



(19)中華民國智慧財產局

(12)發明說明書公告本

(11)證書號數：TW I848291 B

(45)公告日：中華民國 113 (2024) 年 07 月 11 日

(21)申請案號：111118350

(22)申請日：中華民國 111 (2022) 年 05 月 17 日

(51)Int. Cl. : H04W4/08 (2009.01)

H04W4/02 (2018.01)

G06F9/50 (2006.01)

(30)優先權：2021/05/31

中國大陸

202110602003.2

(71)申請人：大陸商華為技術有限公司(中國大陸) HUAWEI TECHNOLOGIES CO., LTD. (CN)  
中國大陸

(72)發明人：彭程暉 PENG, CHENGHUI (CN)；劉哲 LIU, ZHE (CN)；吳建軍 WU, JIANJUN (CN)

(74)代理人：陳寧樺；陳軍宇

(56)參考文獻：

CN 111869303A

CN 112188547A

審查人員：莊榮昌

申請專利範圍項數：30 項 圖式數：13 共 90 頁

(54)名稱

算力資源調度的方法以及相關裝置

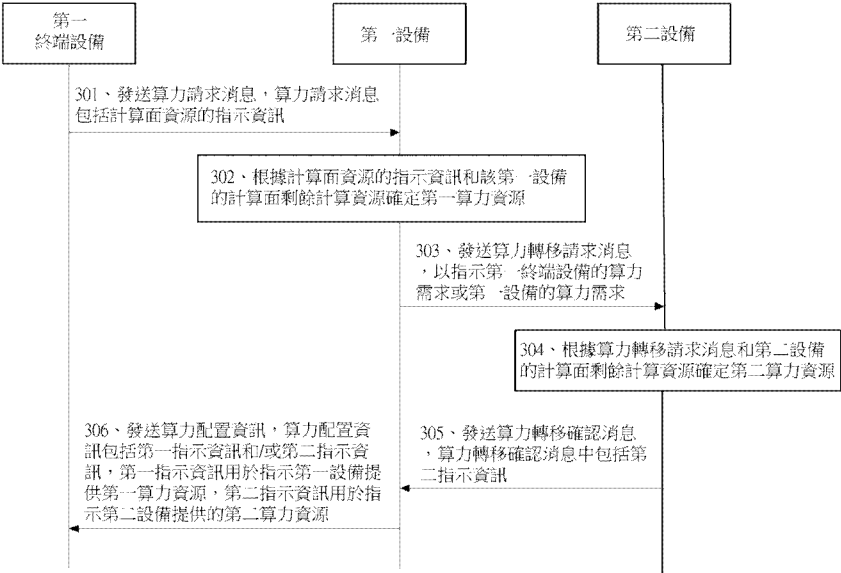
(57)摘要

本申請公開了一種算力資源調度的方法以及相關裝置。前述的方法包括向第一設備發送算力請求消息，算力請求消息包括計算面資源的指示資訊，計算面資源的指示資訊用於指示第一終端設備的算力需求；接收第一設備發送的算力配置資訊，算力配置資訊包括第一指示資訊和/或第二指示資訊，第一指示資訊用於指示第一設備提供的第一算力資源，第二指示資訊用於指示第二設備提供的第二算力資源，第一設備與第二設備不相同。

This application discloses a computing resource scheduling method and apparatus. The method comprises sending a computing resource request message to the first device, which comprises indication information of computing plane resource, wherein the computing plane resource is used to indicate computing resource requirement of the first terminal device. The method further includes receiving computing configuration information sent by the first device. The computing configuration information comprises a first indication and/or a second indication, wherein the first indication is configured to indicate a first computing resource provided by the first device, the second indication is configured to indicate a second computing resource provided by the second device. The first device and the second device are different.

指定代表圖：

符號簡單說明：  
301~306:步驟



【圖3】



I848291

## 【發明摘要】

【中文發明名稱】算力資源調度的方法以及相關裝置

【英文發明名稱】 COMPUTING RESOURCE SCHEDULING  
METHOD AND APPARATUS

【中文】本申請公開了一種算力資源調度的方法以及相關裝置。前述的方法包括向第一設備發送算力請求消息，算力請求消息包括計算面資源的指示資訊，計算面資源的指示資訊用於指示第一終端設備的算力需求；接收第一設備發送的算力配置資訊，算力配置資訊包括第一指示資訊和/或第二指示資訊，第一指示資訊用於指示第一設備提供的第一算力資源，第二指示資訊用於指示第二設備提供的第二算力資源，第一設備與第二設備不相同。

【英文】 This application discloses a computing resource scheduling method and apparatus. The method comprises sending a computing resource request message to the first device, which comprises indication information of computing plane resource, wherein the computing plane resource is used to indicate computing resource requirement of the first terminal device. The method further includes receiving computing configuration information sent by the first device. The computing configuration information comprises a first indication and/or a second indication, wherein the first indication is configured to indicate a first computing resource provided by the first

device, the second indication is configured to indicate a second computing resource provided by the second device. The first device and the second device are different.

【指定代表圖】圖3。

【代表圖之符號簡單說明】

301～306：步驟

【特徵化學式】

無

## 【發明說明書】

【中文發明名稱】 算力資源調度的方法以及相關裝置

【英文發明名稱】 COMPUTING RESOURCE SCHEDULING  
METHOD AND APPARATUS

### 【技術領域】

【0001】 本申請實施例涉及通訊技術領域，具體涉及一種算力資源調度的方法以及相關裝置。

### 【先前技術】

【0002】 現有的無線網路主要提供連接服務，無線網路架構主要解決的是終端和終端之間、或者終端和伺服器之間的連接服務問題。而且隨著未來無線網路面對即時性人工智慧（artificial intelligence，AI）、高資料隱私安全保護等新場景、新需求的出現，無線網路架構需要內生的支援各設備之間直接的 AI 協作機制。

【0003】 然而，在現有的無線網路架構中，相應的網路資料分析功能（network data analytics function，NWDAF）網路元素主要從終端、NF 網路元素、基地台等收集資料，通過資料分析來學習模型，並為 NF 網路元素、基地台等提供資料分析等 AI 優化服務。但是，這樣的系統架構沒有充分利用終端和基地台等各網路功能的自身 AI 能力，無法很好的支撐無線網路中網路設備、終端設備等設備之間的 AI 協作，面臨著低延時 AI 服務無法得到快速且及時地回

應；而且，NWDAF 網路元素從終端或其他網路功能收集資料，可能面臨資料隱私性的問題。

【0004】 因此，在突發性的需要大量算力的學習、增量學習、推理場景等 AI 場景下，基於現有的 NWDAF 網路元素在終端設備的算力資源不足時，無法很好的支撐無線網路中網路設備、終端設備等設備之間的 AI 協作。

#### 【發明內容】

【0005】 本申請實施例提供了一種算力資源調度的方法以及相關裝置，旨在第一終端設備的算力資源不足以支撐與其他網路設備之間的 AI 協作時，通過向第一設備請求相應的算力資源，進而使得第一終端設備能夠有充足的算力資源支撐與其他網路設備之間的 AI 協作。

【0006】 第一方面，本申請實施例中提供了一種算力資源調度的方法，該方法可以應用於第一終端設備，該方法可以包括：向第一設備發送算力請求消息，所述算力請求消息包括計算面資源的指示資訊，所述計算面資源的指示資訊用於指示所述第一終端設備的算力需求；接收所述第一設備發送的算力配置資訊，所述算力配置資訊包括第一指示資訊和/或第二指示資訊，所述第一指示資訊用於指示所述第一設備提供第一算力資源，所述第二指示資訊用於指示第二設備提供的第二算力資源，所述第一設備與所述第二設備不相同。

【0007】 通過上述方式，在第一終端設備的算力資源不足以支撐與其他網路設備之間的 AI 協作時，可以向第一設備發送算力請求消息，進而由第一設備根據該算力請求消息中的計算面資源的指示資訊和該第一設備的計算面剩餘計算資源確定第一算力資源。然後由第一設備向第一終端設備發送相應的算力配置資訊，該算力配置資訊中包括第一指示資訊和/或第二指示資訊。這樣，第一終端設備便能夠根據第一指示資訊從第一設備得到第一算力資源的支援、和/或根據第二指示資訊從第二設備得到第二算力資源的支援，進而支撐與其他網路設備之間的 AI 協作。

【0008】 可選的，在一些實施例中，所述第一設備為第一基地台，向所述第一設備發送算力請求消息，可以包括：通過廣播 RRC 信令向所述第一設備發送算力請求消息。

【0009】 可選的，在另一些實施例中，所述第一設備為核心網設備，向所述第一設備發送算力請求消息，可以包括：通過 NAS 信令向所述第一設備發送算力請求消息。

【0010】 可選的，在另一些實施例中，所述第一設備為第一基地台，接收所述第一設備發送的算力配置資訊，可以包括：接收所述第一設備通過廣播 RRC 信令發送的算力配置資訊。

【0011】 可選的，在另一些實施例中，所述第一設備為核心網設備，接收所述第一設備發送的算力配置資訊，可以包括：接收所述第一設備通過 NAS 信令發送的算力配置資訊。

【0012】 可選的，在一些實施例中，所述計算面資源的指示資訊包

括算力大小、執行時間、算力任務類型、計算模型、計算演算法、執行結果傳輸觸發類型和觸發類型中的一個或多個。

【0013】 第二方面，本申請實施例提供了另一種算力資源調度的方法，該方法可以應用於第一設備，該第一設備可以包括第一基地台、或核心網設備，該方法可以包括：接收第一終端設備發送的算力請求消息，所述算力請求消息包括計算面資源的指示資訊，所述計算面資源的指示資訊用於指示所述第一終端設備的算力需求；根據所述計算面資源的指示資訊和所述第一設備的計算面剩餘計算資源確定第一算力資源；向所述第一終端設備發送算力配置資訊，所述算力配置資訊包括所述第一指示資訊和/或第二指示資訊，所述第一指示資訊用於指示所述第一設備提供所述第一算力資源，所述第二指示資訊用於指示第二設備提供的第二算力資源，所述第一設備與所述第二設備不相同。

【0014】 可選的，在一些實施例中，該方法還可以包括：向第二設備發送算力轉移請求消息，以指示所述第一終端設備的算力需求或所述第一設備的算力需求；接收所述第二設備發送的算力轉移確認消息，所述算力轉移確認消息中包括所述第二指示資訊。

【0015】 可選的，在一些實施例中，所述第一設備包括第一基地台，所述第二設備包括第二基地台；向第二設備發送算力轉移請求消息，包括：通過 X2 口信令向所述第二設備發送算力轉移請求消息。

【0016】 可選的，在一些實施例中，所述第一設備包括第一基地台，所述第二設備包括第二基地台；接收所述第二設備發送的算力轉

移確認消息，包括：接收所述第二設備通過 X2 口信令發送的算力轉移確認消息。

【0017】 可選的，在一些實施例中，所述第一設備包括第一基地台，所述第二設備包括核心網設備；向第二設備發送算力轉移請求消息，包括：通過 N2 口信令向所述第二設備發送算力轉移請求消息。

【0018】 可選的，在一些實施例中，所述第一設備包括第一基地台，所述第二設備包括核心網設備；接收所述第二設備發送的算力轉移確認消息，可以包括：接收所述第二設備通過 N2 口信令發送的算力轉移確認消息。

【0019】 可選的，在一些實施例中，所述第一設備為所述第一基地台，所述接收第一終端設備發送的算力請求消息，包括：接收所述第一終端設備通過廣播 RRC 信令發送的算力請求消息。

【0020】 可選的，在一些實施例中，所述第一設備為所述第一基地台，向所述第一終端設備發送算力配置資訊，包括：通過廣播 RRC 信令向所述第一終端設備發送算力配置資訊。

【0021】 可選的，在一些實施例中，所述第一設備為核心網設備，所述第二設備包括第一基地台；所述接收所述第一終端設備發送的算力請求消息，包括：接收所述第一終端設備通過 NAS 信令發送的算力請求消息。

【0022】 可選的，在一些實施例中，所述第一設備為核心網設備，所述第二設備包括第一基地台；向所述第一終端設備發送算力配置資訊，包括：通過 NAS 信令向所述第一終端設備發送算力配置

資訊。

【0023】 可選的，在一些實施例中，所述第一設備包括第一基地台，所述第二設備包括第二終端設備；向第二設備發送算力轉移請求消息，包括：通過廣播 RRC 信令向所述第二設備發送算力轉移請求消息。

【0024】 可選的，在一些實施例中，所述第一設備包括第一基地台，所述第二設備包括第二終端設備；接收所述第二設備發送的算力轉移確認消息，可以包括：接收所述第二設備通過廣播 RRC 信令發送的算力轉移確認消息。

【0025】 可選的，在一些實施例中，所述計算面資源的指示資訊包括算力大小、執行時間、算力任務類型、計算模型、計算演算法、執行結果傳輸觸發類型和觸發類型中的一個或多個。

【0026】 協力廠商面，本申請實施例提供了另一種算力資源調度的方法，該方法可以應用於第二設備，該方法可以包括：接收第一設備發送的算力轉移請求消息；根據所述算力轉移請求消息和所述第二設備的計算面剩餘計算資源確定第二算力資源；向所述第一設備發送算力轉移確認消息，所述算力轉移確認消息中包括所述第二指示資訊，所述第二指示資訊用於指示所述第二設備提供的所述第二算力資源。

【0027】 可選的，在一些實施例中，所述第一設備包括第一基地台，所述第二設備包括第二基地台；接收第一設備發送的算力轉移請求消息，包括：接收所述第一設備通過 X2 口信令發送的算力轉移

