(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
20 October 2016 (20.10.2016)

WIPO | PCT

(10) International Publication Number
WO 2016/167544 A1

(51) International Patent Classification:
G06Q 20/40 (2012.01) *G06Q 20/38* (2012.01)

(21) International Application Number:
PCT/KR2016/003844

(22) International Filing Date:
12 April 2016 (12.04.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/147,012 14 April 2015 (14.04.2015) US
14/962,365 8 December 2015 (08.12.2015) US
10-2016-0044908 12 April 2016 (12.04.2016) KR

(71) Applicant: **SAMSUNG ELECTRONICS CO., LTD.** [KR/KR]; 129, Samsung-ro, Yeongtong-gu, Suwon-si, Gyeonggi-do 16677 (KR).

(72) Inventors: **PATEL, Kunal M.**; 1000 Escalon Avenue, Apt. D3027, Sunnyvale, California 94085 (US). **KANG, Abraham J.**; 23581, Summit Road, Los Gatos, California 95033 (US). **KASMAN, Bulent**; 5114, Watkings Way, Antioch, California 94531 (US). **NING, Peng**; 20070, Edinburgh Drive, Saratoga, California 95070 (US). **GRACE, Michael C.**; 1003, Fuller Terrace, Sunnyvale, California 94086 (US).

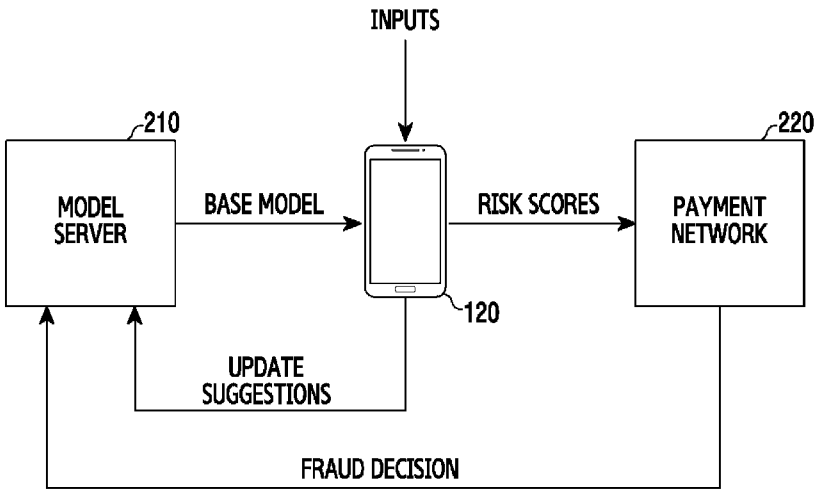
(74) Agents: **KWON, Hyuk-Rok** et al.; 2F, 28, Gyeonghuigung-gil, Jongro-gu, Seoul 03175 (KR).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report (Art. 21(3))*

(54) Title: APPARATUS AND METHOD FOR FRAUD DETECTION IN A MOBILE DEVICE



(57) Abstract: A user device comprises a processor and a communication unit. The processor is configured to determine a risk metric indicating a fraudulent activity associated with a payment, according to a user input associated with the payment, based on confidential information associated with the user. The communication unit is configured to transmit to a payment server information related to the payment and information related to the risk metric that is associated with the payment.

WO 2016/167544 A1

Description

Title of Invention: APPARATUS AND METHOD FOR FRAUD DETECTION IN A MOBILE DEVICE

Technical Field

[1] The present application relates generally to mobile communication devices and, more specifically, to an apparatus and method of fraud detection in a mobile device to preserve user privacy.

Background Art

[2] Mobile devices, such as smartphone and tablets, typically include one or more mobile payment applications, such as Samsung Pay and the like, that allow a user to register one or more credit cards or debit cards on his or her mobile device in order to make payments from the credit card account through the mobile device. Payment systems dealing with credit card or debit card information historically have had to deal with fraudulent card use, whether through the malicious theft of information or through the malicious usage of a stolen credit/debit card (fraudulent payment card use).

[3] Detecting fraudulent payment card use typically requires access to a large amount of both personally identifiable information of the cardholder and private information of the cardholder. However, it could potentially be damaging to the user if the private information of the user becomes associated with personally identifiable information of the user. As a result, conventional fraud detection systems seek to balance the accuracy of fraud detection with the need to collect personally identifiable information and/or private information.

[4] Therefore, there is a need in the art for improved apparatuses and methods for preventing fraud in mobile payment systems that use credit and/or debit cards. In particular, there is a need for smartphone payment systems and applications that minimize the exposure to payment systems of personally identifiable information of the cardholder and private information of the cardholder.

Disclosure of Invention

Technical Problem

[5] An embodiment of the present disclosure provides an apparatus and method of fraud detection in a payment system of mobile communication device.

Solution to Problem

[6] According to an embodiment, it is a primary object to provide a user device comprising: i) transmit path circuitry and receive path circuitry configured to communicate with a payment server; and ii) processing circuitry configured to control the transmit path circuitry and receive path circuitry and further configured to: a) receive a

user input related to a payment process; b) calculate a risk score indicative of a likelihood of fraudulent activity associated with the payment process, wherein the risk score calculation is based on confidential information associated with the user that is stored on the user device; and c) transmit to the payment server a payment action and the risk score associated with the payment action without disclosing the confidential information.

[7] In one embodiment, the confidential information comprises personally identifiable information of the user.

[8] In another embodiment, the confidential information comprises private information of the user.

[9] In still another embodiment, the processing circuitry calculates the risk score using a risk base model received from a model server.

[10] In yet another embodiment, the processing circuitry is further configured to transmit to the model server a suggested parameter update usable by the model server to improve the accuracy of the risk base model.

[11] In a further embodiment, the risk base model is based on a neural network.

[12] In a yet further embodiment, the risk base model is based on a decision tree.

[13] In a still further embodiment, the processing circuitry is further configured to transmit to the payment server a justification corresponding to each risk score.

[14] It is another object to provide a method for use in a user device, the method comprising: i) communicating with a payment server; ii) receiving a user input related to a payment process; iii) calculating a risk score indicative of a likelihood of fraudulent activity associated with the payment process, wherein the risk score calculation is based on confidential information associated with the user that is stored on the user device; and iv) transmitting to the payment server a payment action and the risk score associated with the payment action without disclosing the confidential information.

[15] It is further object to provide a non-transitory computer readable medium configured to control a processor to perform a method of processing payments, the method comprising: i) communicating with a payment server; ii) receiving a user input related to a payment process; iii) calculating a risk score indicative of a likelihood of fraudulent activity associated with the payment process, wherein the risk score calculation is based on confidential information associated with the user that is stored on the user device; and iv) transmitting to the payment server a payment action and the risk score associated with the payment action without disclosing the confidential information.

[16] In other embodiment, a user device domprises a processor configured to determine a risk metric indicating a fraudulent activity associated with a payment, according to a

user input associated with the payment, based on confidential information associated with the user, and a communication unit configured to transmit to a payment server information related to the payment and information related to the risk metric that is associated with the payment.

[17] In other embodiment, a method of operating a user device comprises determining a risk metric indicating a fraudulent activity associated with a payment, according to a user input associated with the payment, based on confidential information associated with the user, and transmitting to a payment server information related to the payment and information related to the risk metric that is associated with the payment.

Advantageous Effects of Invention

[18] According to an embodiment of the present disclosure, an apparatus and method of the present disclosure may effectively preserve confidential information related to a user of a mobile device and allow to a secure payment using the mobile device.

Brief Description of Drawings

[19] For a more complete understanding of the present disclosure and its advantages, reference is now made to the following description taken in conjunction with the accompanying drawings, in which like reference numerals represent like parts:

[20] FIGURE 1 illustrates a mobile payment system according to the principles of the disclosure.

[21] FIGURE 2 illustrates a method of updating parameters of a risk model according to an embodiment of the disclosure.

[22] FIGURE 3 illustrates in greater detail exemplary of a user device that calculates fraud detection risk scores according to an embodiment of the disclosure.

[23] FIGURE 4 illustrates one of a disjunction of step functions according to an exemplary embodiment of the disclosure.

