(19) Евразийское патентное ведомство

(11) 042566 (13) B1

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОМУ ПАТЕНТУ

(45) Дата публикации и выдачи патента
2023.02.27

(21) Номер заявки
202092232

(22) Дата подачи заявки
2020.10.20

(51) Int. Cl. ***G06F 17/00*** (2019.01)
G06Q 20/40 (2012.01)

(54) СПОСОБ И СИСТЕМА НАХОЖДЕНИЯ СХОЖИХ МОШЕННИЧЕСКИХ ГРУПП ПО ГРАФОВЫМ МОДЕЛЯМ

(31) **2020117652**
(32) **2020.04.28**
(33) **RU**
(43) **2021.10.29**

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

(72) Изобретатель:
Оболенский Иван Александрович,
Сысоев Валентин Валерьевич,
Харитонов Александр Сергеевич,
Ключников Александр Валерьевич
(RU)

(74) Представитель:
Герасин Б.В. (RU)

(56) CN-A-107730262
CN-A-110111110
CN-A-109993538
US-B1-7562814

(57) Изобретение относится к способам обработки данных с помощью компьютерных систем, в частности к сравнению графовых моделей между собой с целью выявления схожих графовых моделей. Техническим результатом является выявление схожих (но не обязательно изоморфных) мошеннических схем, которые в дальнейшем можно соотнести к мошенническим группировкам или на основе анализа выделить новые. Заявленный результат достигается за счет компьютерно-реализуемого способа сравнения двух графов, выполняемый с помощью процессора и оперативной памяти, при котором: для каждой из графовых моделей, составленной из транзакционного потока составляется каркас графовой модели путем переноса без дублирования в каркас графовой модели ребер и вершин из множества путей. Где путь в конкретном случае - цепь графовой модели, имеющая расстояние, равное диаметру графа. Для составления множества путей применяется итерационный алгоритм - для каждой вершины строятся цепи со всеми другими вершинами графовой модели, длина которых равна их расстоянию. Наибольшая длина таких цепей и будет диаметром графовой модели, а сами цепи и будут составлять при переносе всех их вершин и ребер без повторения каркас графовой модели. Далее рассчитывается коэффициент подобия, показывающий насколько графовые модели подобны друг другу. Коэффициент подобия POD рассчитывается по формуле POD= DG×KLW×PV, где DG - отношение диаметров обоих графов; KLW - отношение количеств путей в графовых моделях; PV - отношение отношений количества вершин, входящих в каркас к общему количеству вершин в графовой модели. При коэффициенте подобия POD, равном 100%, две графовые модели изоморфны друг другу. На основании коэффициента подобия и определяется принадлежность мошеннических схем к той или иной мошеннической группе или заводится новая группа в картотеке.

042566 B1

042566 B1

Область техники

Изобретение относится к способам обработки данных с помощью компьютерных систем, в частности к способу и системе поиска схожих мошеннических групп по графовым моделям транзакционных данных.

Уровень техники

Проблема анализа банковских транзакций на предмет выявления сомнительных и/или мошеннических операций заключается в том, что для эффективной работы необходимо осуществить большой объем вычислений данных транзакций за небольшой промежуток времени. В частности, есть необходимость в поиске похожих мошеннических групп при известных образах мошеннических транзакций.

Известно решение для отображения и анализа транзакционных потоков, в котором для обработки данных применяется принцип построения графовой модели (US20020156724, PayPal Inc., 24.10.2002). В известном решении графовая модель применяется для анализа узлов совершения транзакций, чтобы отслеживать движение транзакционного потока и визуально представлять маршрут их движения с помощью графовой модели.

В известном решении принцип построения маршрута движения транзакционных потоков также может использоваться для выявления мошеннической активности или узлов графа, которые являются сомнительными и подлежат дополнительной проверке вне работы системы.

Из патента RU 2699677 известно решение нахождения наикратчайших путей методом построения итераций для каждого узла графовой модели (ПАО Сбербанк, 06.09.2019).

В известном решении для каждого узла графовой модели составляется итерационный список. На основе итерационных списков строятся расстояния между каждым узлом графовой модели, что позволяет оценивать, как далеко, исходя из расстояния, располагаются узлы графа. Однако приведенное в аналоге решение не позволяет численно оценить степень схожести той или иной графовой модели исходя из построения узлов и ребер для целей анализа принадлежности узлов графа к известным мошенническим реквизитам.