請求消息。

【0028】 可選的，在一些實施例中，所述第一設備包括第一基地台，所述第二設備包括第二基地台；向所述第一設備發送算力轉移確認消息，包括：通過 X2 口信令向所述第一設備發送算力轉移確認消息。

【0029】 可選的，在一些實施例中，所述第一設備包括第一基地台，所述第二設備包括核心網設備；接收所述第一設備發送的算力轉移請求消息，包括：接收所述第一設備通過 N2 口信令發送的算力轉移請求消息。

【0030】 可選的，在一些實施例中，所述第一設備包括第一基地台，所述第二設備包括核心網設備；向所述第一設備發送算力轉移確認消息，包括：通過 N2 口信令向所述第一設備發送算力轉移確認消息。

【0031】 可選的，在一些實施例中，所述第一設備包括第一基地台，所述第二設備包括第二終端設備；接收所述第一設備發送的算力轉移請求消息，包括：接收所述第一設備通過廣播 RRC 信令發送的算力轉移請求消息。

【0032】 可選的，在一些實施例中，所述第一設備包括第一基地台，所述第二設備包括第二終端設備；向所述第一設備發送算力轉移確認消息，可以包括：通過廣播 RRC 信令向所述第一設備發送算力轉移確認消息。

【0033】 第四方面，本申請實施例提供一種第一終端設備，該所述

第一終端設備可以包括：發送單元，用於向第一設備發送算力請求消息，所述算力請求消息包括計算面資源的指示資訊，所述計算面資源的指示資訊用於指示所述第一終端設備的算力需求；接收單元，用於接收所述第一設備發送的算力配置資訊，所述算力配置資訊包括第一指示資訊和/或第二指示資訊，所述所述第一指示資訊用於指示所述第一設備提供第一算力資源，所述第二指示資訊用於指示第二設備提供的第二算力資源，所述第一設備與所述第二設備不相同。

【0034】 可選的，在一些實施例中，所述第一設備為第一基地台；所述發送單元，用於通過廣播 RRC 信令向所述第一設備發送算力請求消息。

【0035】 可選的，在一些實施例中，所述第一設備為第一基地台；所述接收單元，用於接收所述第一設備通過廣播 RRC 信令發送的算力配置資訊。

【0036】 可選的，在一些實施例中，所述第一設備為核心網設備；所述發送單元，用於通過 NAS 信令向所述第一設備發送算力請求消息。

【0037】 可選的，在一些實施例中，所述第一設備為核心網設備；所述接收單元，用於接收所述第一設備通過 NAS 信令發送的算力配置資訊。

【0038】 可選的，在一些實施例中，所述計算面資源的指示資訊包括算力大小、執行時間、算力任務類型、計算模型、計算演算法、

執行結果傳輸觸發類型和觸發類型中的一個或多個。

【0039】 第五方面，本申請實施例提供一種第一設備，該第一設備可以包括：獲取模組，用於接收第一終端設備發送的算力請求消息，所述算力請求消息包括計算面資源的指示資訊，所述計算面資源的指示資訊用於指示所述第一終端設備的算力需求；處理模組，用於根據所述計算面資源的指示資訊和所述第一設備的計算面剩餘計算資源確定第一算力資源；發送模組，用於向所述第一終端設備發送算力配置資訊，所述算力配置資訊包括所述第一指示資訊和/或第二指示資訊，所述第一指示資訊用於指示所述第一設備提供第一算力資源，所述第二指示資訊用於指示第二設備提供的第二算力資源，所述第一設備與所述第二設備不相同。

【0040】 可選的，在一些實施例中，所述發送模組，用於向第二設備發送算力轉移請求消息，以指示所述第一終端設備的算力需求或所述第一設備的算力需求；所述獲取模組，用於接收所述第二設備發送的算力轉移確認消息，所述算力轉移確認消息中包括所述第二指示資訊。

【0041】 可選的，在一些實施例中，所述第一設備包括第一基地台，所述第二設備包括第二基地台；所述發送模組，用於通過 X2 口信令向所述第二設備發送算力轉移請求消息。

【0042】 可選的，在一些實施例中，所述第一設備包括第一基地台，所述第二設備包括第二基地台；所述獲取模組，用於接收所述第二設備通過 X2 口信令發送的算力轉移確認消息。

【0043】 可選的，在一些實施例中，所述第一設備包括第一基地台，所述第二設備包括核心網設備；所述發送模組，用於通過 N2 口信令向所述第二設備發送算力轉移請求消息。

【0044】 可選的，在一些實施例中，所述第一設備包括第一基地台，所述第二設備包括核心網設備；所述獲取模組，用於接收所述第二設備通過 N2 口信令發送的算力轉移確認消息。

【0045】 可選的，在一些實施例中，所述第一設備為所述第一基地台；所述獲取模組，用於接收所述第一終端設備通過廣播 RRC 信令發送的算力請求消息。

【0046】 可選的，在一些實施例中，所述第一設備為所述第一基地台；所述發送模組，用於通過廣播 RRC 信令向所述第一終端設備發送算力配置資訊。

【0047】 可選的，在一些實施例中，所述第一設備為核心網設備，所述第二設備包括第一基地台和/或第二基地台；所述獲取模組，用於接收所述第一終端設備通過 NAS 信令發送的算力請求消息。

【0048】 可選的，在一些實施例中，所述第一設備為核心網設備，所述第二設備包括第一基地台；所述發送模組，用於通過 NAS 信令向所述第一終端設備發送算力配置資訊。

【0049】 可選的，在一些實施例中，所述第一設備包括第一基地台，所述第二設備包括第二終端設備；所述發送模組，用於通過廣播 RRC 信令向所述第二設備發送算力轉移請求消息。

【0050】 可選的，在一些實施例中，所述第一設備包括第一基地台，

所述第二設備包括第二終端設備；所述獲取模組，用於接收所述第二設備通過廣播 RRC 信令發送的算力轉移確認消息。

【0051】 可選的，在一些實施例中，所述計算面資源的指示資訊包括算力大小、執行時間、算力任務類型、計算模型、計算演算法、執行結果傳輸觸發類型和觸發類型中的一個或多個。

【0052】 第六方面，本申請實施例提供一種第二設備，該第二設備可以包括：第一接收模組，用於接收第一設備發送的算力轉移請求消息；第一處理模組，用於根據所述算力轉移請求消息和所述第二設備的計算面剩餘計算資源確定第二算力資源；第一發送模組，用於向所述第一設備發送算力轉移確認消息，所述算力轉移確認消息中包括所述第二指示資訊，所述第二指示資訊用於指示所述第二設備提供的所述第二算力資源。

【0053】 可選的，在一些實施例中，所述第一接收模組，用於接收所述第一設備通過 X2 口信令發送的算力轉移請求消息。

【0054】 可選的，在一些實施例中，所述第一設備包括第一基地台，所述第二設備包括第二基地台；所述第一發送模組，用於通過 X2 口信令向所述第一設備發送算力轉移確認消息。

【0055】 可選的，在一些實施例中，所述第一接收模組，用於接收所述第一設備通過 N2 口信令發送的算力轉移請求消息。

【0056】 可選的，在一些實施例中，所述第一設備包括第一基地台，所述第二設備包括核心網設備；所述第一發送模組，用於通過 N2 口信令向所述第一設備發送算力轉移確認消息。

【0057】 可選的，在一些實施例中，所述第一設備包括第一基地台，所述第二設備包括第二終端設備；所述第一接收模組，用於接收所述第一設備通過廣播 RRC 信令發送的算力轉移請求消息。

【0058】 可選的，在一些實施例中，所述第一設備包括第一基地台，所述第二設備包括第二終端設備；第一發送模組，用於通過廣播 RRC 信令向所述第一設備發送算力轉移確認消息。

【0059】 第七方面，本申請實施例提供了一種第一終端設備，可以包括：記憶體，用於儲存電腦可讀指令。還可以包括，與記憶體耦合的處理器，用於執行記憶體中的電腦可讀指令從而執行如第一方面或者第一方面任意一種可能的實施方式中所描述的方法。

【0060】 第八方面，本申請實施例提供了一種第一設備，可以包括：記憶體，用於儲存電腦可讀指令。還可以包括，與記憶體耦合的處理器，用於執行記憶體中的電腦可讀指令從而執行如第二方面或者第二方面任意一種可能的實施方式中所描述的方法。

【0061】 第九方面，本申請實施例提供了一種第二設備，可以包括：記憶體，用於儲存電腦可讀指令。還可以包括，與記憶體耦合的處理器，用於執行記憶體中的電腦可讀指令從而執行如協力廠商面或者協力廠商面任意一種可能的實施方式中所描述的方法。

【0062】 第十方面，本申請實施例提供了一種電腦可讀儲存媒體，包括指令，當其在電腦上運行時，使得電腦執行如第一方面、第二方面、協力廠商面，或者第一方面任意一種、第二方面任意一種或協力廠商面任意一種可能實現方式的方法。

【0063】 第十一方面，本申請實施例提供了一種包含指令的電腦程式產品，當其在電腦上運行時，使得電腦執行如第一方面、第二方面、協力廠商面，或者第一方面任意一種、第二方面任意一種或協力廠商面任意一種可能實現方式的方法。

【0064】 本申請第十二方面提供一種晶片系統，該晶片系統可以包括處理器，用於支援第一終端設備實現上述第一方面、第一方面任意一種可能的實施方式中所描述的方法中所涉及的功能，或者支援第一設備實現上述第二方面，第二方面任意一種可能的實施方式中所描述的方法中所涉及的功能；或者支援第二設備實現上述協力廠商面，協力廠商面任意一種可能的實施方式中所描述的方法中所涉及的功能。

【0065】 可選地，結合上述第十二方面，在第一種可能的實施方式中，晶片系統還可以包括記憶體，記憶體，用於保存第一終端設備、第一設備、第二設備必要的程式指令和資料。該晶片系統，可以由晶片構成，也可以包含晶片和其他分立器件。其中，晶片系統可以包括專用積體電路（application specific integrated circuit，ASIC）、現成可程式設計閘陣列（field programmable gate array，FPGA）或者其他可程式設計邏輯器件等。進一步，晶片系統還可以可以包括介面電路等。

【0066】 上述第四方面、第七方面、第十方面至第十二方面所獲得的技術效果與第一方面中對應的技術手段獲得的技術效果近似，在這裡不再贅述。

【0067】 上述第五方面、第八方面、第十方面至第十二方面所獲得的技術效果與第二方面中對應的技術手段獲得的技術效果近似，在這裡不再贅述。

【0068】 上述第六方面、第九方面、第十方面至第十二方面所獲得的技術效果與協力廠商面中對應的技術手段獲得的技術效果近似，在這裡不再贅述。

【0069】 從以上技術方案可以看出，本申請實施例具有以下優點：

【0070】 在本申請實施例中，在第一終端設備的算力資源不足以支撐與其他網路設備之間的 AI 協作時，可以向第一設備發送算力請求消息，進而由第一設備根據該算力請求消息中的計算面資源的指示資訊和該第一設備的計算面剩餘計算資源確定第一算力資源。然後由第一設備向第一終端設備發送相應的算力配置資訊，該算力配置資訊中包括第一指示資訊和/或第二指示資訊。這樣，第一終端設備便能夠根據第一指示資訊從第一設備得到第一算力資源的支援、和/或根據第二指示資訊從第二設備得到第二算力資源的支援，進而支撐與其他網路設備之間的 AI 協作。

#### 【圖式簡單說明】

【0071】 為了更清楚地說明本申請實施例的技術方案，下面將對實施例描述中所需要使用的附圖作簡單地介紹，顯而易見地，下面描述中的附圖僅僅是本申請的一些實施例。

圖 1a 為本實施例提供的一種系統架構示意圖。

圖 1b 為一種介面示意圖。

圖 2a 為本申請實施例提供的一種通訊系統架構示意圖。

圖 2b 為本申請實施例提供的另一種通訊架構示意圖。

圖 3 為本申請實施例中提供的一種算力資源調度的方法流程示意圖。

圖 4a 為本申請實施例中提供的另一種算力資源調度的方法流程示意圖。

圖 4b 為本申請實施例提供的一種算力轉移的示意圖。

圖 5a 為本申請實施例中提供的另一種算力資源調度的方法流程示意圖。

圖 5b 為本申請實施例提供的另一種算力轉移的示意圖。

圖 6a 為本申請實施例中提供的另一種算力資源調度的方法流程示意圖。

圖 6b 為本申請實施例提供的一種算力轉移的示意圖。

圖 7a 為本申請實施例中提供的另一種算力資源調度的方法流程示意圖。

圖 7b 為本申請實施例提供的另一種算力轉移的示意圖。

圖 8 為本申請實施例中提供的另一種算力資源調度的方法流程示意圖。

圖 9a 為本申請實施例中提供的另一種算力資源調度的方法流程示意圖。

圖 9b 為本申請實施例提供的另一種算力轉移的示意圖。

圖 10 為本申請實施例提供了一種第一終端設備的結構示意圖。

圖 11 為本申請實施例提供了一種第一設備的結構示意圖。

圖 12 為本申請實施例提供了一種第二設備的結構示意圖。

圖 13 為本申請實施例提供的通訊設備的硬體結構示意圖。

### 【實施方式】

【0072】 本申請實施例提供了一種算力資源調度的方法以及相關裝置，旨在第一終端設備的算力資源不足以支撐與其他網路設備之間的 AI 協作時，通過向第一設備和/或第二設備請求相應的算力資源，進而使得第一終端設備能夠有充足的算力資源支撐與其他網路設備之間的 AI 協作。

