

(19) **DANMARK**

(10) **DK/EP 3662695 T3**



(12) **Oversættelse af
europæisk patentskrift**

Patent- og
Varemærkestyrelsen

-
- (51) Int.Cl.: **H 04 L 47/724 (2022.01)** **H 04 L 47/125 (2022.01)** **H 04 L 47/80 (2022.01)**
- (45) Oversættelsen bekendtgjort den: **2025-02-03**
- (80) Dato for Den Europæiske Patentmyndigheds bekendtgørelse om meddelelse af patentet: **2025-01-15**
- (86) Europæisk ansøgning nr.: **18840920.5**
- (86) Europæisk indleveringsdag: **2018-07-16**
- (87) Den europæiske ansøgnings publiceringsdag: **2020-06-10**
- (86) International ansøgning nr.: **IL2018050780**
- (87) Internationalt publikationsnr.: **WO2019026061**
- (30) Prioritet: **2017-08-03 US 201762540583 P**
- (84) Designerede stater: **AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR**
- (73) Patenthaver: **Drivenets Ltd., 4 HaSheizaf Street, 4366411 Raanana, Israel**
- (72) Opfinder: **KRAYDEN, Amir, 22 Shdema Street, Even Yehuda, Israel**
GARTSBEIN, Anton, 2 Dekel Street, 3605653 Kiryat Tivon, Israel
ZOLKOVER, Gal, 6 Hasport Street, 3457406 Haifa, Israel
ZAKIN, Ori, P.O.Box 541, 4493500 Sde Varburg, Israel
LEV, Yuval, 12/1 Hakad Street, 7087015 Gan Yavne, Israel
MOSHE, Yuval, 26 Neomi Shemer Street, Ra'anana 4334358, Israel
SADEH, Or, 114 Hanadiv Street, Pardes Hanna 3711915, Israel
- (74) Fuldmægtig i Danmark: **Zacco Denmark A/S, Arne Jacobsens Allé 15, 2300 København S, Danmark**
- (54) Benævnelse: **ET NETVÆRKSBEVIDST ELEMENT OG EN FREMGANGSMÅDE TIL ANVENDELSE DERAFF**
- (56) Fremdragne publikationer:
GB-A- 2 481 719
US-A1- 2007 180 119
US-A1- 2007 220 521
US-A1- 2016 191 348
US-A1- 2016 301 582
US-A1- 2017 048 115
US-B1- 10 116 732

DESCRIPTION

Description

TECHNICAL FIELD

[0001] The present disclosure relates generally to the field of cloud computing and networking, and in particular, to a device and a method for controlling allocation of network resources to a software application operating in a cloud environment.

BACKGROUND

[0002] Cloud computing is a technique that is being used for a wide variety of enterprise and end-user applications. Providers offer varying service portfolios that differ in resource configurations and services that can be provided. A comprehensive solution for allocation of resources is fundamental to any cloud computing service provider. Any resource allocation model has to consider available computational resources as well as network resources, to accurately reflect practical demands. Resource allocation algorithms aim to accomplish the task of scheduling virtual machines on the servers residing in data centers, in order to schedule available network resources in a way that complies with the problem constraints. Several external and internal factors that affect the performance of resource allocation models have been considered in the past.

[0003] However, as cloud computing platforms adopt virtualization as an enabling technology, it becomes an increasingly challenging task to distribute system resources to each user according to the diverse requirements.

[0004] Furthermore, although ample research has already been conducted in order to meet QoS requirements, still, the proposed solutions lack simultaneous support for multiple networks' policies, degrade the aggregated throughput of network resources, and incur CPU overhead.

[0005] One typical characteristic that is shared by the various solutions, is based on management of the computing resources by applications which are implemented in a cloud environment (either private or public environment), for example, the management is carried out by the application through adding and removing CPUs that are allocated to that specific application.

[0006] US 20110320606 describes for example a data processing system for managing network resources. Upon receiving a request for resources in the network data processing system, which comprises requested values for a plurality of attributes for the resources in the network data processing system, the potential resources in the network data processing system are identified. The request is evaluated based on proximity scores for the plurality of attributes in the potential resources relative to an ideal allocation of the potential resources and resource weights for the plurality of attributes in the potential resources.