[24] FIGURE 5 illustrates updating thresholds for one of a disjunction of step functions according to another exemplary embodiment of the disclosure.

[25] FIGURE 6 illustrates an operating flowchart of a user device according to an exemplary embodiment of the disclosure.

Best Mode for Carrying out the Invention

[26] Before undertaking the DETAILED DESCRIPTION below, it may be advantageous to set forth definitions of certain words and phrases used throughout this patent document: the terms "include" and "comprise," as well as derivatives thereof, mean inclusion without limitation; the term "or," is inclusive, meaning and/or; the phrases "associated with" and "associated therewith," as well as derivatives thereof, may mean to include, be included within, interconnect with, contain, be contained within, connect to or with, couple to or with, be communicable with, cooperate with, interleave,

juxtapose, be proximate to, be bound to or with, have, have a property of, or the like. Definitions for certain words and phrases are provided throughout this patent document, those of ordinary skill in the art should understand that in many, if not most instances, such definitions apply to prior, as well as future uses of such defined words and phrases.

[27] FIGURES 1 through 6, discussed below, and the various embodiments used to describe the principles of the present disclosure in this patent document are by way of illustration only and should not be construed in any way to limit the scope of the disclosure. Those skilled in the art will understand that the principles of the present disclosure may be implemented in any suitably arranged mobile device.

[28] FIGURE 1 illustrates mobile payment system 100 according to an embodiment of the present disclosure. Mobile payment system 100 comprises user device 120, cloud server 130, interface server 140, and payment server 150, each of which is configure to communicate with one or more other devices through Internet protocol (IP) network 110, such as Internet 110. In the exemplary embodiment, user device 120 is assumed to be mobile phone 120 as an electronic device. However, alternatively, user device 120 may be any one of a tablet, a smart television, a laptop personal computer, or another type of consumer information appliance.

[29] A payment application on user device 120 may communicate with cloud server 130 to receive risk-related information and updates. The payment application on user device 120 also communicates with interface server 140 and payment server 150 to make on-line payments based on credit cards and/or debits cards. In here, it may mean that user device 120 communicates with cloud server 130, interface server 140, and payment server 150 using the payment application. In other words, user device 120 may communicate with one or more server, such as cloud server 130, interface server 140, and payment server 150 to make on-line payments based on credit cards and/or debits cards. If user device 120 is an Android OS-based system, the payment application may be, for example, Samsung Pay.

[30] In a conventional on-line payment system, one or more interface servers 140 would be responsible for collecting from user device 120 the confidential information (e.g., personally identifiable information, private information) of the user/cardholder in order to calculate a risk score that could be used to detect fraud. In here, the risk score is a value for determining a level of the risk. Thus, the risk score may be referred as 'risk metric', 'value related to the risk', and the like. In here, the confidential information is defined based on contents that are included in information. For example, the confidential information may be information which is representing personal data of the user/cardholder. In other words, the confidential information is unique information of the user/cardholder which allows differentiating between a user of user device 120 and

unspecified third parties. The confidential information includes the unique information of the user/cardholder, thus invasion of personal privacy and the fraudulent activity may arise when the confidential information is revealed. For example, the confidential information may include personally identifiable information, private information, and the like. Interface server 140 would typically report the risk score to payment server 150, which may reject the payment request from user device 120 if the risk score is too high. According to the principles of the disclosure, user device 120 may minimize the exposure of the confidential information (e.g., personally identifiable information, private information) of the cardholder and improve fraud detection accuracy without revealing the personally identifiable or private information. The personally identifiable and private information may include, for example, the name and address of the user, the workplace of the user, the social security number of the user, the location of the user, the recent purchases of the user, movement behavior of the user, usage patterns of certain applications, frequency of changes to telephony information (e.g., phone numbers), and the like.

[31] Accordingly, the personally identifiable and private information remains on user device 120 where that information originated. The payment application on user device 120 is configured to assess risk and detect fraud without revealing the personally identifiable and private information to third parties or to external servers. In one embodiment, risk scores are calculated on user device 120, but the algorithm and algorithm parameters used to calculate risk scores may be controlled by one or more interface servers 140.

[32] User device 120 determines the risk of fraudulent activity using the personally identifiable and private information that user device 120 collects. User device 120 reveals a risk score along with a payment action (e.g., payment request) to interface servers 140 and payment servers 150. In other words, user device 120 may transmit information associated with the risk score to interface servers 140 and payment servers 150. User device 120 optionally provides selected information to central server (i.e., cloud server 130) that allows server 130 to improve the risk calculation algorithm used by user devices 120. However, user device 120 does not reveal the personally identifiable or private information of the user used to improve the risk calculation algorithm to cloud server 130. In other words, the personally identifiable or private information of the user does not transmitted to external servers. In other words, the selected information that is transmitted to servers may not include the personally identifiable or private information.

[33] In an advantageous embodiment, the exemplary outputs of user device 120 comprise: i) one or more risk scores; ii) a set of privacy-preserving justifications corresponding to each risk score; and iii) parameter updates. In the present disclosure, justifications may

indicate filters or at least one set of filter. In other words, the privacy-preserving justifications may be defined as means for converting a monitored activity to information for revealing a risk. For example, the privacy-preserving justifications mean logical functions, modules, algorithms and the like used for converting an activity to the risk score.

[34] Risk Scores - Each output risk score may be used individually as an indicator of various factors expected to be useful for fraud detection. For example, one type of risk score may reflect the likelihood (e.g., a percentage probability) that the owner of user device 120 is not the owner of a card being registered. Another type risk score may reflect the likelihood that the current user of user device 120 is not the legitimate owner of user device 120. The other type of risk score may convey the likelihood that the user is attempting to hide his or her location while using the payment application.

[35] Privacy-preserving justifications for risk scores - Each output risk score may be the outcome of a plurality of highly-indicative justifications that may be easy to understand by human third parties. For example, if an output risk score indicates that the user is likely trying to hide his or her location while using the payment application, at least part of the justification may be that the user is connecting the payment application to payment server 150 or interface server 140 through an internet public proxy.

[36] Parameter updates - Each risk score may be computed based on a given set of randomly distributed parameters. External parties cannot use the parameter updates for the purpose of guessing the private information for a specific user device 120 or a specific user. This ensures that parameter updates alone cannot be used to leak sensitive information about individual users. However, parameter updates in the aggregate across many user devices 120 may be used to identify indicators of fraud. Details of how parameters and parameter updates are calculated are described below.

[37] Due to the privacy-preserving nature of this system, one embodiment of the method for computing the risk score may have the following constraints. First, the parameters involved with computing the risk score must be separable into a vector of private information and a vector of risk parameters. The risk parameters must be independent of private information to ensure that risk parameter updates may be revealed without leaking private information. Second, because the risk score calculation should be updatable as the system becomes more robust, it must be possible for the risk parameters to be updatable. Third, it must be possible for risk parameter update values to be aggregated and for the aggregate risk parameter updates to be usefully applied to any existing vector of risk parameters. Fourth, in order to provide justifications for output risk scores, it must be possible to determine the relative contributions of each component of the vector of private information to the final risk scores. Finally, in order to hide the correlation between the vector of private information and parameter

updates, there must always be multiple possible associated private information vectors for any given parameter update. For example, a relation between any given parameter update and private information vectors is one-to-multiple (1-to-N) or multiple-to-multiple (N-to-N), and thus the correlation between the vector of private information and parameter updates may be hidden.

[38] FIGURE 2 illustrates a method of updating parameters of a risk base model according to an embodiment of the present disclosure. Model server 210 (e.g., cloud server 130) provides a risk base model for determining risk to user device 120. In user device 120, the risk base model may include a basic tool for calculating for the risk score. For example, the risk base model may be implemented as a formula, a function, rules, algorithms, or a combination thereof to calculate the risk score. User device 120 receives from the user multiple inputs, including personally identifiable and private information of the user. In the present disclosure, since the user multiple inputs are directly input to user device 120, the reception of the user multiple inputs by user device 120 may mean that user device 120 identifies or detects the multiple inputs. Using the risk base model and the inputs, user device 120 calculates one or more risk scores, which are transmitted to payment network 220 (e.g., interface server(s) 140 and payment server(s) 150) as part of a payment action. This may include, for example, transmitting the risk score(s) as part of card enrollment data. From time-to-time, user device 120 also transmits parameter update suggestions to model server 210 to improve the accuracy of the risk calculation base model. The risk calculation base model may include a basic tool for calculating the risk.