【0073】 下面將結合本申請實施例中的附圖，對本申請實施例中的技術方案進行清楚、完整地描述，顯然，所描述的實施例僅是本申請一部分實施例，而不是全部的實施例。基於本申請中的實施例，本領域普通技術人員在沒有作出創造性勞動前提下所獲得的所有其他實施例，都屬於本申請保護的範圍。

【0074】 本申請的說明書和申請專利範圍及上述附圖中的術語「第一」、「第二」、「第三」、「第四」等（如果存在）是用於區別類似的物件，而不必用於描述特定的順序或先後次序。應該理解這樣使用的資料在適當情況下可以互換，以便這裡描述的本申請的實施例如能夠以除了在這裡圖示或描述的那些以外的順序實

施。此外，術語「包括」和「具有」以及他們的任何變形，意圖在於覆蓋不排除他的包含。在本申請中，「至少一個」是指一個或者多個，「多個」是指兩個或兩個以上。「和/或」，描述關聯物件的關聯關係，表示可以存在三種關係，例如，A 和/或 B，可以表示：單獨存在 A，同時存在 A 和 B，單獨存在 B 的情況，其中 A，B 可以是單數或者複數。字元「/」一般表示前後關聯物件是一種「或」的關係。「以下至少一項（個）」或其類似表達，是指的這些項中的任意組合，包括單項（個）或複數項（個）中的任意組合。例如，a，b 或 c 中的至少一項（個），可以表示：a，b，c、a 和 b、a 和 c、b 和 c 或者 a 和 b 和 c，其中 a、b 和 c 可以是單個，也可以是多個。值得注意的是，「至少一項（個）」還可以解釋成「一項（個）或多項（個）」。

**【0075】** 現有的無線網路主要提供連接服務，無線網路架構主要解決的是終端和終端之間、或者終端和伺服器之間的連接服務問題。而且隨著未來無線網路面對即時性 AI、高資料隱私安全保護等新場景、新需求的出現，無線網路架構需要內生的支援各設備之間直接的 AI 協作機制。具體可以參閱圖 1a，為本實施例提供的一種系統架構示意圖。如圖 1a 所示，該系統架構為 5G 網路的系統架構示意圖，該系統架構不但支援 3GPP 標準組定義的無線技術存取核心網路側，而且支援支援 non-3GPP 存取技術通過 non-3GPP 轉換功能（non-3GPP interworking function，N3IWF）或下一代存取閘道（next generation packet data gateway，ngPDG）或固網存取

開道或可信 non-3GPP 存取開道存取核心網路側。

【0076】 圖 1a 中的使用者設備（user equipment，UE）、（無線）存取網路（(radio) access network，(R)AN）、用戶面功能（user plane function，UPF）網路元素和資料網路（data network，DN）一般被稱為使用者層網路功能或實體，使用者的資料流程量可以通過 UE 和 DN 之間建立的資料傳輸通道進行傳輸。

【0077】 UE 可以包括：手持終端、車載設備、可穿戴設備、筆記型電腦、使用者單元（subscriber unit）、行動電話（cellular phone）、智慧型電話（smart phone）、無線資料卡、個人數位助理（personal digital assistant，PDA）電腦、平板型電腦、無線數據機（modem）、手持設備（handheld）、桌上型電腦（laptop computer）、無線電話（cordless phone）或者無線區域環路（wireless local loop，WLL）台、機器類型通訊（machine type communication，MTC）終端、使用者駐地設備（customer-premises equipment，CPE）或是其他可以存取網路的設備。UE 與存取網路的設備之間採用某種空中接口技術相互通訊。

【0078】 存取網路（access network，AN），在具體應用中又可稱為無線存取網路（radio access network，RAN），RAN 由存取網路設備組成，負責使用者設備的存取。5G 網路的 RAN 設備可以是下一代（next generation，NG）RAN 設備，也可以是演進型通用陸地無線存取網路（evolved universal terrestrial radio access network，E-UTRAN）設備，5G 網路可同時連接至上述兩種存取網路設備。

RAN 在本申請實施例中可以是下一代基地台（next generation NodeB，gNB）或者下一代演進型基地台（next generation-evolved NodeB，ng-eNB）。其中，gNB 為 UE 提供新廣播（new radio, NR）的用戶面功能和控制面功能，ng-eNB 為 UE 提供演進型通用陸地無線存取（evolved universal terrestrial radio access, E-UTRA）的用戶面功能和控制面功能，需要說明的是，gNB 和 ng-eNB 僅是一種名稱，用於表示支援 5G 網路系統的基地台，並不具有限制意義。

【0079】資料網路（data network，DN）可以為運營商外部網路，也可以為運營商控制的網路，用於向使用者提供業務服務。UE 可通過存取運營商網路來訪問 DN，使用 DN 上的運營商或協力廠商提供的業務。

【0080】圖 1a 中其他的網路元素，則被稱為控制層網路功能或實體，主要負責使用者註冊認證、移動性管理、向用戶面功能（user plane function，UPF）網路元素下發資料包轉發策略、或 QoS 控制策略等，其中，會話管理功能（session management function，SMF）網路元素，主要用於負責用戶面網路元素選擇、用戶面網路元素重定向、互聯網協定（internet protocol，IP）地址分配、承載的建立、修改和釋放等；存取和移動性管理功能（access and mobility management function，AMF）網路元素，主要負責信令處理部分，例如存取控制、移動性管理、附著與去附著以及網路元素選擇等功能。

【0081】具體的，如圖 1a 所示，使用者設備 UE 可以通過 RAN 與

AMF 網路元素連接，AMF 網路元素分別與 SMF 網路元素、統一資料管理功能（unified data management，UDM）等網路元素連接，SMF 網路元素與 UPF 網路元素連接，UPF 網路元素與 DN 連接。系統架構中的介面和連接可以包括：N1、N2、N3、N4 和 N6。其中，連接 N1 為 UE 和 AMF 網路元素之間的控制面連接，用於傳輸使用者設備和核心網控制面之間的控制信令，具體的 N1 連接中的消息可以由 UE 和 RAN 之間的連接、RAN 和 AMF 網路元素之間的 N2 連接進行傳輸。N2 為 RAN 和 AMF 網路元素之間的控制面連接。N3 為 RAN 和 UPF 網路元素之間的連接。N4 為 SMF 網路元素和 UPF 網路元素之間的連接，用於傳遞 SMF 網路元素和 UPF 網路元素之間的控制信令。N6 為 UPF 網路元素和 DN 之間的連接。

【0082】此外，控制層網路功能或實體還可以包括策略控制功能（policy control function，PCF）網路元素，主要支援提供統一的策略框架來控制網路行為，提供策略規則給控制層網路功能，同時負責獲取與策略決策相關的使用者簽約資訊。應用功能（Application Function，AF）網路元素：主要支援與第三代合作夥伴計畫（the 3rd generation partnership project，3GPP）核心網交互來提供服務，例如影響資料路由決策，策略控制功能或者向網路側提供協力廠商的一些服務。網路切片選擇功能（network slice selection function，NSSF）網路元素，主要用於進行網路切片選擇。AUSF（authentication server function）網路元素：主要提供認證和

鑒權功能。統一資料管理（**unified data management**，UDM），可以用於進行位置管理和訂閱管理。網路暴露功能（**network exposure function**，NEF）網路元素：主要支援 3GPP 網路和協力廠商應用安全的交互，NEF 網路元素能夠安全的向協力廠商暴露網路能力和事件，用於加強或者改善應用服務品質，3GPP 網路同樣可以安全的從協力廠商獲取相關資料，用以增強網路的智慧決策；同時該網路元素支援從統一資料庫恢復結構化資料或者向統一資料庫中儲存結構化資料。

【0083】 值得理解的是，雖然上述圖 1a 所示的系統架構中並未體現 NWDAF 網路元素，但可以理解為 NWDAF 網路元素可以與該圖 1a 所示出的所有的功能網路元素之間均有直接的介面，具體體現為 Nnf 介面和 Nnwdaf 介面，如圖 1b 所示的介面示意圖。從圖 1b 中的（a）部分可以看出 NWDAF 網路元素與其他的功能網路元素之間通過 Nnf 介面連接，該 Nnf 介面主要用於 NWDAF 網路元素向其他的功能網路元素（例如：AMF 網路元素、SMF 網路元素等）請求對特定上下文的資料傳輸的訂閱、取消對資料傳遞的訂閱、以及請求對特定上下文的資料的特定報告等；圖 1a 所示的系統架構還允許 NWDAF 網路元素通過調用 OAM（Operations, Administration and Maintenance）運維系統服務，並從 OAM 獲取管理資料。此外，從圖 1b 中的（b）部分可以看出 Nnwdaf 介面是 NWDAF 網路元素向其他功能網路元素提供服務，用於其他網路功能向 NWDAF 請求對特定上下文的網路分析傳遞的訂閱、取消對

網路分析傳遞的訂閱、以及請求對特定上下文的網路分析的特定報告。

【0084】然而，在現有的無線網路架構中，相應的 NWDAF 網路元素主要從終端、其他的功能網路元素、基地台等收集資料，通過資料分析來學習模型，並為其他的功能網路元素、基地台等提供資料分析等 AI 優化服務。但是，這樣的系統架構沒有充分利用終端和基地台等各網路功能的自身 AI 能力，無法很好的支撐無線網路中網路設備、第一終端設備等設備之間的 AI 協作，面臨著低延時 AI 服務無法得到快速且及時地回應；而且，NWDAF 網路元素從終端或其他網路功能收集資料，可能面臨資料隱私性的問題。因此，在突發性的需要大量算力的學習、增量學習、推理場景等 AI 場景下，基於現有的 NWDAF 網路元素在第一終端設備的算力資源不足時，無法很好的支撐無線網路中網路設備、第一終端設備等設備之間的 AI 協作。

【0085】為了解決上述所描述的問題，本申請實施例中提供了一種算力資源調度的方法，該算力資源調度的方法可以應用於 5G、衛星通訊等無線通訊系統中，旨在第一終端設備的算力資源不足時，可以通過向基地台、核心網設備等請求算力資源，進而較好地支撐無線網路中網路設備、第一終端設備等設備之間的 AI 協作。參閱圖 2a，示出了本申請實施例提供的一種通訊系統架構示意圖。從圖 2a 可以看出，該通訊系統架構通常由社區組成，每個社區包括一個基地台（base station，BS），以及至少一個終端設備（如終

端設備 1～終端設備 6)，基地台可以向至少一個終端設備提供通訊服務的基地台。因此，在第一終端設備的算力資源不足時，可以向提供通訊服務的基地台請求算力資源，進而實現第一終端設備能夠在算力資源充足的情況下，支撐無線網路中網路設備、第一終端設備等設備之間的 AI 協作。所描述的第一終端設備可以理解成至少一個終端設備中的一個，此處不做限定。

【0086】而且，隨著 5G 網路的商業部署和 5G 行業應用的興起，將算力資源引入無線網路已成為必然的趨勢。通過在基地台內部引入獨立的計算面，為基地台的其他部分功能（如：控制面、用戶面功能），以及基地台之外的協力廠商提供基礎的公共計算服務，使得在基地台內部署具有即時性需求的功能（如：AI 優化網路、感知能力等），進而使得 AI 行業應用、AR、感知類應用等）等成為不斷發展的應用。具體地，可以參閱圖 2b,為本申請實施例提供的另一種通訊架構示意圖。從圖 2b 可以看出，該通訊架構可以包括至少一個終端設備、至少一個第一基地台、至少一個第二基地台以及核心網設備。所描述的第一基地台可以理解成當前能夠直接為第一終端設備提供服務的基地台，第二基地台可以理解成當前沒有為第一終端設備直接提供服務的基地台。在該通訊架構中，可以通過在第一基地台、第二基地台、第二終端設備和/或核心網設備中增加獨立的計算面，進而提供獨立的 AI 算力資源，並與第一基地台、第二基地台的無線控制面和資料面有邏輯介面，實現為第一終端設備提供即時 AI 服務或擴展（xF）服務，也服務於第一

基地台、第二基地台本身的 AI 服務或感知服務。這樣，在第一終端設備的算力資源不足時，可以向提供通訊服務的第一基地台、核心網設備請求算力資源；也可以在第一基地台的算力資源不足時，通過第一基地台向第二基地台或第二終端設備請求算力資源，或者向核心網設備請求算力資源，進而實現第一終端設備能夠在算力資源充足的情況下，支撐無線網路中網路設備、第一終端設備等設備之間的 AI 協作。

【0087】 需說明，圖 2a 中所描述的基地台可以理解成圖 2b 中的第一基地台。所描述的第一基地台、第二基地台可以包含集中單元（central unit，CU）和分佈單元（distributed unit，DU）。其中，CU 和 DU 可以放置在不同的地方，例如：DU 拉遠，放置於高話務量的區域，CU 放置於中心機房；CU 和 DU 也可以放置在同一機房；CU 和 DU 也可以為一個機架下的不同部件。

【0088】 另外，所描述的第一基地台、第二基地台形態可以呈現出不同的形式，例如：宏基地台、微基地台（也稱為小站）、皮站（Pico，微微站）、中繼站、存取點等，在本申請實施例中不限定限定說明。在採用不同的無線存取技術的系統中，具備基地台功能的設備的名稱可能會有所不同，例如：在 LTE 系統中，稱為演進的節點 B（evolved NodeB，eNB 或 eNodeB）；在第三代（英文：3rd Generation，3G）系統中，稱為節點 B（Node B）。值得理解的是，而隨著未來無線通訊技術的演進，未來出現如 6G、7G 等，也有可能會出現新的基地台名稱，但基地台的不同名稱，並不影響本申請實施例在