[0007] US 20130185729 discloses an apparatus for managing resources assigned to an application or service. A resource manager maintains a set of workload classes and classifies workloads using workload signatures. For example, the resource manager minimizes or reduces resource management costs by identifying a relatively small set of workload classes during a learning phase, determining preferred resource allocations for each workload class, and then during a monitoring phase, classifying workloads and allocating resources based on the preferred resource allocation for the classified workload.

[0008] GB2481719A discloses a mobile telecommunications network including a core and a radio access network having radio means for wireless communication with mobile terminals registered with the network, wherein the radio access network includes control means (fig. 13) operable to control the use of network resources by said mobile terminals, wherein the control means is operable to perform Self Organising Network functions. The control means may monitor the device type, customer types and/or application used. The control means may in real time modify radio specific parameters, traffic handling algorithms and/or processor capacity network functions.

[0009] US 2017048115 A1 discloses an SDN application integration, management and control method, system and device, herein the SDN application integration, management and control system includes: an application integration module configured to integrate an SDN application in an application plane, herein integrating an SDN application includes at least one of registering, deregistering and updating the SDN application and reconstructing a plurality of SDN applications; an application management module configured to manage the SDN application in the application plane; and an application control module configured to be aware of and monitor the SDN application in the application plane. By using the SDN application integration, management and control method, system and device, the management of the SDN application is implemented.

[0010] US 2016301582A1 discloses utilizing collected data from a software-defined networking (SDN) network to diagnose a user experience includes receiving application session quality metrics from an application within a SDN network, collecting network information from network components that carry network traffic of the SDN network, and utilizing the application session quality metrics and the network information to run a diagnostic analysis to improve the user experience in real-time or for future sessions.

[0011] US 2016301582 discloses utilizing collected data from an SDN network to diagnose a

user experience by receiving application session quality metrics from an application running within a SDN network, collecting network information from network components that carry network traffic of the SDN network, and utilizing the application session quality metrics and the network information to run a diagnostic analysis to improve the user experience.

[0012] US 2007180119 describes transmitting data in a broadcast mode to multiple devices operating in a network, a solution that enables efficient utilization of bandwidth while providing a desired level of quality of service for the applications executing on the devices that utilize the broadcasted data.

SUMMARY

[0013] The invention is defined by the appended claims.

[0014] It is an object of the present disclosure to provide a system and a method for improving performance of data processing.

[0015] It is another object of the present disclosure to provide a device and a method for affecting control over network resources via a network element such as a router.

[0016] It is another object of the present disclosure to provide a device and a method to enable a software application to communicate with a router and exchange messages therewith in order to allow the router to control network resources utilized by that software application.

[0017] Other objects of the present disclosure will become apparent from the following description.

[0018] According to an illustrative example, there is provided a network element (e.g. a router) operative in an IP-based communication network (e.g. a cloud computing environment in an IP-based computer network that provides, on demand, networking resources and data to devices comprised in the IP-based communication network, for running at least a first software application and a second software application), and configured to manage networking resources by affecting changes at the communication network level of networking resources associated with the communication network that are being utilized by the first software application, wherein the changes are affected in response to obtaining information retrieved from an application program interface (API) associated with the first software application by users communicating with said first software application, and in case that the second software application has a low demand for networking resources, then the first software application is configured to use these networking resources on request, and wherein the first software application is configured to assess a new user's experience, achieved after the change in the networking resources had been affected, of users communicating with said first software application.

[0019] According to another illustrative example, the information retrieved from the API comprises values of key performance indicators (KPIs) associated with that software application.

[0020] By yet another illustrative example, the network element is a member selected from a group that comprises a router, a switch and the like.

[0021] According to another illustrative example, there is provided a method for use by a network element operative in an IP-based communication network (e.g. a cloud computing environment in an IP-based computer network), for managing resources that are being utilized by a software application, wherein the method comprises affecting changes by the network element in these resources, in response to obtaining information retrieved from an application program interface (API) associated with that software application.

[0022] In accordance with another illustrative example, the information retrieved from the API comprises values of key performance indicators (KPIs) associated with said software application.

[0023] According to another illustrative example of this aspect of the disclosure, the method comprises the steps of:

1. (i) registering the software application with the network element;
2. (ii) providing the software application with information that relates to a user experience, of users communicating with the software application;
3. (iii) providing the network element with information that would enable the network element to affect changes in resources associated with the communication network that are being utilized by the software application; and
4. (iv) affecting one or more changes in these resources based on the information provided in step (iii).