[39] FIGURE 3 illustrates in greater detail exemplary user device 120 that calculates fraud detection risk scores according to an embodiment of the disclosure. In FIGURE 3, user device 120 may be implemented as mobile phone 120. However, this is by way of illustration only and should not be construed to limit the scope of the disclosure. In alternate embodiments, user device 120 may instead be a tablet, a smart TV, a laptop computer, or any other information appliance.

[40] User device 120 comprises core circuitry 300, which includes read-only memory (ROM) 305, random access memory (RAM) 310, central processing unit (CPU) 315, digital signal processor (DSP) 320, digital-to-analog converter (DAC)/analog-to-digital converter (ADC) circuitry 325, baseband (BB) circuitry block 330, codec circuitry block 335, radio frequency (RF) circuitry block 340, transmit (TX)/receive (RX) switch 345, and antenna 395.

[41] In one embodiment, ROM 305 may store a boot-routine and other static data and RAM 310 may store an operating system (not shown), applications 312, and protocol stack 314. In an advantageous embodiment, ROM 305 and RAM 310 may comprise a single electronically erasable memory, such as a Flash memory, that is used in con-

junction with a conventional RAM memory that is used to store dynamic data. Applications 312 in memory may include a social presence application (i.e., RCS Presence) that interacts with carrier SP server 150, an IP multimedia subsystem (IMS) framework that delivers IP multimedia services, a Calendar application that communicates with calendar server 160, and specific Social Network Site (SNS) applications (e.g., Facebook, Twitter), and the like that enable user device 120 to exchange SP information with mobile phones used by other subscribers.

[42] User device 120 further comprises SIM card interface 350, USB interface 355, GPS receiver 360, Bluetooth (BT) transceiver 365, WiFi (or WLAN) transceiver 370, speaker and microphone circuitry block 375, keyboard 380, display 385, and camera 390. In some embodiment, keyboard 380 and display 385 may be implemented together as a touch screen display.

[43] CPU 315 is responsible for the overall operation of user device 120. In an exemplary embodiment, CPU 315 executes applications 312 and protocol stack 314. CPU 315 runs the application layer and a wide variety of applications may be run in a smart phone implementation. Applications 312 may include audio, video, browser, and image/graphics applications. CPU 315 may run applications 312 that support various audio formats such as MP3, MP4, WAV, and rm. CPU 315 may run image applications 312 that support JPEG image formats and video applications 312 that support video formats (e.g., MPEG-1 to MPEG-5). CPU 315 may support various operating systems (not shown), such as Symbian, java, android, RT-Linux, Palm, and the like. For time critical applications, CPU 315 runs a real-time operating system (RTOS). In addition to the physical layer, there are other layers, including protocol stack 314, that enable user device 120 to work with a network base station. In an exemplary embodiment, protocol stack 314 is ported on CPU 315.

[44] DAC/ADC circuitry block 325 converts analog speech signals to digital signals, and vice versa, in user device 120. In the transmit path, the ADC-converted digital signal is sent to a speech coder. Various types of ADCs are available, including sigma delta type. Automatic gain control (AGC) and automatic frequency control (AFC) are used in the receive path to control gain and frequency. AGC helps maintain satisfactory DAC performance by keepings signals within the dynamic range of the DAC circuits. AFC keeps frequency error within limit to achieve better receiver performance.

[45] Baseband (BB) circuitry block 330 may be implemented as part of DSP 320, which executes many of the baseband processing functions (i.e., physical layer, Layer 1, or L1 functions). BB circuitry block 330 may be ported on DSP 320 to meet the latency and power requirements of user device 120. BB circuitry block 330 converts voice and data to be carried over the air interface to I/Q baseband signals.

[46] BB circuitry block 330 may change from modem to modem for various air interface

standards, such as GSM, CDMA, Wimax, LTE, HSPA, and others. BB circuitry block 330 is often referred to as the physical layer, or Layer 1, or L1. For mobile phones that work on GSM networks, the baseband part (Layer 1) running on DSP 320 and the protocol stack 314 running on CPU 315 are based on the GSM standard. For CDMA mobile phones, the Layer 1 and protocol stack 314 are based on the CDMA standard, and so on, for the LTE and HSPA standards-based mobile phones.

[47] For speech or audio inputs, codec circuitry block 335 may compress and decompress the signal to match the data rate to the frame in which the data is sent. By way of example, codec circuitry block 335 may convert speech at an 8 KHz sampling rate to a 13 kbps rate for a full rate speech traffic channel. To do this, a residually excited linear predictive coder (RELP) speech coder may be which compresses 260 bits into a 20 millisecond duration to achieve a 13 kbps rate.

[48] The baseband or physical layer adds redundant bits to enable error detection as well as error correction. Error detection may be obtained with CRC and error correction using forward error correction techniques, such as a convolutional encoder (used in transmitter path) and a Viterbi decoder (used in receive path). Interleaving may be done for the data, which helps in spreading the error over time, thereby helping the receiver de-interleave and decode the frame correctly.

[49] RF circuitry block 340 includes an RF up-converter and an RF down-converter. For a GSM system, the RF up-converter converts modulated baseband signals (I and Q) either at zero intermediate frequency (IF) or some IF to RF frequency (890-915 MHz). The RF down-converter converts RF signals (935 to 960 MHz) to baseband signals (I and Q). For a GSM system, GMSK modulation is used.

[50] Antenna 395 is a metallic object that converts and electro-magnetic signal to and electric signal and vice versa. Commonly used antennas may include a helix type, a planar inverted F-type, a whip, or a patch type. Microstrip patch type antennas are popular among mobile phones due to small size, easy integration on a printed circuit board and multi-frequency band of operation. In a preferred embodiment of user device 120, antenna 395 may support different wire-area standards, including GSM, CDMA, LTE, and WiMAX, as well as short-range standards, including WiFi (WLAN), Bluetooth, and so on.

[51] If antenna 395 comprises only one antenna used for both transmit and receive operations at different times, the TX/RX switch 345 couples both the transmit (TX) path and the receive (RX) path to antenna 395 at different times. TX/RX switch 345 is controlled automatically by DSP 320 based on a GSM frame structure with respect to the physical slot allocated for that particular GSM mobile phone in both the downlink and the uplink. For frequency division duplexing (FDD) systems, TX/RX switch 345 may be implemented as a diplexer that acts as filter to separate various frequency

bands. User device 120 provides connectivity with laptops or other devices using WiFi (or WLAN) transceiver 370, BT transceiver 365, and universal serial bus (USB) interface 355. User device 120 also uses GPS receiver 360 in applications 312 that require position information. If User device 120 is a conventional smart phone, applications 312 may include many popular applications, such as Facebook, Twitter, a browser, and numerous games that come pre-installed with user device 120.

[52] DAC/ADC circuitry block 325, BB circuitry block 330, RF circuitry block 340, TX/RX switch 345, and antenna 395 is components for transmitting and/or receiving signals via wireless channel. Thus, DAC/ADC circuitry block 325, BB circuitry block 330, RF circuitry block 340, TX/RX switch 345, and antenna 395 may be referred as communication unit, transceiver and the like.

[53] Speaker and microphone circuitry block 375 comprises microphone circuitry (or mic) that converts acoustic energy (i.e., air pressure changes caused by speech or other sounds) to electrical signals for subsequent processing. Speaker and microphone 375 further comprises speaker circuitry that converts an electrical audio signal to an audible signal (pressure changes) for human hearing. The speaker circuitry may include an audio amplifier to get required amplification of the audio signal and may further include a volume control circuit to change (increase or decrease) the amplitude of the audio signal. User device 120 preferably includes camera 390. Presently, almost all mobile phones feature a camera module. Camera 390 may comprise a 12 megapixel, 14 megapixel, or a 41 megapixel camera. Display 385 may comprise, by way of example, a liquid crystal display (LCD), a thin-film transistor (TFT) screen, and organic light emitting diode (OLED) display, a thin film diode (TFD) display, or a touch screen of capacitive and resistive type.