Раскрытие изобретения

Заявленным решением предлагается новый подход в решении существующей технической проблемы, который заключается в анализе графовых моделей транзакционных данных известных мошеннических схем, что позволяет выявлять схожие мошеннические схемы и выявлять мошеннические группы по мошенническим схемам и/или указать на то, что в мошеннической схеме была замешана та или иная мошенническая группа.

Технический результат заключается в обеспечении идентификации узлов, связанных со схемами осуществления мошеннических финансовых транзакций.

Заявленный результат достигается за счет компьютерно-реализуемого способа поиска мошеннических транзакций, выполняемого с помощью процессора, при котором:

a) осуществляют получение данных о транзакциях, в которых все данные относятся к мошенническим транзакциям;

b) формируют на основе полученных данных графы, в которых узлами являются данные по транзакциям, а ребрами - выполненные транзакции или связи с атрибутами;

c) для каждого графа, построенного на шаге b), выполняются следующие шаги:

определяются списки итераций для каждой вершины;

определяются для каждого узла графа с помощью итерационного алгоритма расстояния для других узлов;

определяются цепи с максимальным расстоянием;

осуществляют формирование каркаса графа на основе определенных цепей на этапе c), при этом упомянутый каркас состоит из всех вершин и ребер упомянутых цепей;

на основании сформированного каркаса определяют диаметр графа, который представляет собой расстояние любой из его цепей;

определяют количество цепей в каркасе графа;

определяют плотность каркаса как соотношение количества вершин каркаса графа к общему количеству вершин в графе;

d) осуществляют сравнение графов, полученных на этапе b), при котором выполняется

i. расчет отношения диаметров графов;

ii. расчет отношения количества цепей каркаса;

iii. расчет отношения плотностей каркасов сравниваемых графов;

iv. расчет коэффициентов подобия на основании значений отношений, полученных на этапах i), ii) и iii);

e) определяется по меньшей мере один граф, схожий с по меньшей мере одной известной мошеннической схемой, представленной в виде графа; и

f) определяют данные транзакций, связанные с по меньшей мере одной известной мошеннической схемой.

В одном из частных вариантов осуществления способа данные транзакций выбираются из группы:

идентификатор устройства, IP-адрес, номер счета, PAN платежной карты, номер телефона, данные плательщика или получателя платежа, или их сочетания. В другом частном варианте осуществления на основании вычисленного коэффициента подобия в базе данных создается запись о реквизитах, относящихся к новой группировке мошенников, или формируют связь анализируемых данных транзакций с по меньшей мере одной известной мошеннической группировкой.

Заявленное изобретение также реализуется за счет компьютерной системы поиска мошеннических транзакций, причем система содержит по меньшей мере один процессор и по меньшей мере одну память, содержащую машиночитаемые инструкции, которые при их выполнении с помощью процессора осуществляют вышеуказанный способ.

Описание фигур

Фиг. 1 иллюстрирует блок-схему реализации представленного способа.

Фиг. 2 иллюстрирует множество анализируемых графов.

Фиг. 3 иллюстрирует пример определения каркасов графов.

Фиг. 4 иллюстрирует пример сравнения графов.

Фиг. 5 иллюстрирует общий вид вычислительного устройства.

Осуществление изобретения

Дальнейшее описание примера осуществления заявленного решения будет представлено в соответствие с отсылками к представленным фигурам чертежей.

Согласно фиг. 1 заявленный способ сравнения мошеннических транзакций (100) выполняется с помощью вычислительного устройства, например компьютера.

На первом шаге (101) осуществляется сбор множества данных по транзакциям, в которых существуют мошеннические схемы. Транзакция - это банковская операция между двумя субъектами. Данные по транзакциям могут поступать из различных источников информации, например, из платежных систем, POS-терминалов, процессинговых систем и др., а также могут передаваться по любому протоколу из стэка TCP/IP. Транзакции аккумулируются и хранятся, как правило, в базе данных (БД) компьютерного устройства, например сервера.

Следующим шагом является создание графов по транзакциям на этапе (102). Как представлено на фиг. 2 по полученным транзакциям формируется множество графов

$$GM=\{G_1...G_n\} \quad (200),$$

где каждый граф (201)-(203) G_i - невзвешенный неориентированный граф

$$G:=(V, E),$$

где V - непустое множество узлов, а

E - непустое множество неупорядоченных ребер,

n - количество графов.