[0024] By still another embodiment, the method further comprising:

(v) retrieving information for assessing results of affecting at least one of the one or more changes; and

(vi) providing the software application with information, based on the information retrieved in step (v).

[0025] According to yet another embodiment of this aspect of the invention, the network element is a router.

[0026] In accordance with still another illustrative example, the software application is

registered with the router via a network API to enable the router to identify the user experience, based on retrieved values of (KPIs) associated with the software application.

[0027] According to another illustrative example, the information provided in step (ii) is information that relates to a user experience as reflected by the router and/or by the communication network.

[0028] By still another illustrative example, the information that would enable the network element to affect changes in the resources provided in step (iii), comprises one or more indications that respective pre-defined KPI thresholds have been crossed.

[0029] According to still another illustrative example, the changes affected in step (iv) is a member of a group that consists of: changing QoS profile, changing routing metric, and changing virtual routing and forwarding (VRF) of packets.

[0030] In accordance with another illustrative example, the method provided is incorporated in a firewall activity to reduce processing load (e.g. calculations) that would otherwise be required to be carried out by the firewall in the case of an attack on the software application, which in turn would result in a request that would be initiated by the firewall to divert or block traffic at the network level.

[0031] By still another illustrative example, the software application is a storage software application, and the method is configured to enable routing data being conveyed for storage by the storage software application, in case that an approach to a respective storage device is currently adversely affected by a network congestion.

[0032] According to another illustrative example, the method is configured to enable allocating additional bandwidth to the software application in order to overcome a temporary performance problem associated therewith.

[0033] In accordance with yet another illustrative example, the software application is configured to affect one or more temporary network operational changes in order to enhance user experience.

BRIEF DESCRIPTION OF THE DRAWINGS

[0034] The accompanying drawings, which are incorporated herein and constitute a part of this specification, illustrate several illustrative examples and, together with the description, serve to explain the principles of the illustrative examples disclosed herein.

FIG. 1 illustrates a high-level schematic view of a system configured to operate by executing a closed loop algorithm for managing network resources controlled by a software application, according to an illustrative example of the present invention; and

FIG. 2 presents a flow chart, illustrating a method of carrying out an illustrative example of the present invention.

DESCRIPTION OF EXEMPLARY EMBODIMENTS

[0035] Some of the specific details and values in the following detailed description refer to illustrative examples of the disclosure.

[0036] Cloud computing is a form of Internet-based computing that provides shared computer processing resources and data to computers and other devices, on demand. It is a model for enabling ubiquitous, on-demand access to a shared pool of configurable computing resources (e.g., computer networks, servers, storage, applications and services), which can be rapidly provisioned and released while applying minimal management effort. Basically, cloud computing allows the users and enterprises with various capabilities to store and process their data in either privately owned cloud, or on a third-party server in order to make data accessing mechanisms much more easy and reliable. Cloud computing relies on sharing of resources to achieve coherence and economy of scale, similar to a utility (like the electricity grid) over an electricity network.

[0037] Applications that are running in a cloud environment are by definition adapted to share computing, storage and networking resources for better utilization of the available resources. The underlying assumption is that if a first application has a low demand for resources, then a second application may use these resources on request. In today's environment, computing resources may indeed be managed this way. Yet, networking resources are more complicated to manage in this manner, mainly due to the fact that the networking layer is traditionally not an integrated component of servers at which the application is installed, but is comprised of networking elements like routers and switches that are managed and dimensioned based on static configurations and pre-assumptions of the required resources for all hosted applications at a given server or data center.

[0038] FIG. 1 illustrates a high-level schematic view of a system (5) construed in accordance with an illustrative example. System 5 is configured to operate by executing a closed loop algorithm for managing network resources that are controlled by a software application, whereas FIG. 2 exemplifies steps that are taken in order to carry out an illustrative example of the present invention, by which it becomes possible to link between network security, storage, cloud computing services and other network services, thereby enabling a unique user experience with an overall better performance.