基地台中的應用。

【0089】 前面所提及到的第一基地台的計算面、第二基地台的計算面，可以理解成是基於通用 CPU（如：X86 或 ARM 等）、AI 晶片、GPU 晶片或 FPGA 等構建的，對基地台內的其他邏輯功能或基地台外的協力廠商提供計算相關的任務服務。所描述的任務的供給是基於各種不同形式和細微性的計算資源虛擬切割，包括集群 Cluster、主機、虛擬機器（VM）、虛擬節點（POD，可包含多容器）、容器（Container）、或其他更細細微性的可運行一段邏輯代碼/函數的資源（如雲原生的無伺服器 serverless）等。

【0090】 前述圖 2a 和圖 2b 所提及的第一終端設備，具體可以參照前述圖 1a 所示的使用者設備進行理解，此處不做贅述說明。

【0091】 為了便於理解，請參閱圖 3，為本申請實施例中提供的一種算力資源調度的方法流程示意圖。從圖 3 可知，該算力資源調度的方法可以包括如下步驟：

【0092】 301、第一終端設備向第一設備發送算力請求消息，算力請求消息包括計算面資源的指示資訊。

【0093】 在該示例中，第一終端設備可以在確定出自身的計算面資源不足以支撐與其他網路設備之間的 AI 協作時，可以通過算力請求消息向第一設備請求相應的算力資源。需說明，該算力請求消息中包括計算面資源的指示資訊，而且該計算面資源的指示資訊用於指示出該第一終端設備的算力需求。這樣，第一設備在接收到該算力請求消息後，通過解析該算力請求消息得到計算面資源的

指示資訊後，便可以獲知需要向第一終端設備提供多少算力資源了。

【0094】 可選的，在另一些示例中，所描述的計算面資源的指示資訊包括算力大小、執行時間、算力任務類型、計算模型、計算演算法、執行結果傳輸觸發類型和觸發類型中的一個或多個。

【0095】 需說明，算力大小可以包括但不限於雜湊計算速度（hash/second）或者浮點計算速度（floating-point operations per second，FLOPS），此處不做限定說明。所描述的雜湊計算的速度可以包括但不限於 1kH/s、1MH/s、1GH/s、1TH/s、1PH/s、1EH/8；所描述的浮點計算速度包括但不限於 1k FLOPS、1M FLOPS、1G FLOPS、1T FLOPS、1P FLOPS、1E FLOPS，此處不做限定說明。需說明，上述所描述的  $k=10^3$ ， $M=10^6$ ， $G=10^9$ 、 $T=10^{12}$ 、 $P=10^{15}$ 、 $E=10^{18}$ 。

【0096】 上述的執行時間可以包括起始時間資訊和結束時間資訊。

【0097】 計算模型可以理解成協定預定義模型，通過模型標識來指示模型資訊、或者指示模型的下載位址資訊,如統一資源定位符（uniform resource locator，URL）、IP 位址資訊等等。

【0098】 計算演算法可以是監督學習、非監督學習、神經網路和/或強化學習等，此處不做限定說明。上述的監督學習可以包括以下一種或多種具體的演算法：支援向量機（support vector machine，SVM）、決策樹（decision tree）、樸素貝葉斯分類（naive Bayesian classification）、和 k-近鄰演算法（k-nearest neighbor, KNN）。上

述的非監督學習可以包括以下一種或多種具體的演算法：主成分分析（principal component analysis，PCA）、奇異值分解（singular value decomposition，SVD）、和 k-均值聚類。上述的強化學習可以包括以下一種或多種具體的演算法：Q-learning 演算法、狀態-動作-獎勵-狀態-動作（state-action-reward-state-action，SARSA）演算法、深度 Q 網路（Deep Q Network，DQN）演算法、深度確定性策略梯度（deep deterministic policy gradient，DDPG）演算法等等。

【0099】 所提及的執行結果傳輸觸發類型可以是指週期性地傳輸執行結果，如：1ms、2ms、5ms、10ms、20ms、50ms、100ms 等等；或者，也可以是指 RRM 事件的執行結果，如：A1、A2 等各種 RRM 事件，或者終端從 RRC 空閒態或 RRC 非啟動態變成 RRC 連接態。

【0100】 上述的觸發類型可以是指週期性地執行（如：1ms、2ms、5ms、10ms、20ms、50ms、100ms 等等），也可以是指 RRM 事件的執行（如：A1、A2 等各種 RRM 事件，或者終端從 RRC 連接態變成 RRC 空閒態或 RRC 非啟動態）。

【0101】 302、第一設備根據計算面資源的指示資訊和該第一設備的計算面剩餘計算資源確定第一算力資源。

【0102】 在該示例中，所描述的第一設備的計算面剩餘計算資源用於指示該第一設備剩餘的計算面資源。第一設備在接收到第一終端設備發送的算力請求消息後，也可以根據自身的計算面剩餘

計算資源來確定是否需要回應該第一終端設備的算力請求消息。

在第一設備確定回應該算力請求消息時，該第一設備便可以根據該計算面的指示資訊和計算面剩餘計算資源確定第一算力資源。

【0103】 需說明，在一些實施例中，第一設備也可以拒絕回應該第一終端設備的算力請求消息。那麼，這樣第一終端設備就需要從其他的網路設備、基地台等設備中獲取相應的算力資源的指示資訊，進而根據相應的算力資源的指示資訊，從其他的網路設備、基地台等設備得到相應的算力資源的支持。

【0104】 303、第一設備向第二設備發送算力轉移請求消息，以指示第一終端設備的算力需求或第一設備的算力需求。

【0105】 在該實例中，在該第一設備確認回應第一終端設備的算力請求消息時，既可以將所確定出的第一算力資源的指示資訊（後續稱為第一指示資訊）提供給第一終端設備，也可以通過向第二設備發送算力轉移請求消息，進而從其他的第二設備獲取第二算力資源的指示資訊（後續稱為第二指示資訊）。所描述的第一指示資訊可以用來指示第一設備提供的第一算力資源，第二指示資訊可以用來指示第二設備提供的第二算力資源。

【0106】 具體地，在另一些實施例中，在第一設備的計算面剩餘計算資源小於計算面資源的指示資訊中所請求的計算面資源時，第一設備向第二設備發送算力轉移請求消息，以指示第二設備確定第二算力資源。

【0107】 也就是理解成，如果第一設備自身剩餘的計算面計算資

源不足以滿足該第一終端設備所需求的計算面資源時，該第一設備可以向其他具有計算面資源的第二設備發送算力轉移請求消息，進而指示第二設備在確定出相應的第二算力資源後，能夠提供相應的第二指示資訊給第一終端設備。

【0108】 或者，在另一些示例中，在第一終端設備發生社區切換的過程中，第一設備也可以向第二設備發送算力轉移請求消息，指示第二設備確定相應的第二算力資源。此時，就需要第二設備直接向第一終端設備提供相應的第二指示資訊了。

【0109】 304、第二設備根據算力轉移請求消息和第二設備的計算面剩餘計算資源確定第二算力資源。

【0110】 在該示例中，第二設備在接收到第一設備發送的算力轉移請求消息後，便可以基於該算力轉移請求消息和第二設備自身的計算面剩餘計算資源確定第二算力資源了。

【0111】 305、第二設備發送算力轉移確認消息，算力轉移確認消息中包括第二指示資訊。

【0112】 該示例中，在第二設備確定出第二算力資源後，能夠將第二算力資源相應的第二指示資訊包含於算力轉移確認消息中，並將該算力轉移確認消息發送至第一終端設備或者第一設備。所描述的第二指示資訊用於指示第二設備提供的第二算力資源，即指示該第二設備能夠提供的第二算力資源的具體算力資源情況。

【0113】 306、第一設備向第一終端設備發送算力配置資訊，算力配置資訊包括第一指示資訊和/或第二指示資訊，第一指示資訊用

於指示第一設備提供第一算力資源，第二指示資訊用於指示第二設備提供的第二算力資源。

【0114】 在該示例中，在第一設備確定出第一算力資源，以及由第二設備提供相應的第二指示資訊後，便可以向第一終端設備發送算力配置資訊，即該算力配置資訊中包括第一指示資訊和/或第二指示資訊。這樣，第一終端設備便可以獲取到第一指示資訊和/或第二指示資訊，進而根據第一指示資訊得到第一設備中的第一算力資源的支援、和/或根據第二指示資訊得到第二設備中的第二算力資源的支援。

【0115】 需說明，前述所描述的第一指示資訊可以用來指示第一設備提供的第一算力資源，即指示該第一設備能夠提供的第一算力資源的具體算力資源情況。前述的第一算力資源可以是處理資料的過程中需要使用的計算資源，包括但不限於軟體資源和/或硬體資源，例如：CPU、GPU、I/O 等，也可以是一個用於影像處理的神經網路演算法模型資源等，也可以是記憶體等，此處不做限定說明。前述所描述的第二算力資源也可以參照上述的第一算力資源的內容進行理解，此處不做贅述。

【0116】 值得理解的是，上述所描述的第一設備可以是第一基地台、也可以是核心網設備，第二設備可以是第二基地台，也可以是核心網設備等。而第一終端設備可以向不同的提供設備，如：第一基地台、核心網設備、第二基地台等請求算力資源，具體可以參照後續的圖 4a～圖 8 所描述的實施例進行理解，此處先不做贅述。

【0117】 本申請實施例中，在第一終端設備的算力資源不足以支撐與其他網路設備之間的 AI 協作時，可以向第一設備發送算力請求消息，進而由第一設備根據該算力請求消息中的計算面資源的指示資訊和該第一設備的計算面剩餘計算資源確定第一算力資源。然後由第一設備向第一終端設備發送相應的算力配置資訊，該算力配置資訊中包括第一指示資訊和/或第二指示資訊。這樣，第一終端設備便能夠根據第一指示資訊得到第一設備中的第一算力資源的支援、和/或根據第二指示資訊得到第二設備中的第二算力資源的支援，進而支撐與其他網路設備之間的 AI 協作。而且，通過共用的算力資源，提升了網路收益，降低第一終端設備的能耗和算力資源的配置要求，可有效地提升算力資源的複用程度，達到節能減排的效果。

【0118】 為了進一步理解本申請實施例提供的算力資源調度的方法，下面將以第一設備包括第一基地台為例進行說明。具體可以參閱圖 4a，為本申請實施例提供的另一種算力資源調度的方法流程圖。如圖 4a 所示，該算力資源調度的方法可以包括：

【0119】 401、第一終端設備通過廣播 RRC 信令向第一基地台發送算力請求消息，算力請求消息包括計算面資源的指示資訊。

【0120】 在該示例中，第一終端設備可以在確定出自身的計算面資源不足以支撐與其他網路設備之間的 AI 協作時，可以通過廣播 RRC 信令，向第一基地台發送算力請求消息，請求第一基地台提供相應的算力資源。需說明，該算力請求消息中包括計算面資源的

指示資訊，而且該計算面資源的指示資訊用於指示出該第一終端設備的算力需求。這樣，第一基地台在接收到該算力請求消息後，通過解析該算力請求消息得到計算面資源的指示資訊後，便可以獲知需要向第一終端設備提供多少算力資源了。

【0121】 具體地，第一終端設備可以在無線資源控制（radio resource control, RRC）連接建立請求中，告知第一基地台建立 RRC 連接的原因，即在 Establishment Cause 信元中攜帶計算請求存取原因值，即攜帶上述的計算面資源的指示資訊等。或者，第一終端設備也可以在上行專用控制通道-消息類型（uplink-dedicated control channel-messageType, UL-DCCH-MessageType）中定義上行的算力請求消息，由該上行的算力請求消息攜帶計算面資源的指示資訊。

【0122】 另外，所描述的廣播可以理解成第一終端設備與基地台之間的交互介面。

【0123】 需說明，上述所描述的計算面資源的指示資訊，具體可以參照圖 3 中的步驟 301 中所描述的內容進行理解，此處不做贅述。

【0124】 402、第一基地台根據計算面資源的指示資訊和第一基地台的計算面剩餘計算資源確定第一算力資源。

【0125】 在該示例中，所描述的第一基地台的計算面剩餘計算資源用於指示該第一基地台剩餘的計算面資源。第一基地台在接收到第一終端設備發送的算力請求消息後，也可以根據自身的計算面剩餘計算資源來確定是否需要回應該第一終端設備的算力請求

消息。在第一基地台確定回應該算力請求消息時，該第一基地台便可以根據該計算面的指示資訊和計算面剩餘計算資源確定第一算力資源。

【0126】 需說明，在一些實施例中，第一基地台也可以拒絕回應該第一終端設備的算力請求消息。那麼，這樣第一終端設備就需要從其他的網路設備、基地台等提供設備中獲取算力資源。

【0127】 403、第一基地台向第一終端設備發送算力配置資訊，算力配置資訊包括第一指示資訊，第一指示資訊用於指示第一基地台提供第一算力資源。

【0128】 在該實例中，在該第一基地台確認回應第一終端設備的算力請求消息，如果該第一基地台自身的計算面剩餘計算資源大於或等於計算面資源的指示資訊中所請求的算力資源時，則第一基地台可以直接從自身的計算面剩餘計算資源中確定出第一算力資源。這樣，在第一基地台確定出第一算力資源後，便可以將該第一指示資訊包含於算力配置資訊中，隨後向第一終端設備發送算力配置資訊。這樣，第一終端設備便可以獲取到算力配置資訊，進而得到該算力配置資訊中所包括的第一指示資訊。