В качестве данных по транзакции могут выступать: идентификатор устройства (например, смартфона), IP-адрес, данные геолокации, номер счета, PAN платежной карты, номер телефона, данные плательщика или получателя платежа, или их сочетания. Данные по транзакции позволяют точно определить отправителя и получателя денежных средств.

Для каждого графа G (201)-(203) из множества полученных графов GM (200) выполняется последовательный алгоритм, который включает в себя следующие этапы. На этапе (103) для каждой вершины каждого из графов G_i (201)-(203) находится список окрестностей V:

$$OKR_V=\{OKR_{[0]}, OKR_{[1]} ... OKR_{[i]}\},$$

где i - целое число, индекс удаленности от узла V, определяемый количеством ребер между исследуемыми вершинами.

На этапе (104) выполняется определение расстояния для каждой вершины графов (201)-(203) до других вершин соответствующего графа. Данный этап реализуется с помощью итерационного алгоритма поиска кратчайшей цепи, описанного в патенте RU2699677 (ПАО Сбербанк, 06.09.2019), $\forall V$:

$$L_V=\{L_1...L_j\},$$

где j - целое число, j=i-1.

Цепь - представляет собой маршрут, все рёбра которого различны; число рёбер определяет длину цепи.

Выявляются все цепи с максимальным расстоянием (этап 105):

$$W=\{L_{V1}...L_{Vm}\},$$

где m - целое число, количество цепей с максимальным расстоянием.

Далее, как показано на фиг. 3, на этапе (106) для каждого графа (201)-(203) формируется каркас графа (2011, 2021, 2031):

$$KG:=(VK, EK),$$

где $VK=VW_1 \cup VW_2 \cup ... \cup VW_k$ - объединение вершин всех цепей множества W каждого графа G_i (201)-(203),

$EK=EW_1 \cup EW_2 \cup ... \cup EW_k$ - объединение ребер всех цепей множества W каждого графа G_i (201)-(203),

k - целое число, количество цепей множества W каждого из графов (201)-(203).

Далее на этапе (107) выполняется определение диаметра DG каждого графа (201)-(203) из множества GM (200). Диаметр графа DG равен расстоянию любой из цепей в множестве W для соответствующего графа из множества GM (200). На этапе (108) определяется количество цепей KLW из множества W в полученных каркасах KG (2011, 2021, 2031) каждого графа G_i (201)-(203),

$$KL_W=|W|.$$

По итогу вычисления количества цепей KLW на этапе (109) определяется отношение количества вершин VK в каркасах графов KG (2011, 2021, 2031) к общему количеству вершин V в соответствующем графе G_i (201)-(203), т.е. PV=|VK|/|V|, $0<PV\leq 1$. Коэффициент PV отображает плотность графа и показывает, как много у графов (201)-(203) вершин, не входящих в сформированный каркас KG (2011, 2021, 2031), и, соответственно, как сильно каркас отличается от соответствующего графа, для которого он был сформирован. При PV=1 граф и его каркас изоморфны, и чем больше значение PV отличается от 1, тем больше вершин не входят в каркас.

Далее осуществляется сравнение графов из множества GM (200). Сравнение графов выполняется попарно, для этого для каждой пары графов $\{G_1, G_2\} \in GM$ определяется следующее. На этапе (110) выполняется вычисление соотношения диаметров графов G_1 и G_2

$$dDG=\min(DG_1,DG_2)/\max(DG_1,DG_2).$$

Далее на этапе (111) выполняется определение соотношения количеств цепей множеств W в графах G_1 и G_2

$$dKLW=\min(KLW_1,KLW_2)/\max(KLW_1,KLW_2),\ 0<dKLW\leq 1.$$

На этапе (112) определяется соотношение коэффициентов плотности каркасов PV графов G_1 и G_2

$$dPV=\min(PV_1,PV_2)/\max(PV_1,PV_2),\ 0<dPV\leq 1.$$

После чего на этапе (113) вычисляется коэффициент подобия графов

$$POD=dDG\times dKLW\times dPV,\ 0<POD\leq 1.$$

По итогам на этапе (114) выполняется сравнение коэффициентов подобия двух графов, по итогам которого чем ближе коэффициент подобия POD к 1, тем более похожи графы G_1 и G_2 между собой, при коэффициенте подобия POD=1 графы G_1 и G_2 изоморфны.