[0039] Network element 10 of FIG. 1 is a network aware manager, which, according to the present invention may be a router or a switch or any other applicable network element which can be configured to function in compliance with the principles set forth by the present

disclosure. Network element 10 is depicted in FIG. 1 as a device configured to communicate with network elements 15 and 20 which may be for example physical servers and/or white boxes. The term white box is used herein to denote a commodity, being an open or industry-standard compliant hardware for switches and/or routers within the forwarding plane. White boxes provide users with the foundational hardware elements of a network.

[0040] Typically, an application program interface (hereinafter: "API") is a set of routines, protocols, and tools for building software applications. An API specifies how software components should interact. Additionally, APIs may be used when programming graphical user interface (GUI) components. A good API makes it easier to develop a program by providing all the building blocks, and then incorporating these blocks together. Most operating environments, such as MS-Windows™ for example, provide APIs, thereby allowing programmers to write applications consistent with the relevant operating environment.

[0041] According to an illustrative example, the network element (e.g. router 10) retrieve from the relevant API of one or more software applications, information that relates to at least the following two characteristics associated with the respective software application:

1. a) Network visibility, e.g. KPIs (Key Performance Indicators) that serve as indication of the software application usage, such as TCPIP window size, retransmission of packets, and the like.
2. b) Network related control and commands.

[0042] The term performance indicator or key performance indicator (KPI) as used herein throughout the specification and claims is used to denote a performance measurement. Key performance indicators define a set of values against which to measure. These raw sets of values, which are fed to systems in charge of summarizing the information, are referred to as indicators. KPIs may evaluate the performance of a particular activity associated therewith, and such an evaluation may often lead to the identification of potential improvements.

[0043] FIG. 2 exemplifies a method for using router 10 in accordance with an illustrative example.

[0044] Step 1 - Service identification. In this step, the application is registered with the router via the network visibility API so that the router is able to identify the user experience, based on relevant KPIs once retrieved.

[0045] Step 2 - Service notification. Based on the retrieved KPIs, the application is notified about the user experience, as reflected by the network/router.

[0046] Step 3 - Taking an appropriate action by the network element (the router). In case that pre-defined KPIs threshold have been crossed (e.g. exceeded), the router takes an action at

the network level in order to enhance the application's user experience. Such actions are for example, changing QoS profile, changing routing metric, changing virtual routing and forwarding ("VRF"), and the like.

[0047] Step 4 - Feedback. After carrying out a change that has been initiated by the network element (the router), the latter evaluates the results of the affected change and then initiates conveyance of information toward the application, which enables the application to assess the new user experience achieved after the change has been affected.

[0048] Following are number of cases exemplifying various implementations of the above-described solution.

Security

[0049] The proposed solution may be used by having it integrated with firewalls. In the case of an attack on the application, the firewall may request a traffic diversion or blocking at the network level, and using the proposed solution helps in reducing the amount of processing that would otherwise be required for the firewall to carry out, thereby allowing more services to be carried out by the firewall, and consequently, to enable increasing the firewall's capacity.

Storage

[0050] The proposed solution may be used by having it integrated with storage applications. For example, in case where there is a plurality of storage devices and the approach to one or more of this plurality of storage devices is currently adversely affected by a network congestion, the respective one or more storage applications may request that the router would change the route of data being conveyed for storage by the one or more respective storage applications, to a better performing route.

Performance

[0051] The proposed solution may be used by having it integrated with the operative router, so that any application communicating with that router may request allocation of additional bandwidth for its operations in order to overcome temporary performance problems.

User experience

[0052] The proposed solution may be used by having it integrated with the operative router, so that any application communicating with that router which suffers from a diminished user

experience (as reflected by values of the relevant KPIs retrieved), may request initiating of a series of network changes in order to enhance the user experience, e.g. affecting a temporarily change to the QoS profile.