[54] In a simple embodiment, keypad 380 may comprise a simple matrix type keypad that contains numeric digits (0 to 9), alphabetic characters (A to Z), special characters, and specific function keys. In a more advanced embodiment for a smart phone implementation, keypad 380 may be implemented in the mobile phone software, so that keyboard 380 appears on display 385 and is operated by the user using the touch of a finger tip.

[55] Model Overview - The underlying model abstraction that is used is as follows:

[56] $$risk = f(input; parameters), \qquad \text{[Eqn. 1]}$$

[57] where the value "risk" represents the vector of output risk scores, the variable "parameters" represents the vector of model parameters that - when combined with function "f" - completely define the risk calculation, and the variable "input" represents the vector of both input private information and input non-private information. In the case of a linear model, the risk value may be a vector of scalar values ranging from - ∞

to + ∞, the parameters variable may contain the coefficients of each input variable, and the input variable may be a vector of probabilities.

[58] Example Model - Disjunction of Step Functions - FIGURE 4 illustrates a risk model based on one of a disjunction of step functions according to an exemplary embodiment. In this exemplary model, the risk function f is the maximum of a set of step functions parameterized by thresholds. Each step function is treated independently. In this scenario, user device 120 compares an input (e.g., the number of times an account has been registered in the last week) to the threshold value. In here, when the number of times an account has been registered is to be used as the example, the number of times registration about a same account is counted, or the number of times registration about accounts which are different from each other is only counted. If the input value is greater than the threshold, the action is considered a high risk. Otherwise, the action is considered a low risk.

[59] Example Model - Neural Network - In this exemplary risk calculation model, the risk function f is a neural network parameterized by weights. As an example, the inputs are all converted to numerical values and then converted to a risk score as follows:

[60] $$hidden_1 = sigmoid(\sum input_1 * weight_{1i}) \quad \text{[Eqn. 2]}$$

[61] $$hidden_2 = sigmoid(\sum hidden_{1i} * weight_{2i}) \quad \text{[Eqn. 3]}$$

[62] $$risk = sigmoid(\sum hidden_{2i} * weight_{3i}) \quad \text{[Eqn. 4]}$$

[63] If the output risk score is close to one, the action is considered to be high risk. Otherwise, the action is considered to be low risk.

[64] Example Model - Decision Tree - In this exemplary risk calculation model, the risk function f is a decision tree parameterized by: i) the hierarchy of decisions; ii) the parameters required for each individual decision; and iii) the risk score output by each leaf of the decision tree. In this risk model, user device 120 computes a series of Boolean values based on the input. For example, the first decision may be to determine whether or not user device 120 has been used to register more than some threshold number of accounts. If the result is true, then the second decision may be to determine whether or not a card exists on user device 120. Otherwise, the second decision may be to determine whether or not the cardholder name matches the name associated with the new account. At least one decision process that is described above continues until a decision outputs whether or not an action is a high risk.

[65] Exemplary Model Inputs - According to the principles of the disclosure, user device 120 may use certain observable actions or conditions as indications of: i) the user attempting to obfuscate identity of the user; ii) the use of stolen identities; or iii) unusual activity. Attempting to obfuscate the identity of the user may mean causing a situation where the identity of the user is recognized as the third party. In another case, at-

tempting to obfuscate the identity of the user may mean causing a situation where information about the identity of the user is not recognized by uncertainly processing the identity of the user. These observable actions or conditions may include: a) the use of an Internet proxy server. In this case, the actual information being hidden by the proxy may be considered private information and may be used an input to the risk calculation model; b) attempts to register many accounts with different information; c) the existence of many different cardholder names. The cardholder names and relationship implied may be considered private information and may be used as an input to the risk calculation model; d) earlier reports of fraudulent activity associated with user device 120; and e) repeatedly resetting or updating payment-related or account-related data.

[66] Selecting Model Parameters - According to the principles of the disclosure, user device 120 makes use of device-generated parameter updates. To maximize the usefulness of such parameter updates, these parameter updates are based on the aggregate of many device parameters that result in some known statistic. To accomplish maximizing the usefulness of the parameter updates, model server 210 (e.g., cloud server 130) may generate base model parameters (see FIGURE 2). The base model parameters may represent a basic tool for configuring to model parameters. These parameters may be randomly generated or manually selected by a system operator. In order to hide the correlation between the private input vector and the public output risk vector, user device 120 may perturb the model parameters based on a distribution passed down from model server 210, such as:

[67] $$random_{device} - random(;distribution_{server}) \quad \text{[Eqn. 5]}$$

[68] $$parameters_{device} = parameters_{server} + random_{device} \quad \text{[Eqn. 6]}$$

[69] $$risk = f(input;parameters_{device}) \quad \text{[Eqn. 7]}$$

[70] Model server 210 ensures that when the random distribution that model server 210 generates is combined with the parameters model server 210 generates, the random distribution will have the statistics necessary for parameter updates to be useful. For example, model server 210 may require the mean value of a given user device 120 parameter to be equal to a selected value. Thus, model server 210 may send a server parameter equal to the selected value along with a random distribution with a mean value of zero. In cases where private information will not be leaked regardless of the contents of the parameter updates, model server 210 may send a distribution that takes a constant value. For example, the distribution may always take the value of zero.

[71] Example Parameter Selection - Disjunction of Step Functions - As mentioned above, the threshold values are considered parameters when dealing with a disjunction of step functions. In the present disclosure, the disjunction of step functions may mean a role or an operation of step functions that classifies input values or states corresponding to

the input values based on at least one threshold value. In this scenario, user device 120 may perturb the parameters by a random value, for example, one selected from a Gaussian distribution. A Gaussian distribution is itself described by two parameters, a mean and a variance, which are selected by model server 210. User device 120 generates a random number using the server (i.e., model server 210)-provided mean and variance and adds that random number value to a threshold value. User device 120 repeats this process for each threshold parameter. User device 120 uses the perturbed threshold values when computing a risk score.

[72] Example Parameter Selection - Neural Network - As mentioned above, the weights are considered parameters in a neural network. As in the case of disjunction of step functions, user device 120 may perturb each weight by a separate random number selected from the Gaussian distribution. User device 120 uses the perturbed values when computing a risk score.

[73] Example Parameter Selection - Decision Tree - As mentioned above, the hierarchy of decisions and the output risk scores are the parameters of a decision tree. Each decision may implicitly contain a threshold, which may be perturbed by user device 120 by a random number selected from the Gaussian distribution. User device 120 may also perturb the output risk score by drawing from another Gaussian distribution. In addition, information drawn from the another Gaussian distribution is another random number selected. In this example, user device 120 would not randomly modify the hierarchy of decisions in the decision tree.

[74] Identifying Justifications - User device 120 provides justifications for a given risk score based on components of the private input vector that most drastically affect the result. For example, user device 120 may determine justifications by the following methods. In one method, user device 120 directly tests a set of rules against a private input vector. The rules may be parameterized. For example, the time since last usage may be tested against some parameterized threshold. In a second method, user device 120 may re-compute the risk score using the negation(s) of each input component. The negation(s) means properties of the other state which is different from an original state. For example, if it is true that the fraudulent activity is considered a high risk, the negation(s) of the high risk (i.e., the fraudulent activity is not considered a high risk) is false. The negations that most drastically affect a risk score are tested against a set of rules.

[75] If user device 120 re-computes a risk score using negations of an input vector component, the component may have a finite number of negations. For example, a component that takes categorical values may have a negation that represents all other values in the same category. A component that represents a probability of an event occurring may have a negation that represents the probability of that event not

occurring. In cases where the negation of a component is not well defined, user device 120 may determine the contribution of a parameter to the risk score by a more complex operation. For example, determining the contribution of a time parameter may involve integrating the distribution of potential risk scores, weighted by a Gaussian distribution, with respect to time.