Выявление схожих графов с помощью представленного алгоритма можно рассмотреть на следующем примере, представленном на фиг. 4.

Выполняется получение данных по транзакциям между субъектами, характеризующимися реквизитами и атрибутами транзакций. Реквизиты и атрибуты транзакций представляют собой в частном случае идентификаторы транзакций, по которым можно определить отправителя и получателя транзакции, т.е. узлы, между которыми произошел денежный перевод. В рассматриваемом примере реквизиты выбираются из группы: номер счета, PAN платежной карты, номер телефона, данные плательщика или получателя платежа, или их сочетания, а атрибуты из группы: идентификатор устройства (например, смартфон), IP-адрес, данные геолокации или их сочетания. Из вышеописанных данных формируются графовые модели G_1 (201), G_2 (202), G_3 (203). При этом существует отдельная база данных, содержащая данные по транзакциям мошеннических групп, имеющая так же реквизиты и атрибуты. Из этой базы формируется графовая модель мошеннических транзакций GH (210).

Все сформированные графовые модели представляют собой множество графов $GM=\{GH,G_1,G_2,G_3\}$, по которым составляются каркасы каждого графа:

a. для графа GH - каркас графа KGH (2101),

b. для графа G_1 - каркас графа KG_1 (2011),

c. для графа G_2 - каркас графа KG_2 (2021),

d. для графа G_3 - каркас графа KG_3 (2031).

Определяем характеристики каждого графа:

a. для графа GH (210) вычисляются следующие характеристики:

i. диаметр графа DG=3,

ii. количество цепей KLW=3,

iii. плотность графа PV=0.87;

b. для графа G_1 (201):

i. диаметр графа DG=1,

ii. количество цепей KLW=4,

iii. плотность графа PV=1;

c. для графа G_2 (202):

i. диаметр графа DGH=4,

ii. количество цепей KLW=2,

iii. плотность графа PV=0.75;

d. для графа G_3 (203):

i. диаметр графа DGH=1,

ii. количество цепей KLW=5,

iii. плотность графа PV=1.

Далее выполняется сравнение графов множества GM, сравнение выполняется попарно на основании чего вычисляется коэффициент подобия по каркасам графов:

a. сравнивая GH и G_1 - POD=0.22,

b. сравнивая GH и G_2 - POD=0.43,

c. сравнивая GH и G_3 - POD=0.08,

d. сравнивая G_1 и G_2 - POD=0.09,

e. сравнивая G_1 и G_3 - POD=0.8,

f. сравнивая G_2 и G_3 - POD=0.08.

В результате расчетов при сравнении графов GH (210) и G_2 (202) есть подозрение, что в мошеннической схеме G_2 (202) причастна группа мошенников из мошеннической схемы GH (210), ввиду того что коэффициент подобия из всех сравниваемых попарно графов - больше всего.

Смысл коэффициента подобия - чем он ближе к 1, тем выше вероятность, что мошеннические схемы похожи и к их осуществлению причастна одна и та же группа лиц.

При коэффициенте подобия, равном 1, графы являются изоморфными, а, следовательно, графовые схемы, построенные на основе данных транзакций - идентичными.

По результату работы алгоритма сравнения графов (100) определяются мошеннические схемы, похожие друг на друга с точки зрения данных по транзакциям, выявляются реквизиты и атрибуты, используемые обеими мошенническими схемами.

Исходя из сравнительного анализа характера изменения используемых мошенниками номеров счетов, PAN платежных карт, номеров телефонов из схожих мошеннических схем принимается решение об массовой блокировке группы (пулы) реквизитов и атрибутов, которые являются обобщением выявленных реквизитов и атрибутов у схожих мошеннических схем.

Также, в результате выполнения способа (100) принимается решение о создании в каталоге мошеннических схем новой группы мошенников, или причисление мошеннической схемы к существующей группе.

С помощью заявленного способа (100) появляется возможность выявлять схожие мошеннические схемы, к которым могут применяться однотипные меры противодействия на этапе их формирования, а также объединять выявленные мошеннические схемы в преступные сообщества и выявлять организаторов преступных сообществ на основе дополнительного анализа связей схожих мошеннических схем с помощью анализа графов, формируемых на основании данных о транзакциях.