REFERENCES CITED IN THE DESCRIPTION

Cited references

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- [US20110320606A \[0006\]](#)
- [US20130185729A \[0007\]](#)
- [GB2481719A \[0008\]](#)
- [US2017048115A1 \[0009\]](#)
- [US2016301582A1 \[0010\]](#)
- [US2016301582A \[0011\]](#)
- [US2007180119A \[0012\]](#)

Patentkrav

5 **1.** Netværkselement (10) til et cloud-computer-miljø til et IP-baseret kommunikationsnetværk, hvor cloud-computer-miljøet er et internetbaseret computer-miljø, der efter behov leverer netværksressourcer og data til indretninger, der er omfattet af det IP-baserede kommunikationsnetværk, til at køre mindst en første softwareapplikation og en anden softwareapplikation, og hvor netværkselementet (10) er konfigureret til at styre netværksressourcer ved at påvirke ændringer på netværksniveauet af netværksressourcer tilknyttet kommunikationsnetværket, der bliver anvendt af den første softwareapplikation, hvor ændringerne påvirkes som reaktion på opnåelse af information hentet fra en applikationsprogramgrænseflade, API, tilknyttet den første softwareapplikation, og hvor den opnåede information omfatter netværksrelateret kontrol og kommandoer og information, der relaterer til brug af den første softwareapplikation af brugere, der kommunikerer med den første softwareapplikation, og i tilfælde af at den anden softwareapplikation har et lavt behov for netværksressourcer, så er den første softwareapplikation konfigureret til at bruge disse netværksressourcer efter anmodning, og hvor den første softwareapplikation er konfigureret til at vurdere en ny brugers oplevelse af brugere, der kommunikerer med den første softwareapplikation, opnået efter ændringen i netværksressourcerne er blevet påvirket.

25 **2.** Netværkselement (10) ifølge krav 1, hvor informationen hentet fra API'en omfatter værdier af nøgleydelsesindikatorer, KPI'er, tilknyttet softwareapplikationen.

3. Netværkselement (10) ifølge krav 1 eller 2, hvor netværkselementet (10) er en router.

30 **4.** Fremgangsmåde til brug af et netværkselement (10), der fungerer i et cloud-computer-miljø i et IP-baseret kommunikationsnetværk, hvor cloud-computer-

- miljøet er et internetbaseret computermiljø, der efter behov leverer netværksressourcer og data til indretninger, der er omfattet af det IP-baserede kommunikationsnetværk, til at køre mindst en første softwareapplikation og en anden softwareapplikation, og hvor fremgangsmåden anvendes til at styre netværksressourcer, hvor fremgangsmåden omfatter at påvirke ændringer på netværksniveauet af netværksressourcer tilknyttet kommunikationsnetværket, der bliver anvendt af den første softwareapplikation, hvor ændringerne påvirkes som reaktion på opnåelse af information hentet fra en applikationsprogramgrænseflade, API, tilknyttet den første softwareapplikation, og hvor den opnåede information omfatter netværksrelateret kontrol og kommandoer og information, der relaterer til brug af den første softwareapplikation af brugere, der kommunikerer med den første softwareapplikation, og i tilfælde af at den anden softwareapplikation har et lavt behov for netværksressourcer, så er den første softwareapplikation konfigureret til at bruge disse netværksressourcer efter anmodning, og hvor den første softwareapplikation er konfigureret til at vurdere en ny brugers oplevelse af brugere, der kommunikerer med den første softwareapplikation, opnået efter ændringen i netværksressourcerne er blevet påvirket.
- 5
- 10
- 15
- 20 **5.** Fremgangsmåde ifølge krav 4, hvor informationen hentet fra API'en omfatter værdier af nøglepræstationsindikatorer, KPI'er, tilknyttet softwareapplikationen.
- 25 **6.** Fremgangsmåde ifølge krav 5, hvor fremgangsmåden omfatter trinene:
- 30 (i) at registrere softwareapplikationen med netværkselementet (10);
(ii) at forsyne softwareapplikationen med information, der relaterer til en brugeroplevelse for brugere, der kommunikerer med softwareapplikationen;
(iii) at forsyne netværkselementet (10) med information, der ville gøre det muligt for netværkselementet (10) at påvirke ændringer i ressourcer tilknyttet kommunikationsnetværket, som bliver brugt af softwareapplikationen; og

(iv) at påvirke en eller flere ændringer i nævnte ressourcer baseret på informationen tilvejebragt i trin (iii).

7. Fremgangsmåde ifølge krav 6, hvor fremgangsmåden yderligere omfatter:

5 (v) at indhente information til vurdering af resultater af påvirkning af mindst én af en eller flere ændringer; og

(vi) at forsyne softwareapplikationen med information baseret på informationen hentet i trin (v).