[76] When user device 120 generates risk scores, user device 120 may also generate parameter update values and may send these to model server 210 (i.e., cloud server 130) along with the risk scores. The parameter update values represent a set of changes to device-specific parameters that should be made for the computed risk scores to be more accurate. User device 120 generates two sets of parameter update values for each risk score. A first set of the two sets of parameter update values is the parameter changes to make given that the current action is not fraudulent; and a second set of the two sets of parameter update values is the parameter changes to make given that the current action is fraudulent. The actual calculation for parameter updates depends on the interpretation of the risk score being computed. The parameter update for a given component may be rule-based or may involve computing the gradient of an error term with respect to each private input parameter.

[77] Model server 210 retains the sets of parameter changes for a fixed period of time. If the action is later found to be fraudulent, the second set of parameter changes are used to update the base model parameters of model server 210. The second set of parameter changes means parameter changes which are considered that the action is fraudulent now. If within the fixed period of time the action is not found to be fraudulent, the first set of parameter changes is used. The first set of parameter changes represents parameter changes which are considered that the action is not fraudulent now. All parameter changes for a given model are aggregated and applied to the current model in order to generate a new model. The new model base parameters are pushed to every user device 120 along with a new random distribution. The new random distribution and the new model base parameters are input into all of user devices 120, then each user device 120 selects a new device-specific model based on the updated information and computes risk scores and parameter updates using the new model.

[78] Example Updates - Disjunction of Step Functions - FIGURE 5 illustrates updating thresholds for one of a disjunction of step functions. As mentioned above, the thresholds are considered parameters when dealing with a disjunction of step functions. Referring to FIGURE 2, when computing a risk score, user device 120 may compute update suggestions in the following way. First, if a threshold is not passed (i.e., the action is not determined to be fraudulent), then user device 120 may suggest: i) lowering the threshold by a fixed amount if the action is later found to be fraudulent; and ii) no changes if the action is not found to be fraudulent. Alternatively, if the

threshold is passed (i.e., the action is determined to be fraudulent), then user device 120 may suggest: i) no changes if the action is later found to be fraudulent; and ii) raising the threshold by a fixed amount if the action is later found not to be fraudulent.

[79] Many suggestions may be aggregated on model server 210, which may apply the suggestions to a base set of thresholds. This process may be repeated until there are an equal number of user devices 120 that suggest lowering the threshold as there are user devices 120 that suggest raising the threshold. Because each user device 120 randomizes the threshold, it cannot be stated with certainty that a user device 120 input was greater than or less than the server-provided threshold.

[80] Example Updates - Neural Network - As mentioned above, the weights are considered parameters when dealing with a neural network. When computing a risk score, user device 120 may compute updates in the following way:

[81] Step 1 - User device 120 computes the mean-square error of the output risk score assuming that the action is fraudulent. The error here would be the difference of the output risk score from 1.0. The 1.0 may mean that the action is fraudulent.

[82] Step 2 - User device 120 compare the gradient of the mean-squared error with respect to each weight parameter.

[83] Step 3 - User device 120 multiplies the gradient by a fixed value (the learning rate) and suggests the scaled gradient as the update value for when the action is found to be fraudulent.

[84] Steps 1-3 are repeated assuming the action is not fraudulent. This means treating the error as the difference of the output risk score from 0.0. The 0.0 may mean that the action is not fraudulent. Many gradients may be aggregated on model server 210, which may add the aggregated values to the base set of weights. User device 120 may repeat an above-described process until the aggregate gradient is equal to zero for all parameters.

[85] FIGURE 6 illustrates an operating flowchart of user device 120 according to an exemplary embodiment of the disclosure.

[86] Referring to FIGURE 6, in step 605, user device 120 calculates the risk score. In other words, user device 120 determines the risk metric. In particularly, user device 120 determines the risk metric that indicates a possibility of fraudulent activity associated with a payment based on a user input that associated with the payment. For example, the user input may be a name, an address, a phone number and the like of the user. The user input received by user device 120 may be at least one of the private information of the user and personally identifiable information of the user necessary for the payment. For example, user device 120 may identify the user input through an input device such as a keypad, a touch screen, a microphone, and the like. Furthermore, user device 120 may receive signals representing the user input via an

external communication path. The risk metric is determined using the confidential information related to the user and a basic tool for determining the risk. The confidential information is stored on user device 120 and the basic tool is received from model server 210. In other words, user device 120 determines the risk metric based on at least of the user input, the confidential information associated with the user that is stored on user device 120, and the basic tool.

[87] In step 610, user device 120 transmits payment information and information related to the risk metric associated with the payment. In this case, user device 120 transmits the payment information and the information related to the risk metric associated with the payment without revealing the confidential information. In here, the payment information may include information notifying a request of the payment and information that is necessary for the payment (e.g., accounts or identification information of cards).

[88] Although the present disclosure has been described with an exemplary embodiment, various changes and modifications may be suggested to one skilled in the art. It is intended that the present disclosure encompass such changes and modifications as fall within the scope of the appended claims.

Claims

[Claim 1] A user device comprising:
a processor configured to determine a risk metric indicating a fraudulent activity associated with a payment, according to a user input associated with the payment, based on confidential information associated with the user; and
a communication unit configured to transmit to a payment server information related to the payment and information related to the risk metric that is associated with the payment.

[Claim 2] The user device as set forth in Claim 1, wherein the confidential information comprises private information of the user.

[Claim 3] The user device as set forth in Claim 1, wherein the communication unit is configured to receive a risk model for determining the risk metric from a model server.

[Claim 4] The user device as set forth in Claim 3, wherein the communication unit is configured to transmit to the model server a suggested parameter update usable by the model server to improve the accuracy of the risk model.

[Claim 5] The user device as set forth in Claim 3, wherein the risk model is based on a neural network.

[Claim 6] The user device as set forth in Claim 3, wherein the risk model is based on a decision tree.

[Claim 7] The user device as set forth in Claim 1, wherein the communication unit is configured to transmit to the payment server a filter corresponding to the risk metric.

[Claim 8] A method of operating a user device, the method comprising:
determining a risk metric indicating a fraudulent activity associated with a payment, according to a user input associated with the payment, based on confidential information associated with the user; and
transmitting to a payment server information related to the payment and information related to the risk metric that is associated with the payment.

[Claim 9] The user device as set forth in Claim 1 or the method as set forth in Claim 8, wherein the confidential information comprises personally identifiable information of the user.

[Claim 10] The method as set forth in Claim 8, wherein the confidential information comprises private information of the user.

[Claim 11] The method as set forth in Claim 8, wherein the risk metric is determined by using a risk model, received from a model server, for determining the risk metric.