На фиг. 5 представлен общий вид вычислительной системы, реализованной на базе вычислительного устройства (300). 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) могут выбираться из ассортимента устройств, широко применяемых 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) посред-

ством 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. Компьютерно-реализуемый способ поиска мошеннических транзакций, выполняемый с помощью процессора, при котором:

a) осуществляют получение данных о транзакциях, в которых все данные относятся к мошенническим транзакциям;

b) формируют на основе полученных данных графы, в которых узлами являются данные по транзакциям, а ребрами - выполненные транзакции или связи с атрибутами;

c) для каждого графа, построенного на шаге b), выполняются следующие шаги:

определяются списки итераций для каждой вершины;

определяются для каждого узла графа с помощью итерационного алгоритма расстояния для других узлов;

определяются цепи с максимальным расстоянием;

осуществляют формирование каркаса графа на основе определенных цепей на этапе c), при этом упомянутый каркас состоит из всех вершин и ребер упомянутых цепей;

на основании сформированного каркаса определяют диаметр графа, который представляет собой расстояние любой из его цепей;

определяют количество цепей в каркасе графа;

определяют плотность каркаса как соотношение количества вершин каркаса графа к общему количеству вершин в графе;

d) осуществляют сравнение графов, полученных на этапе b), при котором выполняется

i. расчет отношения диаметров графов;

ii. расчет отношения количества цепей каркаса;

iii. расчет отношения плотностей каркасов сравниваемых графов;

iv. расчет коэффициентов подобия на основании значений отношений, полученных на этапах i), ii) и iii);

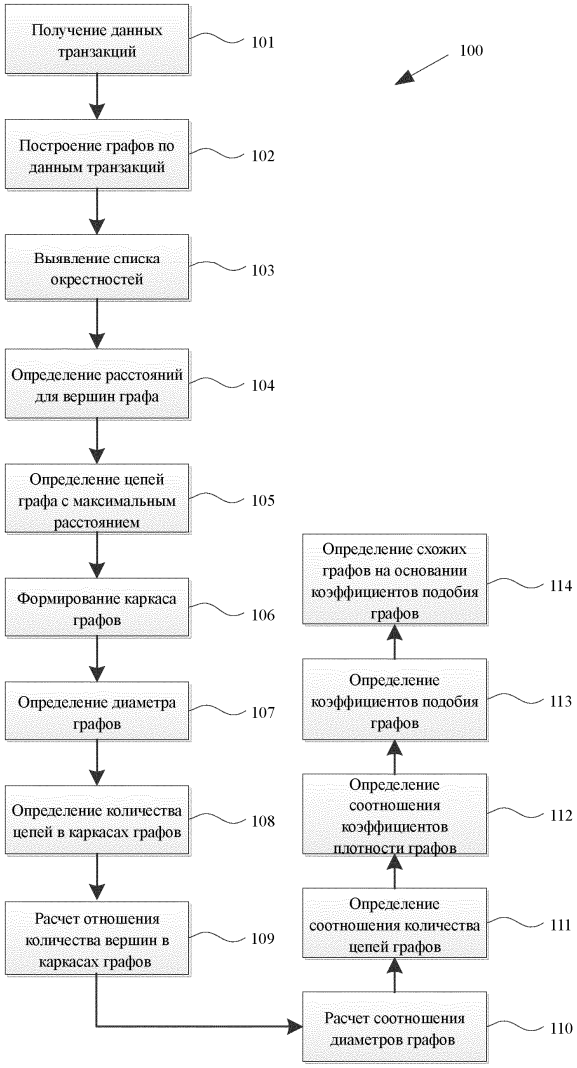
e) определяется по меньшей мере один граф, схожий с по меньшей мере одной известной мошеннической схемой, представленной в виде графа; и

f) определяют данные транзакций, связанные с по меньшей мере одной известной мошеннической схемой.