【0129】 具體地，第一基地台可以通過廣播 RRC 信令向第一終端設備發送算力建立消息，該算力建立消息中可以包括前述的算力配置資訊。所述算力建立消息還可以包含計算無線承載( computing radio bearers, CRB ) 配置資訊，該 CRB 配置資訊用於第一終端設備確定 CRB 資源，所描述的 CRB 資源用於第一終端設備與第一

基地台之間交互算力 offloading 時的資料登錄輸出、或者計算模型資料或計算演算法資料。

【0130】 可選的，上述的 CRB 中承載的資料或模型或演算法映射到分組資料融合協定（packet data convergence protocol，PDCP）層中計算類的協定資料單元（data protocol data unit，data PDU）。

【0131】 可選的，所述 CRB 是使用者面資料無線承載（user data radio bearers，DRB）或信令無線承載（signaling radio bearers，SRB）的子集。

【0132】 具體可以參閱圖 4b，為本申請實施例提供的一種算力轉移的示意圖。從圖 4b 可以看出，第一終端設備向第一基地台的控制面發送算力請求消息，在第一基地台的控制面與第一基地台的計算面中完成第一算力資源的分配和確認，進而由第一基地台的控制面將該第一指示資訊發送至該第一終端設備，以完成算力資源的請求。最終，第一終端設備將算力 offloading 到該第一基地台的計算面，通過該 CRB，該第一終端設備與第一基地台之間可以交互算力 offloading 時的資料登錄輸出、或者計算模型資料或計算演算法資料。

【0133】 需說明，前述的第一算力資源可以參照前述圖 3 中的步驟 306 所描述的內容進行理解，此處不做贅述。

【0134】 為了進一步理解本申請實施例提供的算力資源調度的方法，下面將以第一設備包括第一基地台、第二設備包括第二基地台為例進行說明。具體可以參閱圖 5a，為本申請實施例提供的另一

種算力資源調度的方法流程圖。如圖 5a 所示，該算力資源調度的方法可以包括：

【0135】 501、第一終端設備通過廣播 RRC 信令向第一基地台發送算力請求消息，算力請求消息包括計算面資源的指示資訊。

【0136】 需說明，本實施例中的步驟 501 具體可以參照上述圖 4a 中的步驟 401 所描述的內容進行理解，此處不做贅述。

【0137】 502、第一基地台根據計算面資源的指示資訊和計算面剩餘計算資源確定第一算力資源。

【0138】 需說明，本實施例中的步驟 501～步驟 502 具體可以參照上述圖 4a 中的步驟 401～步驟 402 所描述的內容進行理解，此處不做贅述。

【0139】 503、第一基地台通過 X2 口信令向第二基地台發送算力轉移請求消息，以指示第一終端設備的算力需求或第一基地台的算力需求。

【0140】 在該示例中，在第一基地台確定出自身的計算面剩餘計算資源無法滿足第一終端設備所請求的算力資源，或者第一終端設備需要切換至第二基地台所覆蓋的社區時，該第一基地台可以向第二基地台發送算力轉移請求消息，以指示該第二基地台提供第二算力資源，進而滿足第一終端設備的請求。具體地，第一基地台可以通過 X2 口信令向第二基地台發送算力轉移請求消息。

【0141】 需說明，所描述的算力轉移請求消息，也可以參照圖 3 中的步驟 301 中所描述的算力請求消息進行理解，此處不做贅述。

【0142】 504、第二基地台根據算力轉移請求消息和第二基地台的計算面剩餘計算資源確定第二算力資源。

【0143】 在該實例中，第二基地台接收到第一基地台的算力轉移請求消息後，可以根據該第二基地台的計算面剩餘計算資源和算力轉移請求消息確定第二算力資源。

【0144】 也就是理解成，如果第一基地台自身的計算面計算資源不足以滿足該第一終端設備所需求的算力資源時，該第一基地台可以向其他具有計算面資源的第二基地台發送算力轉移請求消息，進而指示第二基地台能夠提供相應的第二算力資源給第一終端設備。

【0145】 505、第二基地台向第一基地台發送算力轉移確認消息，算力轉移確認消息中包括第二指示資訊，第二指示資訊用於指示第二基地台提供的第二算力資源。

【0146】 該示例中，由於第一基地台是直接服務於第一終端設備的基地台，因此在第二基地台確定出第二算力資源後，可以該通過算力轉移確認消息的方式將第二指示資訊發送至第一基地台。具體地，該第二基地台也可以通過 X2 口信令向第一基地台發送算力轉移確認消息，進而將包含在算力轉移確認消息中的第二指示資訊發送至第一基地台。另外，所描述的 X2 口可以理解成基地台與基地台之間的交互介面。

【0147】 需說明，在另一些示例中，如果是在第一終端設備從第一基地台所覆蓋的社區切換至第二基地台所覆蓋的社區的場景中，

此時可以由第二基地台直接將第二指示資訊發送至第一終端設備，具體此處不做說明。

【0148】 另外，所描述的算力轉移確認消息還可以包含計算無線承載 CRB 配置資訊，該計算無線承載 CRB 配置資訊用於第一終端設備確定 CRB 資源，所描述的 CRB 資源用於第一終端設備與第一基地台之間交互算力 offloading 時的資料登錄輸出、或者計算模型資料或計算演算法資料。所描述的 CRB 可以參照前述圖 4a 中的步驟 403 所描述的 CRB 進行理解，此處不做贅述。

【0149】 506、第一基地台向第一終端設備發送算力配置資訊，算力配置資訊包括第一指示資訊和/或第二指示資訊。

【0150】 具體地，第一基地台也可以通過廣播 RRC 信令，將算力配置資訊發送至第一終端設備。

【0151】 在該示例中，在第一基地台確定出計算面剩餘計算資源不足以滿足第一終端設備所請求的算資源時，通過 X2 口信令向第二基地台發送算力轉移請求消息，由該第二基地台確定出相應的第二算力資源。這樣，第一基地台在獲取到第二基地台發送的第二指示資訊後，便可以結合第一指示資訊，向第一終端設備發送算力配置資訊，即該算力配置資訊中包括第一指示資訊和/或第二指示資訊。這樣，第一終端設備便可以根據第一指示資訊得到第一基地台的第一算力資源的支援、和/或根據第二指示資訊得到第二基地台的第二算力資源的支援，進而支撐與其他網路設備之間的 AI 協作。

【0152】 具體可以參閱圖 5b，為本申請實施例提供的一種算力轉移的示意圖。從圖 5b 可以看出，第一終端設備向第一基地台的控制面發送算力請求消息，在第一基地台的控制面與第一基地台的計算面中完成第一算力資源的分配和確認；並且，在第一基地台的控制面與第二基地台的控制面中完成跨基地台的第二算力資源的分配和確認，進而由第一基地台的控制面將該第一指示資訊和第二指示資訊發送至該第一終端設備，以完成算力資源的請求。

【0153】 最終，第一終端設備將算力 offloading 到該第一基地台的計算面，通過該 CRB，該第一終端設備與第一基地台之間可以交互算力 offloading 時的資料登錄輸出、或者計算模型資料或計算演算法資料。

【0154】 若第一基地台的算力資源不夠支持第一終端設備的算力請求時，第一基地台也將部分算力 offloading 到該第二基地台的計算面，通過 X2 介面交互算力 offloading 時的資料登錄輸出、或者計算模型資料或計算演算法資料。

【0155】 若因第一終端設備的移動性原因導致第一終端設備需要將算力從第一基地台 offloading 到第二基地台時，第一終端設備將算力 offloading 到該第二基地台的計算面，通過該 CRB，該第一終端設備與第二基地台之間可以交互算力 offloading 時的資料登錄輸出、或者計算模型資料或計算演算法資料。

【0156】 為了進一步理解本申請實施例提供的算力資源調度的方法，下面將以第一終端設備包括第一終端設備、第一設備包括第一

基地台、第二設備包括核心網設備為例進行說明。具體可以參閱圖 6a，為本申請實施例提供的另一種算力資源調度的方法流程圖。如圖 6a 所示，該算力資源調度的方法可以包括：

【0157】 601、第一終端設備通過廣播 RRC 信令向第一基地台發送算力請求消息，算力請求消息包括計算面資源的指示資訊。

【0158】 602、第一基地台根據計算面資源的指示資訊和第一基地台的計算面剩餘計算資源確定第一算力資源。

【0159】 需說明，本實施例中的步驟 601-602 具體可以參照上述圖 4a 中的步驟 401-402 所描述的內容進行理解，此處不做贅述。

【0160】 603、第一基地台通過 N2 口信令向核心網設備發送算力轉移請求消息。

【0161】 在該示例中，在第一基地台確定出自身的計算面剩餘計算資源無法滿足第一終端設備所請求的算力資源時，可以從核心網設備中獲取算力資源，進而滿足第一終端設備的請求。具體地，在計算面剩餘計算資源小於計算面資源的指示資訊中所請求的計算面資源時，第一基地台可以通過 N2 口信令向核心網設備發送算力轉移請求消息。另外，所描述的 N2 口可以理解成基地台與 AMF 網路元素之間的交互介面。

【0162】 需說明，所描述的算力轉移請求消息，也可以參照圖 3 中的步驟 301 中所描述的算力請求消息進行理解，此處不做贅述。

【0163】 604、核心網設備確定第二算力資源。

【0164】 在該實例中，核心網設備接收到第一基地台的算力轉移

請求消息後，可以根據該核心網設備的計算面剩餘計算資源和算力轉移請求消息確定第二算力資源。

【0165】 605、核心網設備向第一基地台發送算力轉移確認消息，算力轉移確認消息中包括第二指示資訊，第二指示資訊用於指示核心網設備提供的第二算力資源。

【0166】 該示例中，由於第一基地台是直接服務於第一終端設備的基地台，因此在核心網設備確定出第二算力資源後，可以通過 N2 口信令向第一基地台算力轉移確認消息，算力轉移確認消息中包括第二指示資訊。另外，所描述的算力轉移確認消息可以參照前述圖 5a 中步驟 505 進行理解，此處不做贅述。

【0167】 606、第一基地台向第一終端設備發送算力配置資訊，算力配置資訊包括第一指示資訊和/或第二指示資訊。

【0168】 需說明，本實施例中的步驟 606 具體可以參照上述圖 5a 中的步驟 506 所描述的內容進行理解，此處不做贅述。

【0169】 在該示例中，在第一基地台確定出計算面剩餘計算資源不足以滿足第一終端設備所請求的算資源時，通過 N2 口信令向核心網設備發送算力轉移請求消息，由該核心網設備確定出相應的第二算力資源。這樣，第一基地台在獲取到核心網設備發送的第二指示資訊後，便可以結合第一指示資訊，向第一終端設備發送算力配置資訊，即該算力配置資訊中包括第一指示資訊和/或第二指示資訊。這樣，第一終端設備便可以根據第一指示資訊從第一設備得到第一算力資源的支援、和/或根據第二指示資訊從第二設備得到

第二算力資源的支援，支撐與其他網路設備之間的 AI 協作。

【0170】 具體可以參閱圖 6b，為本申請實施例提供的一種算力轉移的示意圖。從圖 6b 可以看出，第一終端設備向第一基地台的控制面發送算力請求消息，進而在第一基地台的控制面與第一基地台的計算面中完成第一算力資源的分配和確認；並且在核心網設備的計算面與核心網設備的控制面中完成第二算力資源的分配和確認。這樣，第一基地台的控制面將該第一指示資訊和第二指示資訊發送至該第一終端設備，以完成算力資源的請求。

【0171】 最終，第一終端設備將算力 offloading 到該第一基地台的計算面，通過該 CRB，該第一終端設備與第一基地台之間可以交互算力 offloading 時的資料登錄輸出、或者計算模型資料或計算演算法資料。

【0172】 若第一基地台的算力資源不夠支持第一終端設備的算力請求時，第一基地台也將部分算力 offloading 到該核心網設備的計算面，通過 N2 介面交互算力 offloading 時的資料登錄輸出、或者計算模型資料或計算演算法資料。

【0173】 為了進一步理解本申請實施例提供的算力資源調度的方法，下面將以第一設備包括核心網設備、第二設備包括第一基地台為例進行說明。具體可以參閱圖 7a，為本申請實施例提供的另一種算力資源調度的方法流程圖。如圖 7a 所示，該算力資源調度的方法可以包括：

【0174】 701、第一終端設備通過 NAS 信令向核心網設備發送算

力請求消息，算力請求消息包括計算面資源的指示資訊。

【0175】 需說明，所描述的 NAS 口可以理解成第一終端設備與 AMF 網路元素之間的交互介面。

【0176】 702、核心網設備根據計算面資源的指示資訊和核心網的計算面剩餘計算資源確定第一算力資源。

【0177】 需說明，在實際應用中，核心網設備也可以獲取第一基地台和/或第二基地台的計算面資源後，結合計算面資源的指示資訊和自身的計算面剩餘計算資源確定第一任務。另外，上述的計算面資源的指示資訊可以參照前述步驟圖 3 中的步驟 301 所描述的內容進行理解，此處不做贅述。

【0178】 703、在核心網的計算面剩餘計算資源小於計算面資源的指示資訊中所請求的計算面資源時，核心網設備向第一基地台發送算力轉移請求消息、以及核心網設備的計算面剩餘計算資源。

