

(19)



**Евразийское
патентное
ведомство**

(11) **040193**(13) **B1**(12) **ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОМУ ПАТЕНТУ**

(45) Дата публикации и выдачи патента
2022.04.28

(21) Номер заявки
201991423

(22) Дата подачи заявки
2019.07.10

(51) Int. Cl. **G06F 16/00** (2006.01)
H04W 64/00 (2006.01)
G06F 13/14 (2006.01)

(54) **СПОСОБ И СИСТЕМА ПОИСКА ПРИНАДЛЕЖНОСТИ IP-АДРЕСА
ТЕРРИТОРИАЛЬНОМУ КЛАСТЕРУ НА ОСНОВЕ ДАННЫХ ТРАНЗАКЦИЙ**

(31) **2019118544**

(32) **2019.06.14**

(33) **RU**

(43) **2020.12.30**

(56) **RU-C2-2611971**
US-B1-9641604
US-A1-20130246342
WO-A2-2008070415
WO-A1-2013192443

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

(72) Изобретатель:
**Воробьев Иван Александрович,
Пинчук Андрей Михайлович (RU)**

(74) Представитель:
Герасин Б.В. (RU)

(57) Изобретение относится к области обработки данных, в частности к способу и системе поиска принадлежности IP-адресов территориальным кластерам на основе транзакционных данных. Технический результат заключается в обеспечении определения местоположения участников транзакции с помощью используемых IP-адресов при их отнесении к заданному территориальному кластеру. Заявленный результат достигается за счет компьютерно-реализуемого способа поиска принадлежности IP-адреса территориальному кластеру на основе данных транзакций, который выполняется с помощью процессора и содержит этапы, на которых получают данные о транзакциях пользователей, содержащие IP-адреса отправителей и получателей транзакций; выполняют ранжирование частоты использования IP-адресов и количества транзакций между упомянутыми IP-адресами; формируют граф на основании данных ранжирования, в котором вершинами являются IP-адреса, а ребрами - количество переводов за определенный период времени между отправителями и получателями, использующими соответствующие IP-адреса; осуществляют построение территориальных кластеров на основании сформированного графа, причем каждый кластер содержит несколько IP-адресов; получают транзакционную информацию пользователя, содержащую, по меньшей мере, IP-адрес; осуществляют поиск полученного IP-адреса в территориальном кластере и в случае нахождения IP-адреса присваивают пользователю соответствующий территориальный кластер.

B1**040193****040193****B1**

Область техники

Изобретение относится, в общем, к области обработки данных, а в частности к способу и системе поиска принадлежности IP-адресов территориальным кластерам на основе транзакционных данных.

Уровень техники

Все большая доля транзакций проводится клиентами в удаленных каналах обслуживания, в которых единственным параметром, позволяющим определить фактическое расположение клиента, является IP-адрес. На текущий момент существуют справочники, которые сопоставляют конкретный IP-адрес с городом. Такие справочники не позволяют полноценно решать задачи, где необходимо понимать территориальное расположение клиента. Например, для случая противодействия мошенничеству, когда необходимо определение территориальной близости между двумя близкими по времени клиентскими сессиями в удаленном канале обслуживания, выполненными с различными IP-адресами. Основной причиной является то, что справочники устаревают, также IP-адреса могут переходить к другим провайдерам, принцип написания наименования населенных пунктов тоже не является формализованным в привязке к IP-адресам (один и тот же город зачастую имеет несколько написаний), точность для IP-адресов сотовых операторов крайне низкая.

Для целей анализа мошеннических действий при осуществлении транзакций существующие методы основываются на анализе удаленности пользователей от места выполнения транзакции, например, POS-терминалов, банкоматов и т.п. Для целей идентификации местонахождения участника транзакции может использоваться геолокация пользовательских устройств для выполнения транзакции, например, мобильных телефонов.

Такие решения, например, раскрываются в следующих патентных документах: US 20120215701 A1 (Playspan Inc., 23.08.2012), US 20190043054 A1 (Capital One Services LLC, 07.02.2019), US 20120209773 A1 (PayPal Inc., 16.08.2012).

