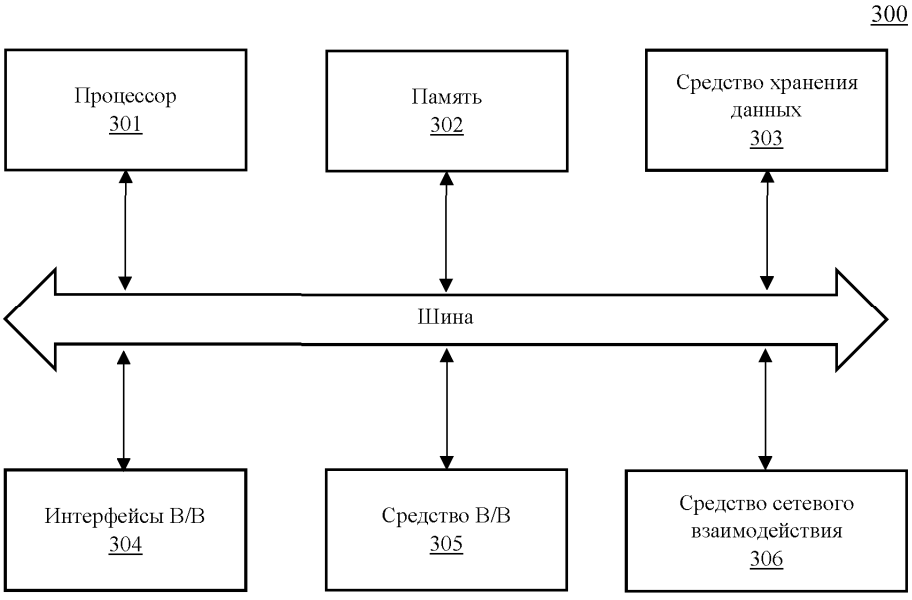
5. Система выявления аномального взаимодействия узлов ИВС, содержащая, по меньшей мере, один процессор и, по меньшей мере, одну память, которая хранит машиночитаемые инструкции, которые при их исполнении процессором осуществляют способ по любому из пп.1-4.


100
Сбор данных обмена между узлами ИВС
101
Формируют граф, отображающий взаимодействие узлов ИВС
102
Определяют кратчайшие пути обмена между узлами ИВС
103
Формирование векторных представлений кратчайших путей
104
Понижение размерности векторных представлений
105
Кластеризация узлов на основании векторных представлений
106
Формирование первой модели взаимодействия узлов
107
Формирование второй модели взаимодействия узлов
108
Сравнение моделей между собой
109
Определение узлов в ИВС, демонстрирующих аномальное взаимодействие
110

Фиг. 1



Фиг. 2



Фиг. 3

Евразийская патентная организация, ЕАПВ
Россия, 109012, Москва, Малый Черкасский пер., 2