【0179】 在該示例中，在核心網設備確定出自身的計算面剩餘計算資源無法滿足第一終端設備所請求的算力資源時，可以請求第一基地台提供第二算力資源，進而滿足第一終端設備的請求。具體地，在計算面剩餘計算資源小於計算面資源的指示資訊中所請求的計算面資源時，核心網設備向第一基地台發送算力轉移請求消息，並且將核心網設備的計算面剩餘計算資源發送至第一基地台。

【0180】 需說明，所描述的算力轉移請求消息，也可以參照圖 3 中的步驟 301 中所描述的算力請求消息進行理解，此處不做贅述。

【0181】 704、第一基地台確定第二算力資源。

【0182】 在該實例中，第一基地台接收到核心網設備的算力轉移請求消息後，可以根據該第一基地台的計算面剩餘計算資源和算力轉移請求消息確定第二算力資源。也就是理解成，如果核心網設備自身的計算面計算資源不足以滿足該第一終端設備所需求的計算面資源時，該核心網設備可以向其他具有計算面資源的第一基地台發送算力轉移請求消息，進而第一基地台能夠提供相應的算力轉移資源的指示資訊給第一終端設備。

【0183】 需說明，在一些實施例中，如果第一基地台自身的計算面計算資源也不足以滿足該第一終端設備所需求的計算面資源時，還可以進一步地向第二基地台請求相應的第二算力資源，具體可以參閱前述圖 5a 中的步驟 502-步驟 505 進行理解。所描述的第一基地台是直接向第一終端設備提供服務的基地台，而第二基地台可以理解成不直接向第一終端設備提供業務服務的基地台。

【0184】 705、第一基地台向第一終端設備發送算力配置資訊，算力配置資訊包括第一指示資訊和/或第二指示資訊。

【0185】 在該示例中，在核心網設備確定出計算面剩餘計算資源不足以滿足第一終端設備所請求的算力資源時，通過向第一基地台發送算力轉移請求消息，由該第一基地台確定出相應的第二算力資源。這樣，第一基地台在確定出第二算力資源後，便可以向第一終端設備發送算力配置資訊，即該算力配置資訊中包括第一指示資訊和/或第二指示資訊。這樣，第一終端設備便可以根據第一指示資訊從第一設備得到第一算力資源的支援、和/或根據第二指

示資訊從第二設備得到第二算力資源的支援，支撐與其他網路設備之間的 AI 協作。

【0186】 具體可以參閱圖 7b，為本申請實施例提供的一種算力轉移的示意圖。從圖 7b 可以看出，第一終端設備向核心網設備的計算面資源管理器發送算力請求消息，進而在核心網設備的計算面資源管理器與核心網的計算面中完成第一算力資源的分配和確認；而且在核心網設備的計算面資源管理器、第一基地台的計算面以及第二基地台的計算面中完成第二算力資源的分配和確認。這樣，第一基地台將該第一指示資訊和第二指示資訊發送至該第一終端設備，以完成算力資源的請求。

【0187】 最終，第一終端設備將算力 offloading 到該第一基地台的計算面，通過該 CRB，該第一終端設備與第一基地台之間可以交互算力 offloading 時的資料登錄輸出、或者計算模型資料或計算演算法資料。

【0188】 若第一基地台的算力資源不夠支持第一終端設備的算力請求時，第一基地台也將部分算力 offloading 到該核心網設備的計算面，通過 N2 介面交互算力 offloading 時的資料登錄輸出、或者計算模型資料或計算演算法資料；和/或，第一基地台也可以將部分算力 offloading 到該第二基地台的計算面，通過 X2 介面交互算力 offloading 時的資料登錄輸出、或者計算模型資料或計算演算法資料。

【0189】 為了進一步理解本申請實施例提供的算力資源調度的方

法，下面將以第一設備包括第一基地台為例進行說明。具體可以參閱圖 8，為本申請實施例提供的另一種算力資源調度的方法流程圖。如圖 8 所示，該算力資源調度的方法可以包括：

【0190】 801、第一基地台廣播第一基地台的算力配置資訊。

【0191】 802、第一終端設備通過向第一基地台發送 RRC 消息或隨機存取過程的消息，RRC 消息或隨機存取過程的消息包括計算面資源的指示資訊。

【0192】 在該示例中，第一終端設備可以在確定出自身的計算面資源不足以支撐與其他網路設備之間的 AI 協作時，可以在第一基地台已經通過廣播的方式廣播該第一基地台的算力資源後，向第一基地台發送 RRC 消息或隨機存取過程的消息，請求第一基地台提供相應的算力資源。需說明，該 RRC 消息或隨機存取過程的消息中包括計算面資源的指示資訊，這樣第一基地台在接收到該 RRC 消息或隨機存取過程的消息後，通過解析該 RRC 消息或隨機存取過程的消息得到計算面資源的指示資訊後，便可以獲知需要向第一終端設備提供多少算力資源了。

【0193】 803、第一基地台根據計算面資源的指示資訊確定第一算力資源。

【0194】 804、第一基地台通過 RRC 回應訊息或隨機存取過程的回應訊息向第一終端設備發送算力配置資訊，算力配置資訊包括第一指示資訊。

【0195】 在該示例中，在第一基地台確定出第一算力資源後，便可

以向第一終端設備發送算力配置資訊。這樣，第一終端設備便可以獲取到算力配置資訊，進而得到該算力配置資訊中所包括的第一指示資訊。

【0196】 可選的，在另一些實施例中，上述的算力配置資訊的更新週期跟系統廣播消息的傳輸週期相同，算力配置資訊統計的是系統廣播消息週期內的平均算力大小，所支援的計算模型或演算法等。

【0197】 可選的，在另一些實施例中，RRC 回應訊息或隨機存取過程的回應訊息中傳輸的是第一基地台的計算面的即時算力統計資訊。

【0198】 可選的，在另一些實施例中，第一基地台廣播的算力配置資訊還可以包括核心網設備的計算能力資訊、和/或第二基地台的計算能力資訊。其中，第一基地台可以通過基地台間介面，如 X2/Xn 介面獲取第二基地台的計算能力資訊；第一基地台也可以通過第一基地台與核心網設備之間的介面獲取核心網設備的計算能力資訊，具體此處不做限定說明。

【0199】 為了進一步理解本申請實施例提供的算力資源調度的方法，下面將以第一設備包括第一基地台、第二設備包括第二終端設備為例進行說明。具體可以參閱圖 9a，為本申請實施例提供的另一種算力資源調度的方法流程圖。如圖 9a 所示，該算力資源調度的方法可以包括：

【0200】 901、第一終端設備通過廣播 RRC 信令向第一基地台發

送算力請求消息，算力請求消息包括計算面資源的指示資訊。

【0201】 902、第一基地台根據計算面資源的指示資訊和計算面剩餘計算資源確定第一算力資源。

【0202】 需說明，本實施例中的步驟 901-902 具體可以參照上述圖 5a 中的步驟 501-502 所描述的內容進行理解，此處不做贅述。

【0203】 903、第一基地台向第二終端設備發送算力轉移請求消息，以指示第一終端設備的算力需求或第一基地台的算力需求。

【0204】 在該示例中，在第一基地台確定出自身的計算面剩餘計算資源無法滿足第一終端設備所請求的算力資源時，該第一基地台可以向第二終端設備發送算力轉移請求消息，以指示該第二終端設備提供第二算力資源，進而滿足第一終端設備的請求。具體地，第一基地台可以通過廣播 RRC 信令向第二終端設備發送算力轉移請求消息。

【0205】 需說明，所描述的算力轉移請求消息，也可以參照圖 3 中的步驟 301 中所描述的算力請求消息進行理解，此處不做贅述。

【0206】 另外，所描述的算力轉移請求消息還可以包含計算無線承載 CRB 配置資訊，該計算無線承載 CRB 配置資訊用於第二終端設備確定 CRB 資源，所描述的 CRB 資源用於第二終端設備與第一基地台之間交互算力 offloading 時的資料登錄輸出、或者計算模型資料或計算演算法資料。所描述的 CRB 可以參照前述圖 4a 中的步驟 403 所描述的 CRB 進行理解，此處不做贅述。

【0207】 可選的，所描述的算力轉移請求消息還可以包含計算無

線承載 CRB 配置資訊，該計算無線承載 CRB 配置資訊用於第二終端設備確定 CRB 資源，所描述的 CRB 資源用於第二終端設備與第一終端之間交互算力 offloading 時的資料登錄輸出、或者計算模型資料或計算演算法資料。

【0208】 904、第二終端設備根據算力轉移請求消息和第二終端設備的計算面剩餘計算資源確定第二算力資源。

【0209】 在該實例中，第二終端設備接收到第一基地台的算力轉移請求消息後，可以根據該第二終端設備的計算面剩餘計算資源和算力轉移請求消息確定第二算力資源。

【0210】 也就是理解成，如果第一基地台自身的計算面計算資源不足以滿足該第一終端設備所需求的算力資源時，該第一基地台可以向其他具有計算面資源的第二終端設備發送算力轉移請求消息，進而指示第二終端設備能夠提供相應的第二算力資源給第一終端設備。

【0211】 905、第二終端設備向第一基地台發送算力轉移確認消息，算力轉移確認消息中包括第二指示資訊，第二指示資訊用於指示第二終端設備提供的第二算力資源。

【0212】 該示例中，由於第一基地台是直接服務於第一終端設備的基地台，因此在第二終端設備確定出第二算力資源後，可以該通過算力轉移確認消息的方式將第二指示資訊發送至第一基地台。具體地，該第二終端設備也可以通過廣播 RRC 信令向第一基地台發送算力轉移確認消息，進而將包含在算力轉移確認消息中的第

二指示資訊發送至第一基地台。

【0213】 906、第一基地台向第一終端設備發送算力配置資訊，算力配置資訊包括第一指示資訊和/或第二指示資訊。

【0214】 具體地，第一基地台也可以通過廣播 RRC 信令，將算力配置資訊發送至第一終端設備。

【0215】 在該示例中，在第一基地台確定出計算面剩餘計算資源不足以滿足第一終端設備所請求的算資源時，通過廣播 RRC 信令向第二終端設備發送算力轉移請求消息，由該第二終端設備確定出相應的第二算力資源。這樣，第一基地台在獲取到第二終端設備發送的第二指示資訊後，便可以結合第一指示資訊，向第一終端設備發送算力配置資訊，即該算力配置資訊中包括第一指示資訊和/或第二指示資訊。這樣，第一終端設備便可以根據第一指示資訊得到第一基地台的第一算力資源的支援、和/或根據第二指示資訊得到第二終端設備中的第二算力資源的支援，進而支撐與其他網路設備之間的 AI 協作。

【0216】 具體可以參閱圖 9b，為本申請實施例提供的一種算力轉移的示意圖。從圖 9b 可以看出，第一終端設備向第一基地台的控制面發送算力請求消息，在第一基地台的控制面與第二終端設備的控制面中完成跨基地台的第二算力資源的分配和確認，進而由第一基地台的控制面將該第一指示資訊和第二指示資訊發送至該第一終端設備，以完成算力資源的請求。

【0217】 最終，第一終端設備將算力 offloading 到該第一基地台的

計算面，通過該 CRB，該第一終端設備與第一基地台之間可以交互算力 offloading 時的資料登錄輸出、或者計算模型資料或計算演算法資料。

【0218】 若第一基地台的算力資源不夠支持第一終端設備的算力請求時，第一基地台也將部分算力 offloading 到該第二終端設備的計算面，通過廣播或 CRB 交互算力 offloading 時的資料登錄輸出、或者計算模型資料或計算演算法資料。

【0219】 若第一終端設備與第二終端設備是臨近終端，第一終端設備與第二終端設備還可以直接通過車聯萬物（vehicle to everything,V2X）或設備到設備通訊（device-to-device communication，D2D）介面交互算力 offloading 時的資料登錄輸出、或者計算模型資料或計算演算法資料。

【0220】 本申請實施例中，在第一終端設備的算力資源不足以支撐與其他網路設備之間的 AI 協作時，可以向第一設備發送算力請求消息，進而由第一設備根據該算力請求消息中的計算面資源的指示資訊和該第一設備的計算面剩餘計算資源確定第一算力資源。然後由第一設備向第一終端設備發送相應的算力配置資訊，該算力配置資訊中包括第一指示資訊和/或第二指示資訊。這樣，第一終端設備便能夠根據第一指示資訊從第一設備得到第一算力資源的支援、和/或根據第二指示資訊從第二設備得到第二算力資源的支援，支撐與其他網路設備之間的 AI 協作。

【0221】 上述主要從方法的角度對本申請實施例提供的方案進行

了介紹。可以理解的是，上述的第一終端設備、第一設備為了實現上述功能，其包含了執行各個功能相應的硬體結構和/或軟體模組。本領域技術人員應該很容易意識到，結合本文中所公開的實施例描述的功能，本申請能夠以硬體或硬體和電腦軟體的結合形式來實現。某個功能究竟以硬體還是電腦軟體驅動硬體的方式來執行，取決於技術方案的特定應用和設計約束條件。專業技術人員可以對每個特定的應用來使用不同方法來實現所描述的功能，但是這種實現不應認為超出本申請的範圍。

**【0222】** 從功能單元的角度，本申請可以根據上述方法實施例對第一終端設備、第一設備進行功能單元的劃分，例如，可以對應各個功能劃分各個功能單元，也可以將兩個或兩個以上的功能整合在一個功能單元中。上述整合的功能單元既可以採用硬體的形式實現，也可以採用軟體功能單元的形式實現。