Общим недостатком известных решений является необходимость использования уточняющих данных, характеризующих конкретные точки выполнения транзакций, в частности известные решения основываются исключительно на использовании геопозиции пользователя и средств выполнения транзакций для анализа точек предположительного местоположения, что не позволяет получить сегментированную информацию о территориальной принадлежности клиента тому или иному кластеру, соответствующему наиболее часто используемым IP-адресам, что позволит более быстро определить местоположение пользователя.

Раскрытие изобретения

Решаемой технической проблемой или технической задачей с помощью изобретения является формирование кластеров, указывающих территориальную принадлежность IP-адресов.

Технический результат заключается в обеспечении определения местоположения участников транзакции с помощью используемых IP-адресов при их отнесении к заданному территориальному кластеру.

Дополнительным техническим результатом является повышение скорости определения местоположения пользователя за счет сформированных кластеров, соответствующих территориальному распределению IP-адресов.

Заявленный результат достигается за счет компьютерно-реализуемого способа поиска принадлежности IP-адреса территориальному кластеру на основе данных транзакций, который выполняется с помощью процессора и содержит этапы, на которых

получают данные о транзакциях пользователей, содержащие IP-адреса отправителей и получателей транзакций;

выполняют ранжирование частоты использования IP-адресов и количества транзакций между упомянутыми IP-адресами;

формируют граф на основании данных ранжирования, в котором вершинами являются IP-адреса, а ребрами - количество переводов за определенный период времени между отправителями и получателями, использующими соответствующие IP-адреса;

осуществляют построение территориальных кластеров на основании сформированного графа, причем каждый кластер содержит несколько IP-адресов;

получают транзакционную информацию пользователя, содержащую по меньшей мере IP-адрес;

осуществляют поиск полученного IP-адреса в территориальном кластере и в случае нахождения IP-адреса присваивают пользователю соответствующий территориальный кластер.

В одном из частных вариантов осуществления способа данные о транзакциях получают от устройств выполнения транзакций.

В другом частном варианте осуществления способа устройства представляют собой мобильные устройства, банкоматы, POS-терминалы или их сочетания.

В другом частном варианте осуществления способа данные о транзакциях дополнительно содержат геокоординаты пользователя.

Изобретение также реализуется за счет компьютерной системы, содержащей по меньшей мере один процессор и память, хранящую исполняемые процессором инструкции для реализации вышеуказанного способа.

Краткое описание чертежей

Признаки и преимущества изобретения станут очевидными из приводимого ниже подробного описания изобретения и прилагаемых чертежей, на которых

фиг. 1 иллюстрирует блок-схему выполнения заявленного способа;

фиг. 2 - пример построения графа на основании IP-адресов;

фиг. 3, 4 - примеры сформированных кластеров;

фиг. 5 - пример вычислительного устройства.

Осуществление изобретения

На фиг. 1 представлен процесс выполнения заявленного способа (100) поиска принадлежности IP-адреса территориальному кластеру. На первом шаге (101) осуществляется сбор данных о транзакциях, которые могут включать в себя историю трат, историю входов в интернет-банк, устройства клиента, финансовое поведение клиента, остатки на счетах и т.д. Данные транзакций пользователя содержат IP-адреса пользователей в формате XXX.XXX.XXX.XXX, участвующих в переводах денежных средств, в частности р2р транзакции (англ. Payer-to-payer). Данные транзакций, как правило, поступают от устройств осуществления транзакций, например, банкоматов, POS-терминалов, мобильных устройств пользователей с установленным приложением для выполнения транзакций и т.п.

По собранным данным транзакций отбираются переводы между пользователями (р2р транзакции). Для каждого клиента определяется его IP-адрес класса С (XXX.XXX.XXX), с которого он осуществлял вход в канал выполнения транзакций, например мобильное приложение Сбербанк Онлайн или использование банкомата наибольшее число раз. Далее на этапе (102) в каждой транзакции клиент-отправитель платежа и клиент-получатель платежа сопоставляются с IP-адресом класса С, полученным ранее. Затем все транзакции агрегируются по уникальной паре IP-адрес отправителя и IP-адрес получателя, а также подсчитывается число транзакций по этой паре. Пример статистики представлен в табл. 1.