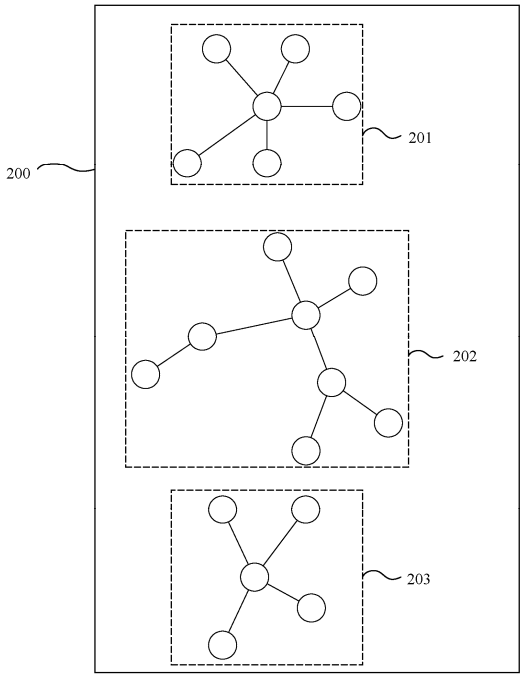
2. Способ по п.1, характеризующийся тем, что данные по транзакции выбираются из группы: идентификатор устройства, и/или IP-адрес, и/или данные геолокации совершения транзакции, и/или номер счета, и/или PAN платежной карты, и/или номер телефона плательщика, и/или данные плательщика или получателя платежа, или их сочетания.

3. Способ по п.1, характеризующийся тем, что на основании вычисленного коэффициента подобия в базе данных создается запись о реквизитах, относящихся к новой группировке мошенников, или формируют связь анализируемых данных транзакций с по меньшей мере одной известной мошеннической группировкой.

4. Система поиска мошеннических транзакций, содержащая по меньшей мере один процессор и по меньшей мере одну память, содержащую машиночитаемые инструкции, которые при их выполнении с помощью процессора осуществляют способ по любому из пп.1-3.

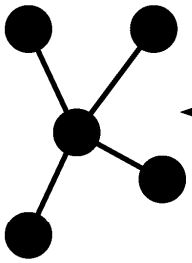

Получение данных транзакций
101
100
Построение графов по данным транзакций
102
Выявление списка окрестностей
103
Определение расстояний для вершин графа
104
Определение цепей графа с максимальным расстоянием
105
Формирование каркаса графов
106
Определение диаметра графов
107
Определение количества цепей в каркасах графов
108
Расчет отношения количества вершин в каркасах графов
109
Расчет соотношения диаметров графов
110
Определение соотношения количества цепей графов
111
Определение соотношения коэффициентов плотности графов
112
Определение коэффициентов подобия графов
113
Определение схожих графов на основании коэффициентов подобия графов
114