**【0223】** 比如，以採用整合的方式劃分各個功能單元的情況下，參見圖 10，為本申請實施例提供了一種第一終端設備的結構示意圖。所描述的第一終端設備可以包括：

**【0224】** 發送單元 1001，用於向第一設備發送算力請求消息，算力請求消息包括計算面資源的指示資訊，計算面資源的指示資訊用於指示第一終端設備的算力需求；

**【0225】** 接收單元 1002，用於接收第一設備發送的算力配置資訊，算力配置資訊包括第一指示資訊和/或第二指示資訊，所述第一指示資訊用於指示所述第一設備提供第一算力資源，所述第二指示

資訊用於指示第二設備提供的第二算力資源，第一設備與第二設備不相同。

【0226】 可選地，在一些示例中，第一設備為第一基地台；發送單元 1001，用於通過廣播 RRC 信令向第一設備發送算力請求消息。

【0227】 可選的，在一些實施例中，所述第一設備為第一基地台；所述接收單元 1002，用於接收所述第一設備通過廣播 RRC 信令發送的算力配置資訊。

【0228】 可選地，在一些示例中，第一設備為核心網設備；發送單元 1001，用於通過 NAS 信令向第一設備發送算力請求消息。

【0229】 可選的，在一些實施例中，所述第一設備為核心網設備；所述接收單元 1002，用於接收所述第一設備通過 NAS 信令發送的算力配置資訊。

【0230】 可選地，在一些示例中，計算面資源的指示資訊包括算力大小、執行時間、算力任務類型、計算模型、計算演算法、執行結果傳輸觸發類型和觸發類型中的一個或多個。

【0231】 上述主要從描述了第一終端設備的結構。下面將從功能單元的角度介紹前述所描述的第一設備。比如，以採用整合的方式劃分各個功能單元的情況下，參見圖 11，本申請實施例提供了一種第一設備的結構示意圖，該第一設備可以包括：

【0232】 獲取模組 1101，用於接收第一終端設備發送的算力請求消息，算力請求消息包括計算面資源的指示資訊，計算面資源的指示資訊用於指示第一終端設備的算力需求；

【0233】 處理模組 1102，用於根據計算面資源的指示資訊和第一設備的計算面剩餘計算資源確定第一算力資源；

【0234】 發送模組 1103，用於向第一終端設備發送算力配置資訊，算力配置資訊包括第一指示資訊和/或第二指示資訊，所述第一指示資訊用於指示所述第一設備提供第一算力資源，所述第二指示資訊用於指示第二設備提供的第二算力資源，第一設備與第二設備不相同。

【0235】 可選地，在一些示例中，發送模組 1103，用於向第二設備發送算力轉移請求消息，以指示所述第一終端設備的算力需求或所述第一設備的算力需求；獲取模組 1101，用於接收第二設備發送的算力轉移確認消息，所述算力轉移確認消息中包括第二指示資訊。

【0236】 可選地，在一些示例中，第一設備包括第一基地台，第二設備包括第二基地台；發送模組 1103，用於通過 X2 口信令向第二設備發送算力轉移請求消息。

【0237】 可選的，在一些實施例中，所述第一設備包括第一基地台，所述第二設備包括第二基地台；所述獲取模組 1101，用於接收所述第二設備通過 X2 口信令發送的算力轉移確認消息。

【0238】 可選地，在一些示例中，第一設備包括第一基地台，第二設備包括核心網設備；發送模組 1103，用於通過 N2 口信令向第二設備發送算力轉移請求消息。

【0239】 可選的，在一些實施例中，所述第一設備包括第一基地台，

所述第二設備包括核心網設備；所述獲取模組 1101，用於接收所述第二設備通過 N2 口信令發送的算力轉移確認消息。

【0240】 可選地，在一些示例中，第一設備為第一基地台；發送模組 1103，用於接收第一終端設備通過廣播 RRC 信令發送的算力請求消息。

【0241】 可選的，在一些實施例中，所述第一設備為所述第一基地台；所述發送模組 1103，用於通過廣播 RRC 信令向所述第一終端設備發送算力配置資訊。

【0242】 可選地，在一些示例中，第一設備為核心網設備，第二設備包括第一基地台和/或第二基地台；發送模組 1103，用於接收第一終端設備通過 NAS 信令發送的算力請求消息。

【0243】 可選的，在一些實施例中，所述第一設備為核心網設備，所述第二設備包括第一基地台；所述發送模組 1103，用於通過 NAS 信令向所述第一終端設備發送算力配置資訊。

【0244】 可選的，在一些實施例中，所述第一設備包括第一基地台，所述第二設備包括第二終端設備；所述發送模組 1103，用於通過廣播 RRC 信令向所述第二設備發送算力轉移請求消息。

【0245】 可選的，在一些實施例中，所述第一設備包括第一基地台，所述第二設備包括第二終端設備；所述獲取模組 1101，用於接收所述第二設備通過廣播 RRC 信令發送的算力轉移確認消息。

【0246】 可選地，在一些示例中，計算面資源的指示資訊包括算力大小、執行時間、算力任務類型、計算模型、計算演算法、執行結

果傳輸觸發類型和觸發類型中的一個或多個。

【0247】 上述主要從描述了第一終端設備、第一設備的結構。下面將從功能單元的角度介紹前述所描述的第二設備。比如，以採用整合的方式劃分各個功能單元的情況下，參見圖 12，本申請實施例提供了一種第二設備的結構示意圖，該第二設備可以包括：

【0248】 第一接收模組 1201，用於接收第一設備發送的算力轉移請求消息；第一處理模組，用於根據算力轉移請求消息和所述第二設備的計算面剩餘計算資源確定第二算力資源；

【0249】 第一發送模組 1202，用於向第一設備發送算力轉移確認消息，所述算力轉移確認消息中包括第二指示資訊，所述第二指示資訊用於指示所述第二設備提供的所述第二算力資源。

【0250】 可選的，在一些實施例中，第一接收模組 1201，用於接收第一設備通過 X2 口信令發送的算力轉移請求消息。

【0251】 可選的，在一些實施例中，所述第一設備包括第一基地台，所述第二設備包括第二基地台；所述第一發送模組 1202，用於通過 X2 口信令向所述第一設備發送算力轉移確認消息。

【0252】 可選的，在一些實施例中，第一接收模組 1201，用於接收第一設備通過 N2 口信令發送的算力轉移請求消息。

【0253】 可選的，在一些實施例中，所述第一設備包括第一基地台，所述第二設備包括核心網設備；所述第一發送模組 1202，用於通過 N2 口信令向所述第一設備發送算力轉移確認消息。

【0254】 可選的，在一些實施例中，所述第一設備包括第一基地台，

所述第二設備包括第二終端設備；所述第一接收模組 1201，用於接收所述第一設備通過廣播 RRC 信令發送的算力轉移請求消息。

【0255】 可選的，在一些實施例中，所述第一設備包括第一基地台，所述第二設備包括第二終端設備；第一發送模組 1202，用於通過廣播 RRC 信令向所述第一設備發送算力轉移確認消息。

【0256】 需要說明的是，上述裝置各模組/單元之間的資訊交互、執行過程等內容，由於與本申請方法實施例基於同一構思，其帶來的技術效果與本申請方法實施例相同，具體內容可參見本申請前述所示的方法實施例中的敘述，此處不再贅述。

【0257】 上面從模組化功能實體的角度對本申請實施例中的第一終端設備、第一設備以及第二設備進行描述。從實體設備角度來描述，上述第一終端設備、第一設備以及第二設備可以由一個實體設備實現，也可以由多個實體設備共同實現，還可以是一個實體設備內的一個邏輯功能單元，本申請實施例對此不做具體限定。

【0258】 例如，上述第一終端設備、第一設備以及第二設備可以由圖 13 中的通訊設備來實現。圖 13 為本申請實施例提供的通訊設備的硬體結構示意圖。該通訊設備包括至少一個處理器 1301、通訊線路 1307、記憶體 1303 以及至少一個通訊介面 1304。

【0259】 處理器 1301 可以是一個通用中央處理器（central processing unit，CPU）、微處理器、特定應用積體電路（application-specific integrated circuit，伺服器 IC）或一個或多個用於控制本申請方案程式執行的積體電路。

【0260】 通訊線路 1307 可包括一通路，在上述元件之間傳送資訊。

【0261】 通訊介面 1304，使用任何收發器一類的裝置，用於與其他裝置或通訊網路通訊，如乙太網路等。

【0262】 記憶體 1303 可以是唯讀記憶體 (read-only memory, ROM) 或可儲存靜態資訊和指令的其他類型的靜態儲存裝置，隨機存取記憶體 (random access memory, RAM) 或者可儲存資訊和指令的其他類型的動態儲存裝置，也可以是電子可抹除程式化唯讀記憶體 (electrically erable programmable read-only memory, EEPROM)、唯讀光碟 (compact disc read-only memory, CD-ROM) 或其他光碟儲存、光碟儲存 (包括壓縮光碟、鐳射碟、光碟、數位通用光碟、藍光光碟等)、磁片儲存媒體或者其他磁存放裝置、或者能夠用於攜帶或儲存具有指令或資料結構形式的期望的程式碼並能夠由電腦存取的任何其他媒體，但不限於此。記憶體可以是獨立存在，通過通訊線路 1307 與處理器相連接。記憶體 1303 也可以和處理器 1301 整合在一起。

【0263】 其中，記憶體 1303 用於儲存執行本申請方案的電腦執行指令，並由處理器 1301 來控制執行。處理器 1301 用於執行記憶體 1303 中儲存的電腦執行指令，從而實現本申請上述實施例提供的算力資源調度的方法。

【0264】 可選的，本申請實施例中的電腦執行指令也可以稱之為應用程式碼，本申請實施例對此不作具體限定。

【0265】 在具體實現中，作為一種實施例，處理器 1301 可以包括

一個或多個 CPU，例如圖 13 中的 CPU0 和 CPU1。

【0266】 在具體實現中，作為一種實施例，通訊裝置可以包括多個處理器，例如圖 13 中的處理器 1301 和處理器 1302。這些處理器中的每一個可以是一個單核（single-CPU）處理器，也可以是一個多核（multi-CPU）處理器。這裡的處理器可以指一個或多個裝置、電路、和/或用於處理資料（例如電腦程式指令）的處理核。

【0267】 在具體實現中，作為一種實施例，通訊裝置還可以包括輸出裝置 1305 和輸入裝置 1306。輸出裝置 1305 和處理器 1301 通訊，可以以多種方式來顯示資訊。輸入裝置 1306 和處理器 1301 通訊，可以以多種方式接收使用者的輸入。例如，輸入裝置 1306 可以是滑鼠、觸控式螢幕裝置或傳感裝置等。

【0268】 上述的通訊裝置可以是一個通用裝置或者是一個專用裝置。在具體實現中，通訊裝置可以是可攜式電腦、移動終端等或有圖 13 中類似結構的裝置。本申請實施例不限定通訊裝置的類型。

【0269】 需說明，圖 13 中的處理器 1301 可以通過調用記憶體 1303 中儲存的電腦執行指令，使得第一終端設備執行如圖 3-圖 9a 對應的方法實施例中第一終端設備所執行的方法。圖 13 中的處理器 1301 可以通過調用記憶體 1303 中儲存的電腦執行指令，使得第一設備執行如圖 3-圖 9a 對應的方法實施例中第一設備所執行的方法。圖 13 中的處理器 1301 可以通過調用記憶體 1303 中儲存的電腦執行指令，使得第二設備執行如圖 3-圖 9a 對應的方法實施例中第二設備所執行的方法。

【0270】 具體的，圖 11 中的處理模組 1102 的功能/實現過程可以通過圖 13 中的處理器 1301 調用記憶體 1303 中儲存的電腦執行指令來實現；圖 10 中的發送單元 1001、接收單元 1002，圖 11 中的獲取模組 1101、發送模組 1103，圖 12 中的第一接收模組 1201、第一發送模組 1202 的功能/實現過程可以通過圖 13 中的通訊介面 1304 來實現。

【0271】 在上述實施例中，可以全部或部分地通過軟體、硬體、軟體或者其任意組合來實現。當使用軟體實現時，可以全部或部分地以電腦程式產品的形式實現。

【0272】 所屬領域的技術人員可以清楚地瞭解到，為描述的方便和簡潔，上述描述的系統，裝置和單元的具體工作過程，可以參考前述方法實施例中的對應過程，在此不再贅述。

【0273】 在本申請所提供的幾個實施例中，應該理解到，所揭露的系統，裝置和方法，可以通過其它的方式實現。例如，以上所描述的裝置實施例僅僅是示意性的，例如，該單元的劃分，僅僅為一種邏輯功能劃分，實際實現時可以有另外的劃分方式，例如多個單元或元件可以結合或者可以整合到另一個系統，或一些特徵可以忽略，或不執行。另一點，所顯示或討論的相互之間的耦合或直接耦合或通訊連接可以是通過一些介面，裝置或單元的間接耦合或通訊連接，可以是電性、機械或其它的形式。

【0274】 該作為分離部件說明的單元可以是或者也可以不是物理上分開的，作為單元顯示的部件可以是或者也可以不是物理單元，

即可以位於一個地方，或者也可以分佈到多個網路單元上。可以根據實際的需要選擇其中的部分或者全部單元來實現本實施例方案的目的。

【0275】 另外，在本申請各個實施例中的各功能單元可以整合在一個處理單元中，也可以是各個單元單獨物理存在，也可以兩個或兩個以上單元整合在一個單元中。上述整合的單元既可以採用硬體的形式實現，也可以採用軟體功能單元的形式實現。