Таблица 1. Статистика использования IP-адресов

IP_адрес отправителя	IP_адрес получателя	Количество
83.220.239	213.87.148	975
213.87.130	31.173.84	902
62.182.73	62.182.73	185

На основе полученных данных по переводам между клиентами и их сессиям в удаленных каналах обслуживания (например, мобильное приложение, WEB-интерфейс) строится граф (этап 103), вершинами которого являются IP-адреса класса xxx.xxx.xxx, ребрами - количество переводов за определенный период времени между клиентами, использующими такие IP-адреса. Пример такого графа представлен на фиг. 2. Граф ненаправленный. В случае если между вершинами количество транзакций меньше заданного порога N (экспериментально N=40, если период 30 дней), то такое ребро исключается.

К построенному графу применяется алгоритм кластеризации (например, label propagation - https://en.wikipedia.org/wiki/Label_Propagation_Algorithm). Алгоритм кластеризации вершин графа позволяет отнести IP-адрес к определенному кластеру. Такая группировка позволяет выявлять IP-адреса, используемые клиентами, имеющими социальные и финансовые связи между собой (осуществление транзакций, обмен сообщениями посредством сети Интернет и т.п.). Алгоритм построения графа может быть реализован на основе алгоритма, раскрытого в статье "Community Detection via Semi-Synchronous Label Propagation Algorithms" (arXiv:1103.4450v1, 23.03.2011). Работа алгоритма основывается на ранжировании и отборе наибольших кластеров, их которых выбирается заданное число кластеров (например, 11 кластеров, что является оптимальных в ходе проведенных экспериментов), оставшиеся кластеры объединяются в один. На основании кластеризации графа формируются территориальные кластеры (104), которые позволяют на основе данных об использовании клиентами средств выполнения транзакций (банкоматов, POS-терминалов, мобильного приложения, посещения офисов банка и др.) определить фактическое местоположение клиента в определенный временной период. Результатом данного алгоритма является многоугольник, заданный геокоординатами на поверхности земли с плотностью распределения точек активности клиента.

Границы многоугольника могут определяться методом, схожим с картографированием ареала живых организмов (контурный метод), при котором на карту наносятся линейные границы ареала, с формированием контурной линии всей территории, в пределах которой можно встретить клиента, проявившего транзакционную активность. Дополнительно определяется центр многоугольника (медиана значений координат точек, входящих в многоугольник) и, в случае нахождения выброса в данных (координата, выделяющаяся от общей выборки по клиенту и определяемая "правилом трех сигм"), многоугольник перестраивается заново без точек-выбросов.

Данный подход позволяет каждого клиента, участвующего в кластеризации, отнести к определенному кластеру следующим способом:

найти его самый часто используемый IP-адрес класса С (XXX.XXX.XXX);

найти, к какому кластеру относится этот IP-адрес;

присвоить клиенту выявленный кластер.

Алгоритм сопоставления геопозиционирования клиента с IP-адресами клиентских сессий - используются для составления базы данных, с помощью которой можно определить территориальный кластер IP-адреса. Территориальный кластер представляет собой многоугольник, заданный геокоординатами на поверхности земли, построенный на основе данных клиентов, использующих IP-адрес с вычисленным кластером и по которым известны точные геокоординаты местонахождения (например, при посещении офиса банка) (105). При получении информации по текущим сессиям клиентов, осуществляющих транзакции, в частности данным использующихся IP-адресов при использовании устройств осуществления транзакций, выполняется сравнение полученного IP-адреса со сформированными территориальными кластерами, и при выявлении соответствующего IP-адреса в том или ином кластере клиенту присваивается территориальное месторасположение в рамках заданного кластера.