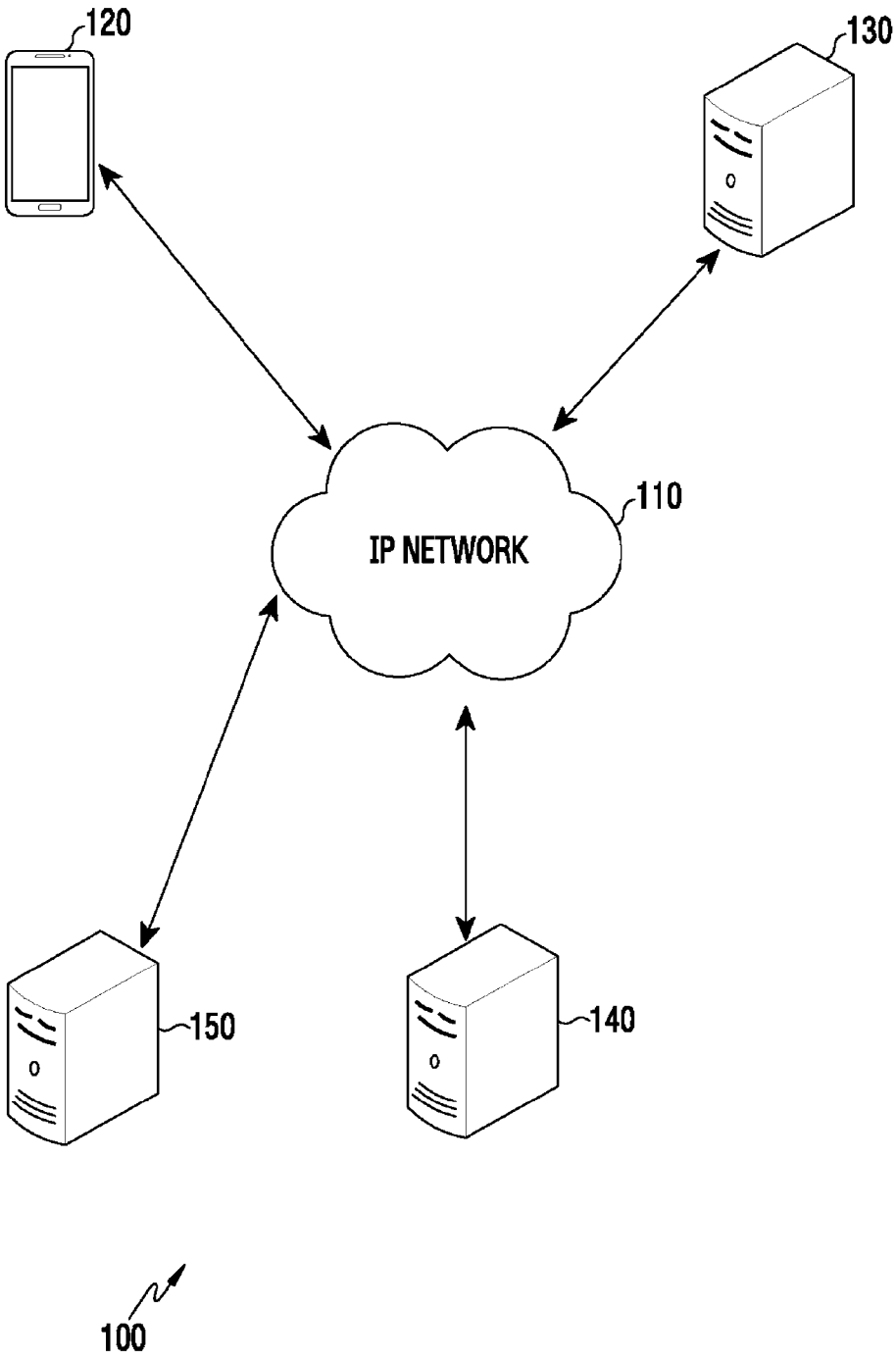
[Claim 12] The method as set forth in Claim 11, further comprising:
transmitting to the model server a suggested parameter update usable by the model server to improve the accuracy of the risk model.

[Claim 13] The method as set forth in Claim 11, wherein the risk model is based on a neural network.

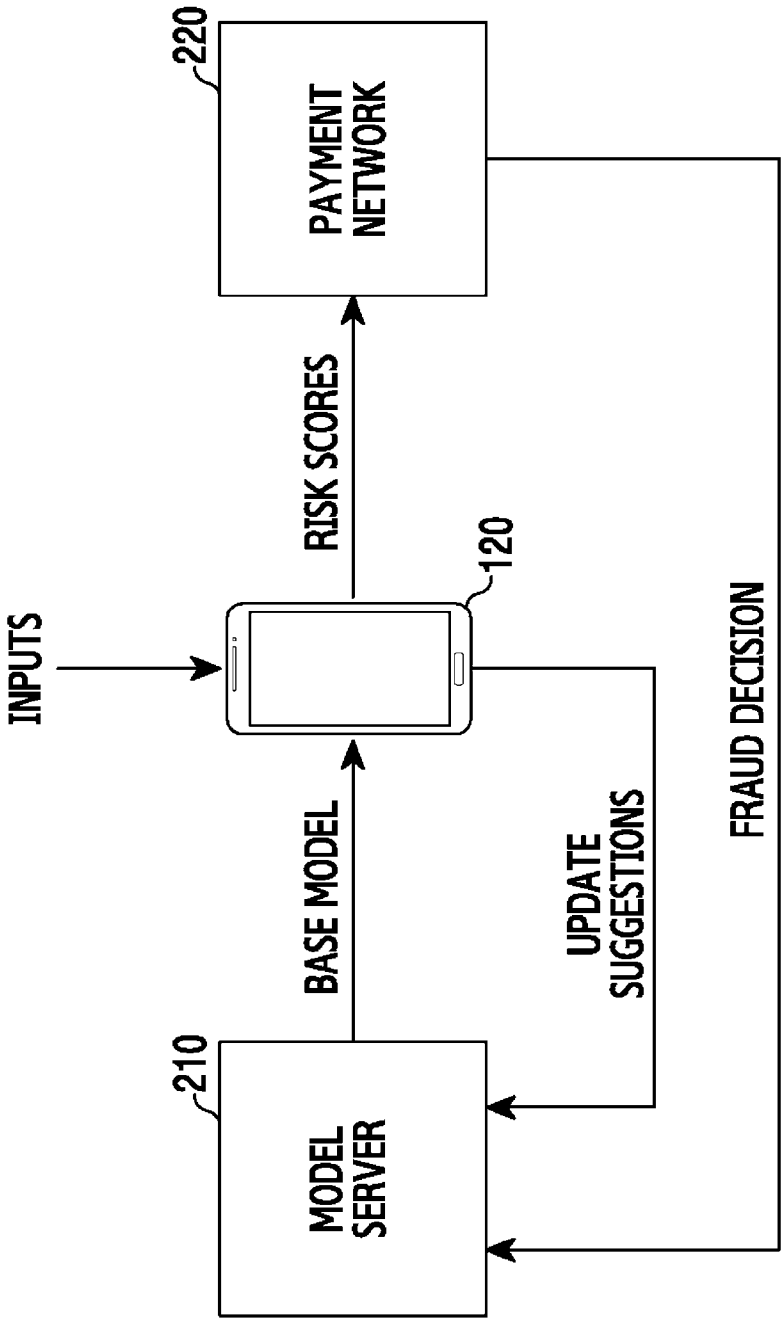
[Claim 14] The method as set forth in Claim 11, wherein the risk model is based on a decision tree.

[Claim 15] The method as set forth in Claim 9, further comprising:
transmitting to the payment server a filter corresponding to the risk metric.