Фиг. 1


200
201
202
203

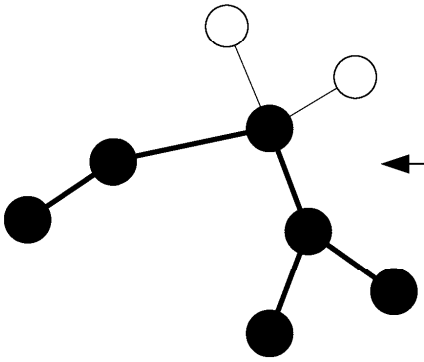
Фиг. 2

DG1=1
kGK1=4
dV1 = |VK|/|V| = 5/5 =1



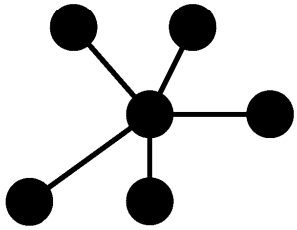
2011

DG2=4
kGK2=2
dV2 = |VK|/|V| = 7/8 =0,75



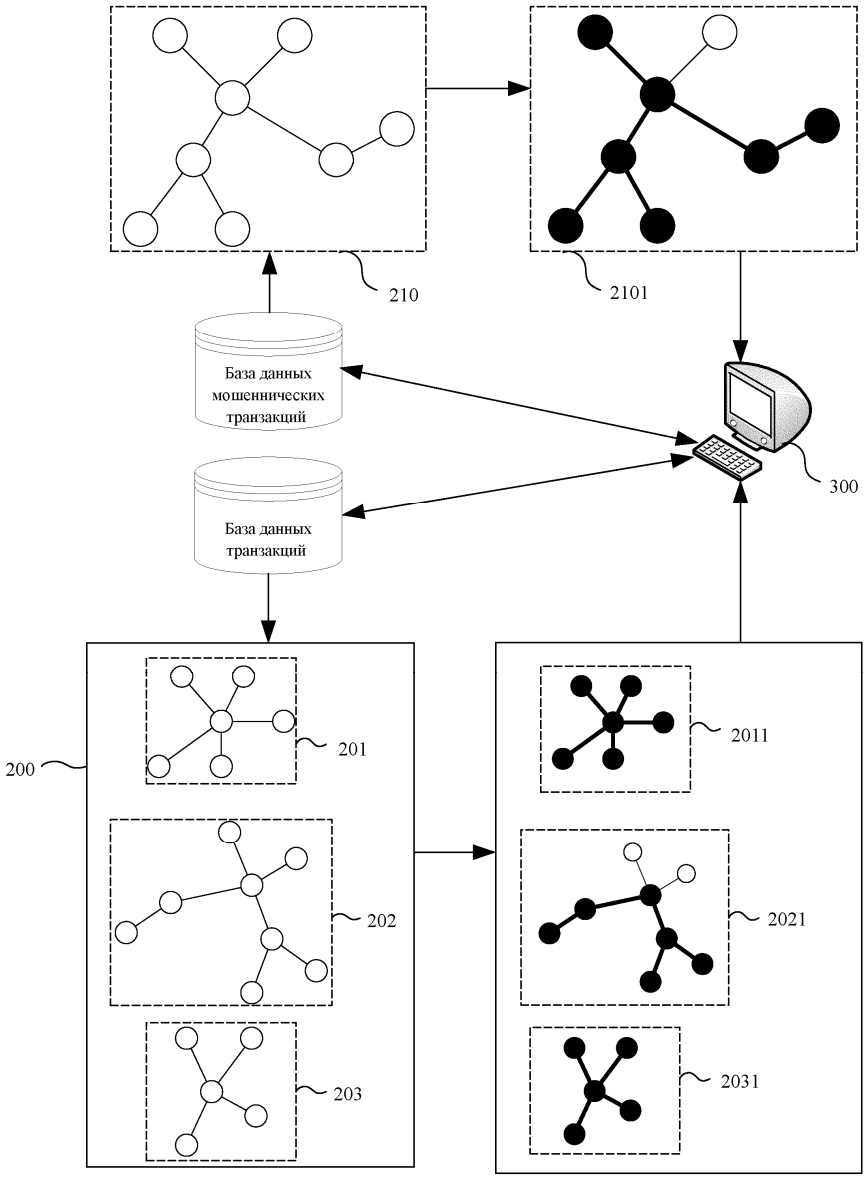
2021

DG3=1
kGK3=5
dV3 = |VK|/|V| = 6/6 =1

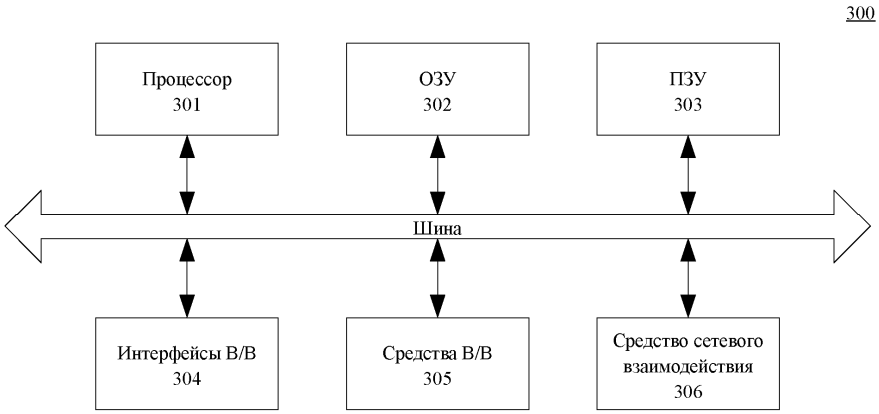


2031

Фиг. 3


210
2101
База данных мошеннических транзакций
300
База данных транзакций
200
201
2011
202
2021
203
2031

Фиг. 4


300
Процессор 301
ОЗУ 302
ПЗУ 303
Шина
Интерфейсы В/В 304
Средства В/В 305
Средство сетевого взаимодействия 306

Фиг. 5