Определение наиболее часто используемого IP-адреса клиентом может выполняться с помощью анализа его транзакционной активности за заданный промежуток времени, например, день, неделя, месяц и т.п. Дополнительно в процессе кластеризации IP-адресов клиентов могут использоваться личные идентификаторы клиентов, например данные паспорта, идентифицирующие территориальную регистрацию клиентов. Примеры сформированных кластеров представлены на фиг. 3, 4. Для валидации полученных результатов работы алгоритма (100) было отобрано по 10000 клиентов из каждого кластера, по каждому клиенту была вычислена его средняя точка обитания (по использованию банкоматов и POS-терминалов). Точки нанесены на географическую карту и разбиты по сформированным кластерам.

На основании полученной информации по кластерам была оценена бизнес-метрика, в ходе которой были отобраны данные транзакций между клиентами, по которым поступала информация по подозрению в осуществлении мошеннических действий (фрод) от системы фрод-мониторинга. За заданный период времени было получено 86519 срабатываний. Уникальное количество клиентов, участвующих в этих транзакциях (отправитель и получатель) - 142847. По 118536 (83%) клиентам был сформирован кластер.

Если из транзакционных переводов исключить транзакции, по которым у отправителя и получателя нет кластера либо он "0" (объединение малых кластеров), то для оценки остается 50199 транзакций (58%). Если ввести признак "равенство кластера отправителя и получателя", то распределение: перевод в разрезе легитимных ("G") и мошеннических ("F") транзакций будет выглядеть, как представлено в табл. 2.

Таблица 2. Распределение легитимных и мошеннических транзакций

	Мошеннические транзакции	Легитимные транзакции
Количество	99.0000	50100.0000
Среднее	0.303030	0.552555
Стандартное отклонение	0.461907	0.497235
0%	0.000000	0.000000
10%	0.000000	0.000000
20%	0.000000	0.000000
30%	0.000000	0.000000
40%	0.000000	0.000000
50%	0.000000	1.000000
60%	0.000000	1.000000
70%	0.600000	1.000000
80%	1.000000	1.000000
90%	1.000000	1.000000
Максимальное значение	1.000000	1.000000

Созданный способ (100) кластеризации клиентов инструмент позволяет по истории входов в интернет-банк (IP-адреса) разделить клиентов по принадлежности к географической общности, не используя знания об их геокоординатах. На сегодняшний момент многие клиенты пользуются только мобильным приложением интернет-банка (узнают остатки денежных средств, осуществляют переводы и т.д.). Есть также клиенты, использующие и мобильное приложение и банкоматную сеть (внос/снятие наличных, подключение мобильного банка и т.д.). При использовании банкомата можно определить, в каком месте находился клиент, получая услугу. Составленный алгоритм позволяет использовать знания о поведении клиентов, использующих как банкоматы, так и мобильное приложение, при обогащении профилей клиентов, не оставляющих географический след. Таким образом, реализуется возможность применимости моделей, в которых есть необходимость понимания фактического месторасположения клиента, на тех,

кто пользуется только интернет-банком с отключенным модулем определения геолокации.

Результаты работы кластеризации могут использоваться в качестве дополнительного признака для повышения эффективности модели скоринга транзакции. При осуществлении легитимных транзакций клиенты чаще осуществляют переводы клиентам, которые находятся в одной географической местности с отправителем. При этом в мошеннических переводах чаще наблюдается паттерн поведения, в котором отправитель и получатель находятся на далеком расстоянии друг от друга: мошенникам не всегда известна локация жертвы и подобрать клиента получателя из того же региона, что и отправитель - затруднительно. При создании модели оценки перевода на возможное мошенничество признак одного географического местоположения отправителя и получателя оказывает сильное влияние на снижение ложных срабатываний антифрод системы.

На фиг. 5 представлен пример общего вида вычислительной системы (200) на базе вычислительного устройства (200), которое обеспечивает реализацию заявленного способа (100). Устройство (200) может являться частью компьютерной системы, например сервером, обрабатывающим необходимые данные для осуществления способа (100).

В общем случае, вычислительное устройство (200) содержит объединенные общей шиной информационного обмена один или несколько процессоров (201), средства памяти, такие как ОЗУ (202) и ПЗУ (203), интерфейсы ввода/вывода (204), устройства ввода/вывода (205), и устройство для сетевого взаимодействия (206).