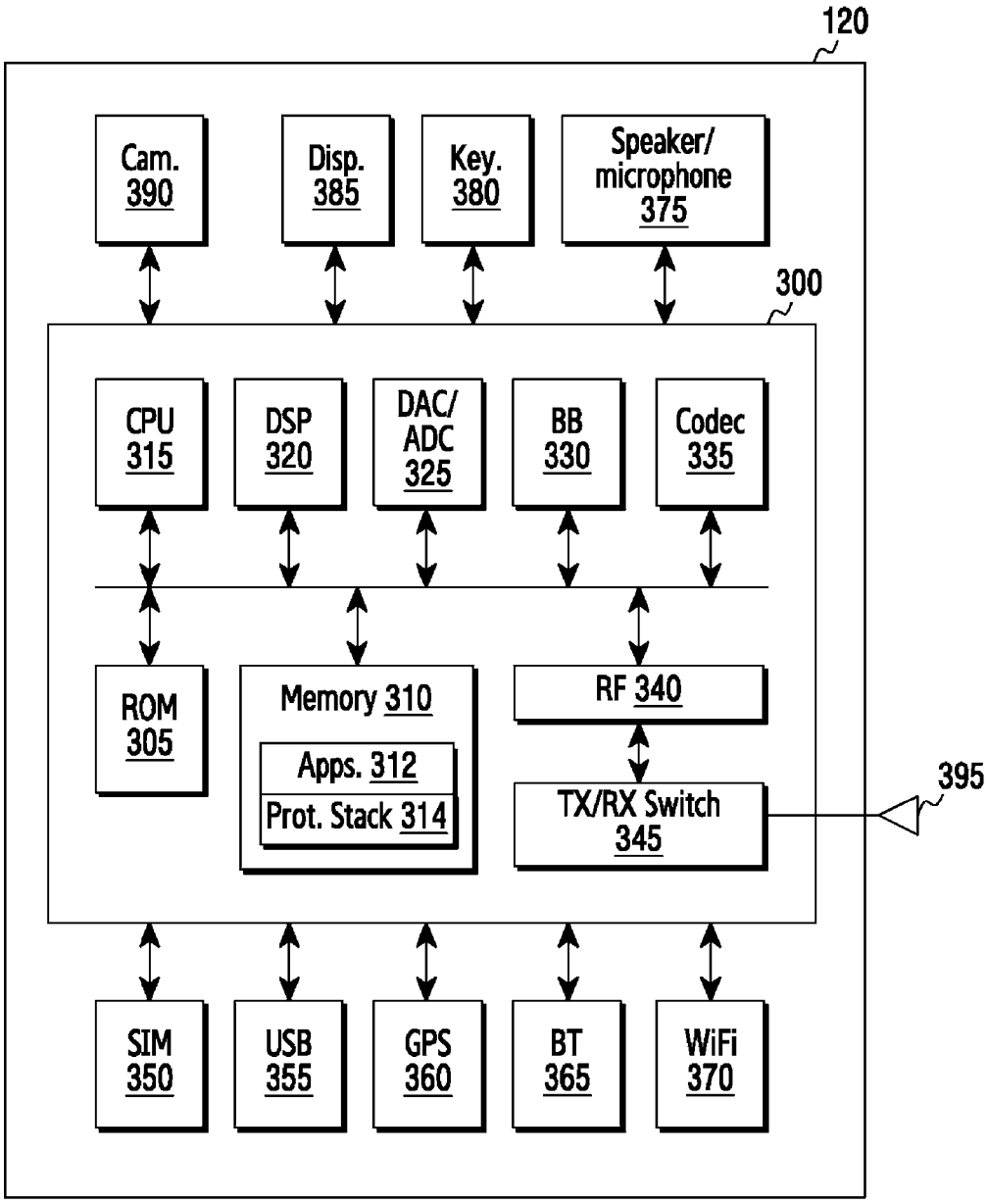
[Fig. 1]



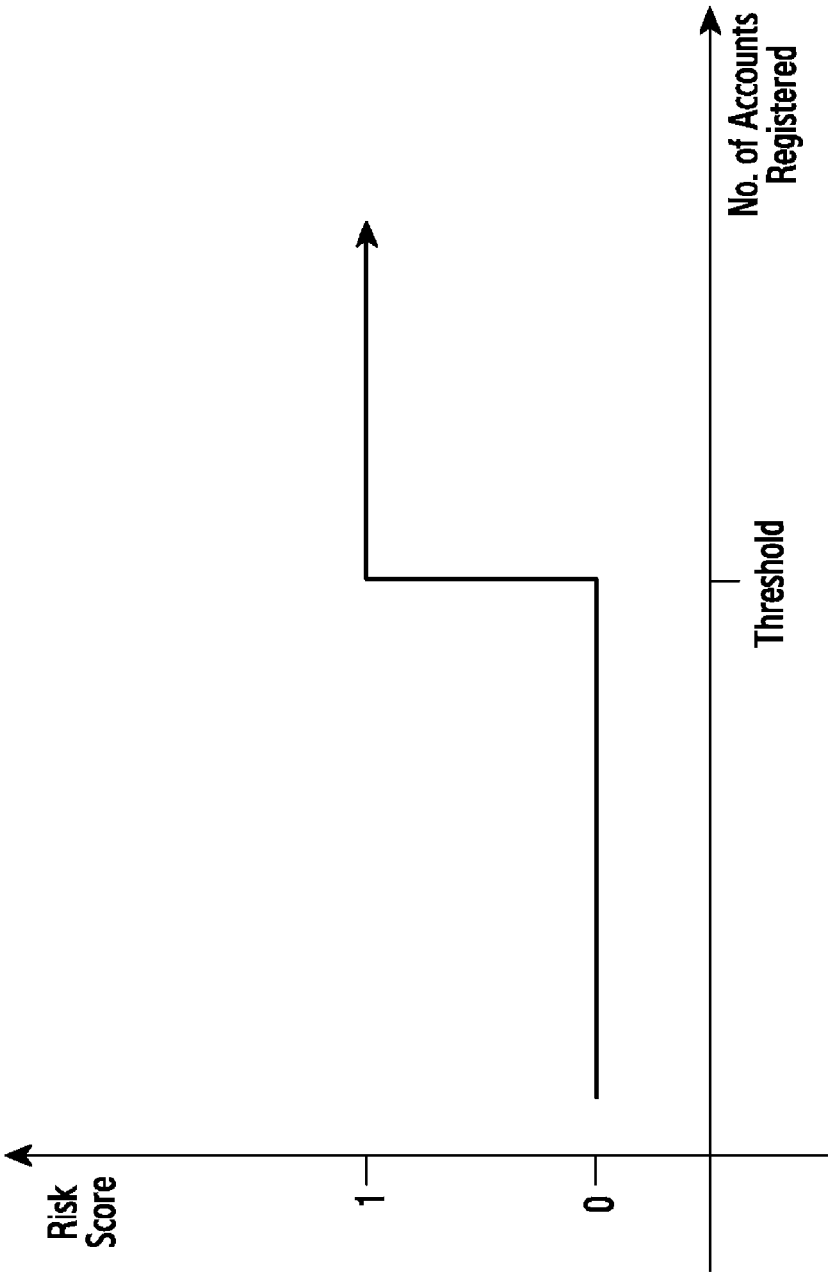
[Fig. 2]


220
PAYMENT NETWORK
RISK SCORES
120
INPUTS
FRAUD DECISION
BASE MODEL
UPDATE SUGGESTIONS
210
MODEL SERVER

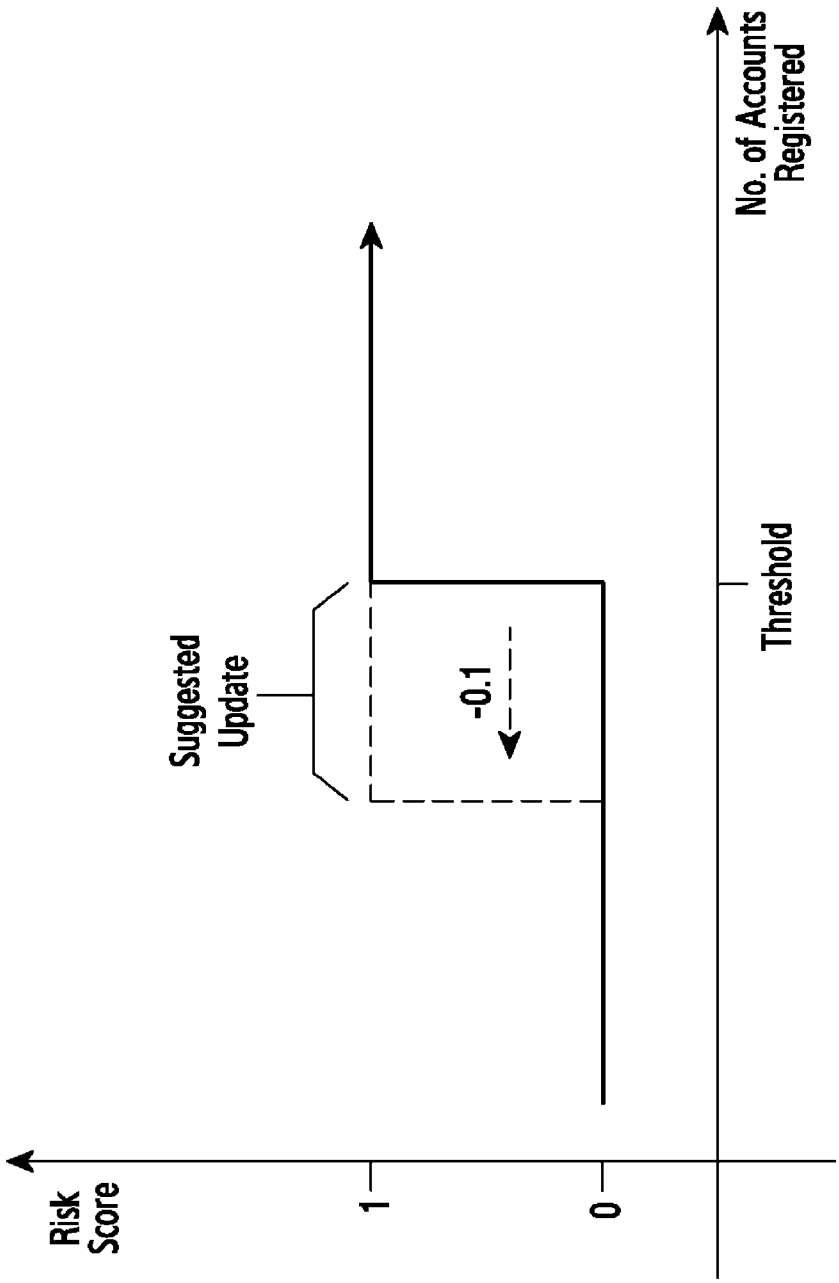
[Fig. 3]