10 **8.** Fremgangsmåde ifølge krav 6 eller 7, hvor netværkselementet (10) er en router.

9. Fremgangsmåde ifølge krav 8, hvor softwareapplikationen er registreret hos routeren via en netværks-API for at gøre det muligt for routeren at identificere brugeroplevelsen, baseret på hentede værdier af KPI'er tilknyttet softwareapplikationen.

15 **10.** Fremgangsmåde ifølge et af kravene 6-9, hvor informationen tilvejebragt i trin (ii) er information, der relaterer til en brugeroplevelse som afspejlet af routeren og/eller af kommunikationsnetværket.

20 **11.** Fremgangsmåde ifølge et af kravene 6-10, hvor informationen, der gør det muligt for netværkselementet (10) at påvirke ændringer i de ressourcer, der er tilvejebragt i trin (iii), omfatter en eller flere indikationer på, at respektive foruddefinerede KPI-tærskler er blevet overskredet.

12. Fremgangsmåde ifølge et af kravene 6-11, hvor ændringerne påvirket i trin (iv) er en del af en gruppe, der består af: ændring af QoS-profil, ændring af routingmetrik og ændring af virtuel routing og videresendelse, VRF, af pakker.

30

13. Fremgangsmåde ifølge et af kravene 6-12, hvor fremgangsmåden er inkorporeret i en firewall-aktivitet for at muliggøre, efter en anmodning initieret af firewallen i tilfælde af et angreb på softwareapplikationen, at omdirigere eller blokere trafik på netværksniveau.

5

14. Fremgangsmåde ifølge et af kravene 6-13, hvor softwareapplikationen er en lagringssoftwareapplikation, og hvor fremgangsmåden muliggør routing af data, der overføres til lagring af lagringssoftwareapplikationen i tilfælde af, at en tilgang til en respektive lagringsenhed i øjeblikket er negativt påvirket ved en overbelastning af nettet.

10

15. Fremgangsmåde ifølge et af kravene 6-14, hvor fremgangsmåden gør det muligt at allokere yderligere båndbredde til softwareapplikationen for at overvinde et midlertidigt ydeevneproblem forbundet dermed.

15

16. Fremgangsmåde ifølge et af kravene 6-15, hvor softwareapplikationen er konfigureret til at påvirke en eller flere midlertidige netværksdriftsændringer for at forbedre en brugeroplevelse.