Процессор (201) (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 (202) представляет собой оперативную память и предназначено для хранения исполняемых процессором (201) машиночитаемых инструкций для выполнения необходимых операций по логической обработке данных. ОЗУ (202), как правило, содержит исполняемые инструкции операционной системы и соответствующих программных компонент (приложения, программные модули и т.п.).

ПЗУ (203) 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 (200) и организации работы внешних подключаемых устройств применяются различные виды интерфейсов В/В (204). 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 (200) применяются различные средства (205) 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 (206) обеспечивает передачу данных устройством (200) посредством внутренней или внешней вычислительной сети, например Интранет, Интернет, ЛВС и т.п. В качестве одного или более средств (206) может использоваться, но не ограничиваясь: 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 (200), например GPS, ГЛОНАСС, BeiDou, Galileo.

Представленные материалы заявки раскрывают предпочтительные примеры реализации изобретения 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 принадлежности IP-адреса территориальному кластеру на основе данных транзакций, выполняемый с помощью процессора и содержащий этапы, на которых получают данные о транзакциях пользователей, содержащие IP-адреса отправителей и получателей транзакций;

выполняют ранжирование частоты использования IP-адресов и количества транзакций между упомянутыми IP-адресами;

формируют граф на основании данных ранжирования, в котором вершинами являются IP-адреса, а ребрами - количество транзакций за определенный период времени между отправителями и получателями, использующими соответствующие IP-адреса;

осуществляют построение территориальных кластеров на основании сформированного графа, причем каждый кластер содержит несколько IP-адресов;

получают транзакционную информацию пользователя, содержащую по меньшей мере IP-адрес;

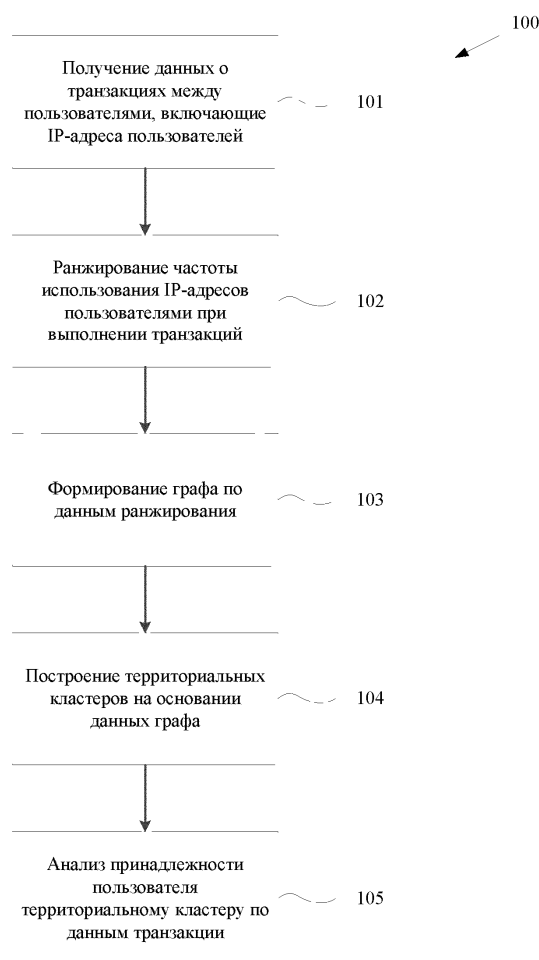
осуществляют поиск полученного IP-адреса в территориальном кластере и в случае нахождения IP-адреса присваивают пользователю соответствующий территориальный кластер.

2. Способ по п.1, характеризующийся тем, что данные о транзакциях получают от устройств выполнения транзакций.

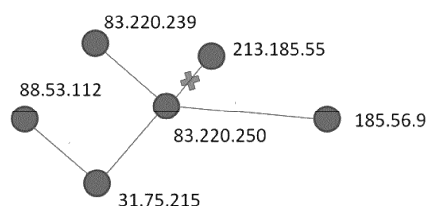
3. Способ по п.2, характеризующийся тем, что устройства представляют собой мобильные устройства, банкоматы, POS-терминалы или их сочетания.