120
Cam. 390
Disp. 385
Key. 380
Speaker/ microphone 375
300
CPU 315
DSP 320
DAC/ ADC 325
BB 330
Codec 335
ROM 305
Memory 310
Apps. 312
Prot. Stack 314
RF 340
TX/RX Switch 345
395
SIM 350
USB 355
GPS 360
BT 365
WiFi 370

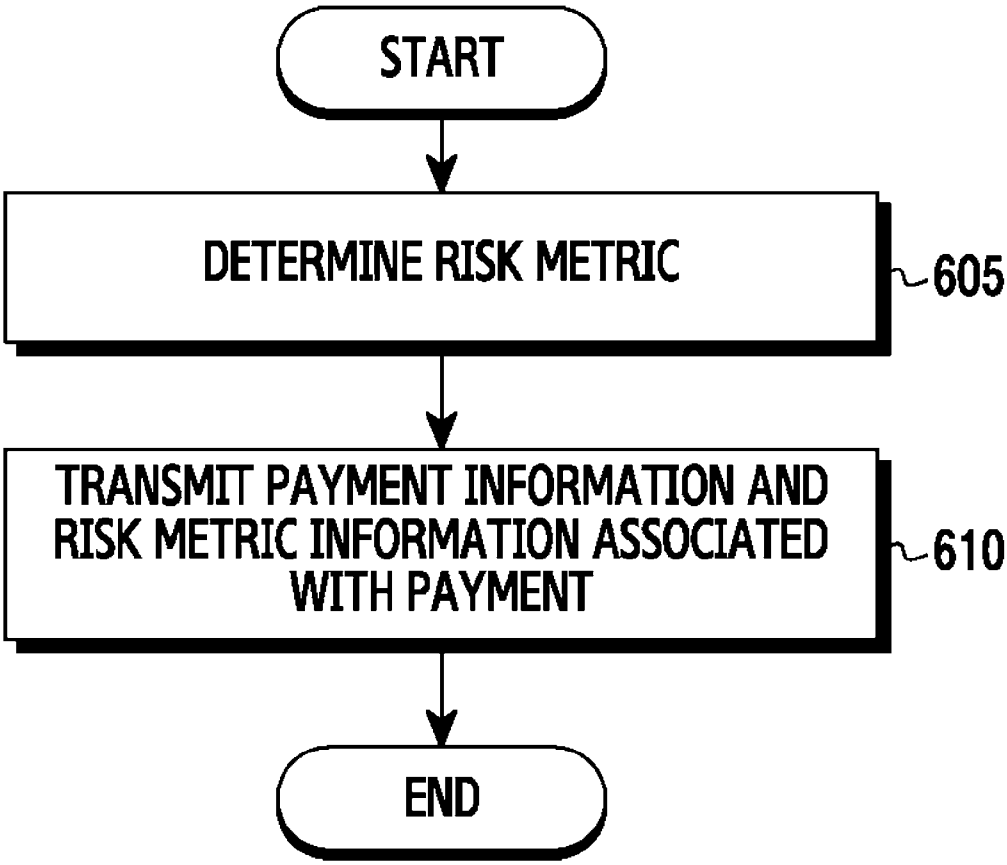
[Fig. 4]



[Fig. 5]


No. of Accounts Registered
Threshold
Suggested Update
-0.1
Risk Score
1
0

[Fig. 6]


START
DETERMINE RISK METRIC
605
TRANSMIT PAYMENT INFORMATION AND RISK METRIC INFORMATION ASSOCIATED WITH PAYMENT
610
END

INTERNATIONAL SEARCH REPORT

International application No.
PCT/KR2016/003844

A. CLASSIFICATION OF SUBJECT MATTER

G06Q 20/40(2012.01)i, G06Q 20/38(2012.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G06Q 20/40; H04W 4/12; G06F 17/60; G06Q 40/00; G06Q 10/00; G06Q 20/38

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
Korean utility models and applications for utility models
Japanese utility models and applications for utility models

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
eKOMPASS(KIPO internal) & Keywords: risk, metric, score, model, confidential, personal, fraudulent

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2013-0024238 A1 (BRETT A. NIELSON et al.) 24 January 2013 See paragraphs [0002],[0008]-[0011], claims 7,21,26 and figures 1-3.	1-15
Y	US 8626663 B2 (BRAD NIGHTENGALE et al.) 07 January 2014 See column 3, lines 62-67, column 5, lines 58-61, column 7, lines 52-55, claims 1-4 and figures 2-3.	1-15
Y	US 2004-0199462 A1 (ED STARRS) 07 October 2004 See abstract, claims 1-5 and figure 3.	5,13
Y	US 2005-0283429 A1 (MICHAEL BATES et al.) 22 December 2005 See paragraph [0022], claims 1-7 and figure 2.	6,14
A	US 2009-0307778 A1 (UPENDRA MARDIKAR) 10 December 2009 See abstract, claims 1-13 and figures 2-3.	1-15

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents,such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
30 June 2016 (30.06.2016)	**30 June 2016 (30.06.2016)**

Name and mailing address of the ISA/KR
International Application Division
Korean Intellectual Property Office
189 Cheongsa-ro, Seo-gu, Daejeon, 35208, Republic of Korea
Facsimile No. +82-42-481-8578

Authorized officer
Jang, Gijeong
Telephone No. +82-42-481-8364

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/KR2016/003844

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2013-0024238 A1	24/01/2013	US 8473318 B2	25/06/2013
US 8626663 B2	07/01/2014	US 2011-0238575 A1	29/09/2011
		US 2014-0095393 A1	03/04/2014
		WO 2011-119761 A2	29/09/2011
		WO 2011-119761 A3	12/04/2012
US 2004-0199462 A1	07/10/2004	WO 2004-090690 A2	21/10/2004
		WO 2004-090690 A3	06/05/2005
US 2005-0283429 A1	22/12/2005	CA 2571251 A1	26/01/2006
		US 2012-0066132 A1	15/03/2012
		US 8082207 B2	20/12/2011
		WO 2006-009541 A1	26/01/2006
US 2009-0307778 A1	10/12/2009	CN 102057386 A	11/05/2011
		CN 102057386 B	01/07/2015
		CN 105046479 A	11/11/2015
		EP 2308014 A1	13/04/2011
		EP 2308014 A4	06/11/2013
		US 2009-0305673 A1	10/12/2009
		US 2009-0307139 A1	10/12/2009
		US 2009-0307140 A1	10/12/2009
		US 2009-0307142 A1	10/12/2009
		US 2012-0089520 A1	12/04/2012
		US 2012-0173434 A1	05/07/2012
		US 2013-0198086 A1	01/08/2013
		US 2015-0056957 A1	26/02/2015
		US 8108318 B2	31/01/2012
		US 8150772 B2	03/04/2012
		US 8417643 B2	09/04/2013
		US 8543091 B2	24/09/2013
		US 8554689 B2	08/10/2013
		WO 2009-149376 A1	10/12/2009
		WO 2010-002541 A1	07/01/2010