DRAWINGS

Drawing

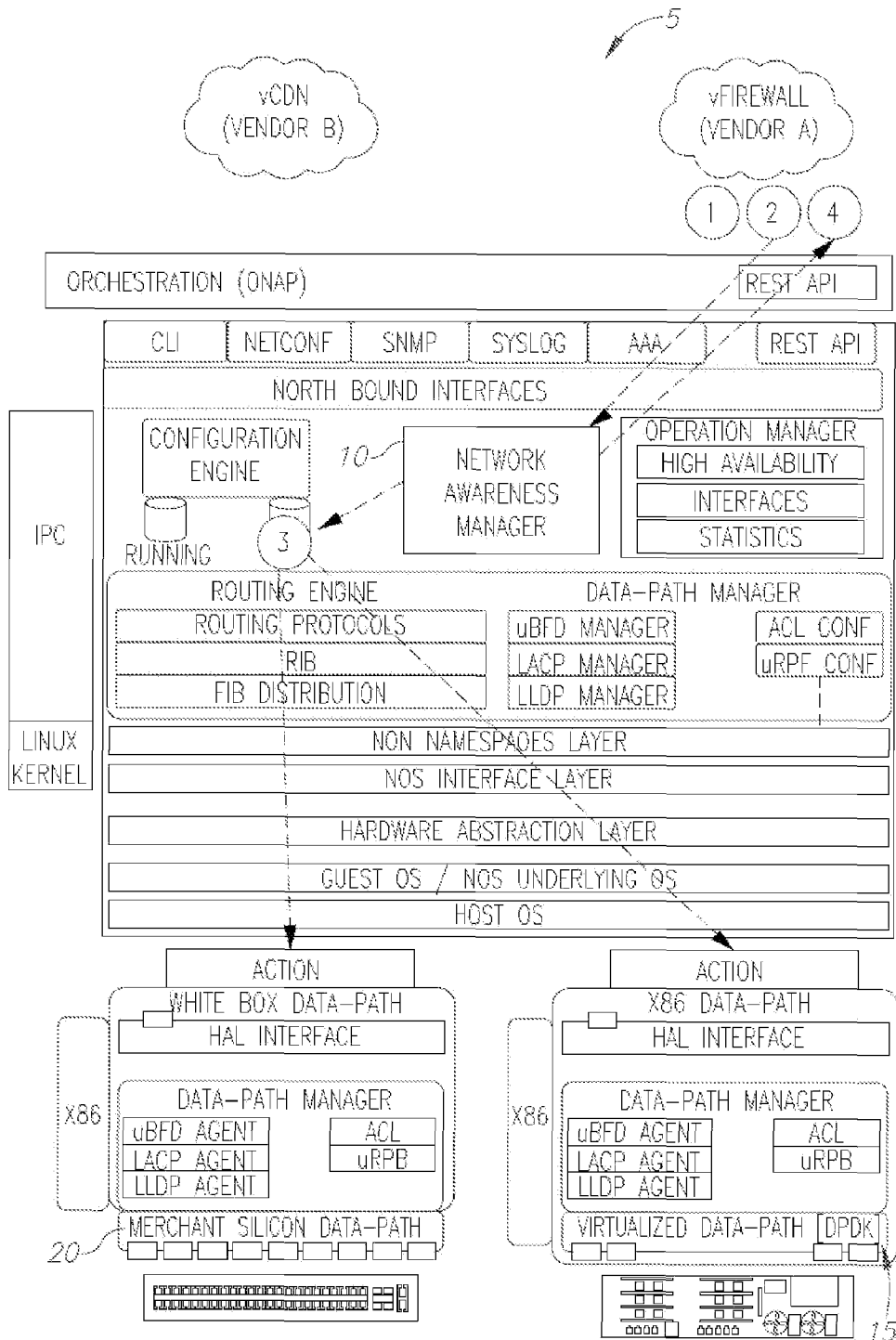


FIG. 1

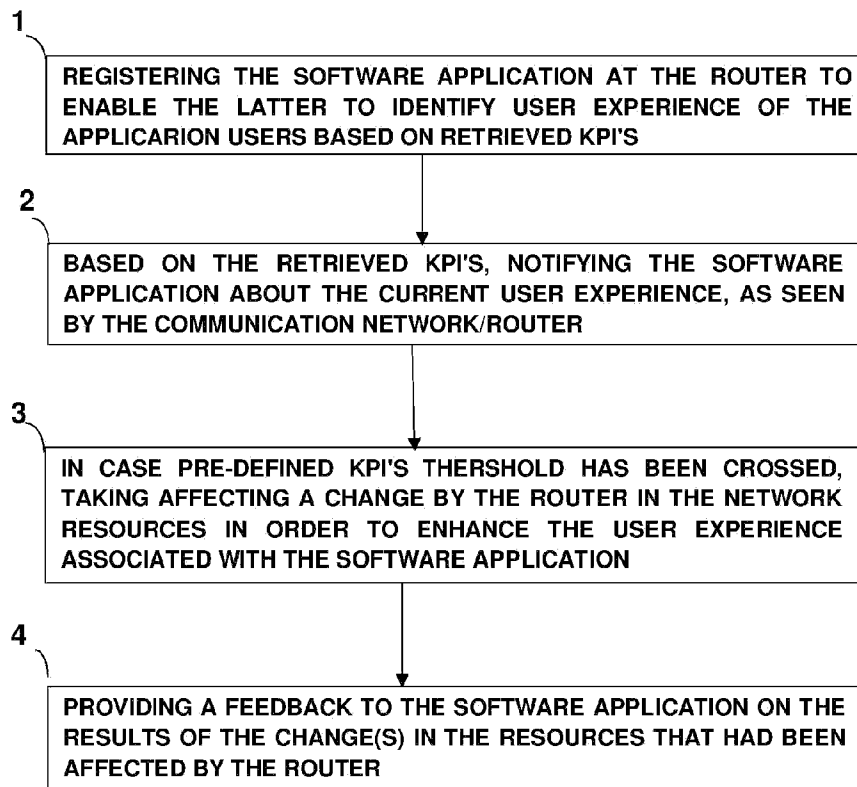


FIG. 2