4. Способ по п.2, характеризующийся тем, что данные о транзакциях дополнительно содержат геокоординаты пользователя.

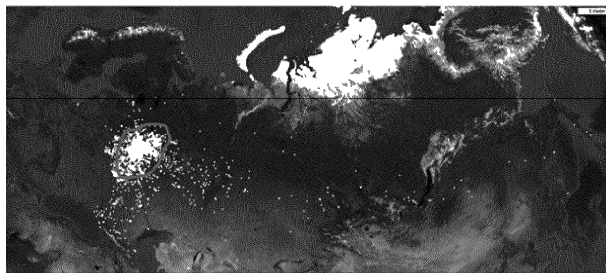
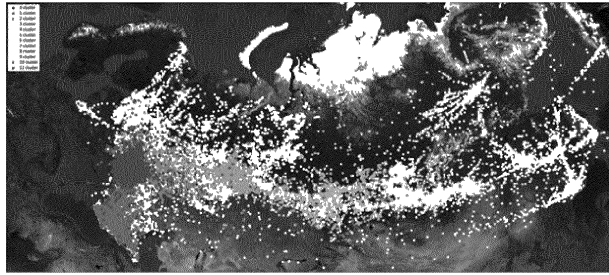
5. Система поиска принадлежности IP-адреса территориальному кластеру на основе данных транзакций, содержащая по меньшей мере один процессор и память, содержащую машиночитаемые инструкции, которые при их выполнении с помощью процессора реализуют способ по пп.1-4.



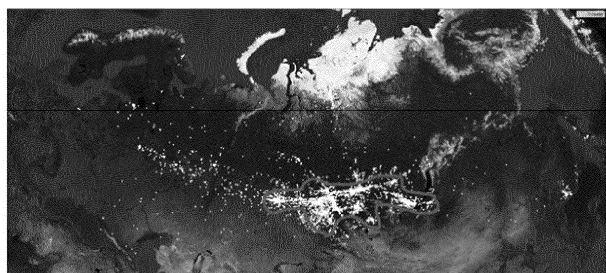
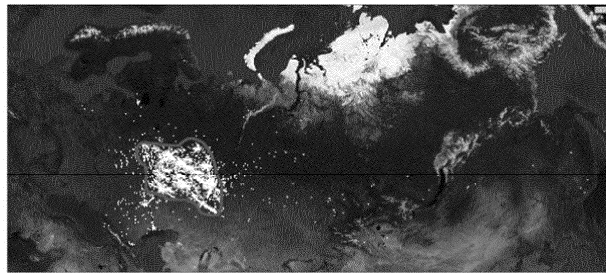
Фиг. 1



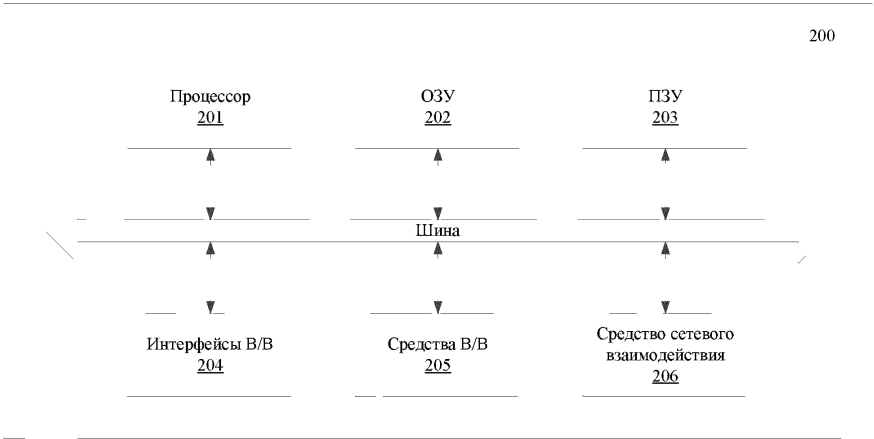
Фиг. 2



Фиг. 3



Фиг. 4



Фиг. 5