【0276】 該整合的單元如果以軟體功能單元的形式實現並作為獨立的產品銷售或使用時，可以儲存在一個電腦可讀取儲存媒體中。基於這樣的理解，本申請的技術方案本質上或者說對現有技術做出貢獻的部分或者該技術方案的全部或部分可以以軟體產品的形式體現出來，該電腦軟體產品儲存在一個儲存媒體中，包括若干指令用以使得一台電腦設備（可以是個人電腦，伺服器，或者網路設備等）執行本申請各個實施例該方法的全部或部分步驟。而前述的儲存媒體包括：隨身碟、移動硬碟、唯讀記憶體（read-only memory，ROM）、隨機存取記憶體（random access memory，RAM）、磁碟或者光碟等各種可以儲存程式碼的媒體。

【0277】 上述實施例，可以全部或部分地通過軟體、硬體、韌體或者其任意組合來實現，當使用軟體實現時，可以全部或部分地以電腦程式產品的形式實現。

【0278】 電腦程式產品包括一個或多個電腦指令。在電腦上載入和執行電腦執行指令時，全部或部分地產生按照本申請實施例的

流程或功能。電腦可以是通用電腦、專用電腦、電腦網路、或者其他可程式設計裝置。電腦指令可以儲存在電腦可讀儲存媒體中，或者從一個電腦可讀儲存媒體向另一電腦可讀儲存媒體傳輸，例如，電腦指令可以從一個網站、電腦、伺服器或資料中心通過有線（例如同軸電纜、光纖、數位用戶線路（DSL））或無線（例如紅外、無線、微波等）方式向另一個網站、電腦、伺服器或資料中心進行傳輸。電腦可讀儲存媒體可以是電腦能夠儲存的任何可用媒體或者是包含一個或多個可用媒體整合的伺服器、資料中心等資料存放裝置。可用媒體可以是磁性媒體，（例如，軟碟、硬碟、磁帶）、光媒體（例如，DVD）、或者半導體媒體（例如 SSD）等。

**【0279】** 以上實施例僅用以說明本申請的技術方案，而非對其限制；儘管參照前述實施例對本申請進行了詳細的說明，本領域的普通技術人員應當理解：其依然可以對前述各實施例所記載的技術方案進行修改，或者對其中部分技術特徵進行等同替換；而這些修改或者替換，並不使相應技術方案的本質脫離本申請各實施例技術方案的精神和範圍。

### **【符號說明】**

#### **【0280】**

1001：發送單元

1002：接收單元

1101：獲取模組

1102：處理模組

1103：發送模組

1201：第一接收模組

1202：第一發送模組

1301、1302：處理器

1303：記憶體

1304：通訊介面

1305：輸出裝置

1306：輸入裝置

1307：通訊線路

301～306、401～403、501～506、601～606、701～705、

801～804、901～906：步驟

AF：應用功能網路元素

AMF：存取和移動性管理功能網路元素

DN：資料網路

N1、N2、N3、N4、N6：連接

NEF：網路暴露功能網路元素

Nnf、Nnwdaf：介面

NWDAF：網路資料分析功能網路元素

NSSF：網路切片選擇功能網路元素

PCF：策略控制功能網路元素

(R)AN：（無線）存取網路

SMF：會話管理功能網路元素

UDM：統一資料管理功能網路元素

UE：使用者設備

UPF：向用戶面功能網路元素

## 【發明申請專利範圍】

【請求項1】 一種算力資源調度的方法，應用於第一終端設備，所述方法包括：

向第一設備發送算力請求消息，所述算力請求消息包括計算面資源的指示資訊，所述計算面資源的指示資訊用於指示所述第一終端設備的算力需求；以及

接收所述第一設備發送的算力配置資訊，所述算力配置資訊包括第二指示資訊、或第一指示資訊和所述第二指示資訊，所述第一指示資訊用於指示所述第一設備提供第一算力資源，所述第二指示資訊用於指示第二設備提供的第二算力資源，所述第一設備與所述第二設備不相同，其中所述方法還包括：

所述第一設備向所述第二設備發送算力轉移請求消息，以指示所述第一終端設備的所述算力需求或所述第一設備的算力需求；以及

所述第一設備接收所述第二設備發送的算力轉移確認消息，所述算力轉移確認消息中包括所述第二指示資訊，所述第二算力資源是所述第二設備根據所述算力轉移請求消息和所述第二設備的計算面剩餘計算資源確定的。

【請求項2】 如請求項1所述的方法，其中第一設備為第一基地台，向所述第一設備發送所述算力請求消息，包括：

通過廣播 RRC 信令向所述第一設備發送所述算力請求消息。

【請求項3】 如請求項1所述的方法，其中所述第一設備為核心網設備，向所述第一設備發送所述算力請求消息，包括：

通過 NAS 信令向所述第一設備發送所述算力請求消息。

【請求項4】 如請求項1-3中任一項所述的方法，其中所述計算面資源的指示資訊包括算力大小、執行時間、算力任務類型、計算模型、計算演算法、執行結果傳輸觸發類型和觸發類型中的一個或多個。

【請求項5】 一種算力資源調度的方法，應用於第一設備，所述方法包括：

接收第一終端設備發送的算力請求消息，所述算力請求消息包括計算面資源的指示資訊，所述計算面資源的指示資訊用於指示所述第一終端設備的算力需求；

根據所述計算面資源的指示資訊和所述第一設備的計算面剩餘計算資源確定第一算力資源；以及

向所述第一終端設備發送算力配置資訊，所述算力配置資訊包括第二指示資訊、或第一指示資訊和所述第二指示資訊，所述第一指示資訊用於指示所述第一設備提供所述第一算力資源，所述第二指示資訊用於指示第二設備提供的第二算力資源，所述第一設備與所述第二設備不相同，其中所述方法還包括：

向所述第二設備發送算力轉移請求消息，以指示所述第一終端設備的所述算力需求或所述第一設備的算力需求；以及

接收所述第二設備發送的算力轉移確認消息，所述算力轉移

確認消息中包括所述第二指示資訊，所述第二算力資源是所述第二設備根據所述算力轉移請求消息和所述第二設備的計算面剩餘計算資源確定的。

【請求項6】 如請求項5所述的方法，其中所述第一設備包括第一基地台，所述第二設備包括第二基地台，其中向所述第二設備發送所述算力轉移請求消息，包括：

通過 X2 口信令向所述第二設備發送所述算力轉移請求消息。

【請求項7】 如請求項5所述的方法，其中所述第一設備包括第一基地台，所述第二設備包括核心網設備，其中向所述第二設備發送所述算力轉移請求消息，包括：

通過 N2 口信令向所述第二設備發送所述算力轉移請求消息。

【請求項8】 如請求項6-7中任一項所述的方法，其中所述第一設備為所述第一基地台，接收所述第一終端設備發送的所述算力請求消息，包括：

接收所述第一終端設備通過廣播 RRC 信令發送的所述算力請求消息。

【請求項9】 如請求項5所述的方法，其中所述第一設備為核心網設備，所述第二設備包括第一基地台，其中所述接收所述第一終端設備發送的所述算力請求消息，包括：

接收所述第一終端設備通過 NAS 信令發送的所述算力請求消息。

【請求項10】 如請求項5所述的方法，其中所述計算面資源的指示資訊包括算力大小、執行時間、算力任務類型、計算模型、計算演算法、執行結果傳輸觸發類型和觸發類型中的一個或多個。

【請求項11】 一種算力資源調度的方法，應用於第二設備，所述方法包括：

接收第一設備發送的算力轉移請求消息，其中，在所述第一設備發送所述算力轉移請求消息之前，所述第一設備接收第一終端設備發送的算力請求消息，所述算力請求消息包括計算面資源的指示資訊，所述計算面資源的所述指示資訊用於指示所述第一終端設備的算力需求；

根據所述算力轉移請求消息和所述第二設備的計算面剩餘計算資源確定第二算力資源；以及

向所述第一設備發送算力轉移確認消息，所述算力轉移確認消息中包括所述第二指示資訊，所述第二指示資訊用於指示所述第二設備提供的所述第二算力資源，其中，所述第一設備向所述第一終端設備發送算力配置資訊，所述算力配置資訊包括所述第二指示資訊、或第一指示資訊和所述第二指示資訊，所述第一指示資訊用於指示所述第一設備提供所述第一算力資源，所述第一設備與所述第二設備不相同。

【請求項12】 如請求項11所述的方法，其中所述第一設備包括第一基地台，所述第二設備包括第二基地台，其中接收所述第一設備發送的所述算力轉移請求消息，包括：

接收所述第一設備通過 X2 口信令發送的所述算力轉移請求消息。

【請求項13】 如請求項11所述的方法，其中所述第一設備包括第一基地台，所述第二設備包括核心網設備，其中接收所述第一設備發送的所述算力轉移請求消息，包括：

接收所述第一設備通過 N2 口信令發送的所述算力轉移請求消息。

【請求項14】 一種第一終端設備，包括：

發送單元，用於向第一設備發送算力請求消息，所述算力請求消息包括計算面資源的指示資訊，所述計算面資源的指示資訊用於指示所述第一終端設備的算力需求；以及

接收單元，用於接收所述第一設備發送的算力配置資訊，所述算力配置資訊包括第二指示資訊、或第一指示資訊和所述第二指示資訊，所述第一指示資訊用於指示所述第一設備提供第一算力資源，所述第二指示資訊用於指示第二設備提供的第二算力資源，所述第一設備與所述第二設備不相同，其中

所述第一設備向所述第二設備發送算力轉移請求消息，以指示所述第一終端設備的所述算力需求或所述第一設備的算力需求；以及

所述第一設備接收所述第二設備發送的算力轉移確認消息，所述算力轉移確認消息中包括所述第二指示資訊，所述第二算力資源是所述第二設備根據所述算力轉移請求消息和所述第二設備

的計算面剩餘計算資源確定的。

【請求項15】 如請求項14所述的第一終端設備，其中所述第一設備為第一基地台；

所述發送單元，用於通過廣播 RRC 信令向所述第一設備發送所述算力請求消息。

【請求項16】 如請求項14所述的第一終端設備，其中所述第一設備為核心網設備；

所述發送單元，用於通過 NAS 信令向所述第一設備發送所述算力請求消息。

【請求項17】 如請求項14-16中任一項所述的第一終端設備，其中所述計算面資源的指示資訊包括算力大小、執行時間、算力任務類型、計算模型、計算演算法、執行結果傳輸觸發類型和觸發類型中的一個或多個。

【請求項18】 一種第一設備，包括：

獲取模組，用於接收第一終端設備發送的算力請求消息，所述算力請求消息包括計算面資源的指示資訊，所述計算面資源的所述指示資訊用於指示所述第一終端設備的算力需求；

處理模組，用於根據所述計算面資源的所述指示資訊和所述第一設備的計算面剩餘計算資源確定第一算力資源；以及

發送模組，用於向所述第一終端設備發送算力配置資訊，所述算力配置資訊包括第二指示資訊、或第一指示資訊和所述第二指示資訊，所述第一指示資訊用於指示所述第一設備提供所述第

一算力資源，所述第二指示資訊用於指示第二設備提供的第二算力資源，所述第一設備與所述第二設備不相同，其中

所述發送模組，用於向所述第二設備發送算力轉移請求消息，以指示所述第一終端設備的所述算力需求或所述第一設備的算力需求；

所述獲取模組，用於接收所述第二設備發送的算力轉移確認消息，所述算力轉移確認消息中包括所述第二指示資訊，所述第二算力資源是所述第二設備根據所述算力轉移請求消息和所述第二設備的計算面剩餘計算資源確定的。

**【請求項19】** 如請求項18所述的第一設備，其中所述第一設備包括第一基地台，所述第二設備包括第二基地台；

所述發送模組，用於通過 X2 口信令向所述第二設備發送所述算力轉移請求消息。

**【請求項20】** 如請求項18所述的第一設備，其中所述第一設備包括第一基地台，所述第二設備包括核心網設備；

所述發送模組，用於通過 N2 口信令向所述第二設備發送所述算力轉移請求消息。

**【請求項21】** 如請求項19-20中任一項所述的第一設備，其中所述第一設備為所述第一基地台；

所述獲取模組，用於接收所述第一終端設備通過廣播 RRC 信令發送的所述算力請求消息。

【請求項22】 如請求項18所述的第一設備，其中所述第一設備為核心網設備，所述第二設備包括第一基地台和/或第二基地台；

所述獲取模組，用於接收所述第一終端設備通過 NAS 信令發送的所述算力請求消息。

【請求項23】 如請求項18所述的第一設備，其中所述計算面資源的指示資訊包括算力大小、執行時間、算力任務類型、計算模型、計算演算法、執行結果傳輸觸發類型和觸發類型中的一個或多個。

【請求項24】 一種第二設備，包括：

第一接收模組，用於接收第一設備發送的算力轉移請求消息，其中，在所述第一設備發送算力轉移請求消息之前，所述第一設備接收第一終端設備發送的算力請求消息，所述算力請求消息包括計算面資源的指示資訊，所述計算面資源的所述指示資訊用於指示所述第一終端設備的算力需求；

第一處理模組，用於根據所述算力轉移請求消息和所述第二設備的計算面剩餘計算資源確定第二算力資源；以及

第一發送模組，用於向所述第一設備發送算力轉移確認消息，所述算力轉移確認消息中包括第二指示資訊，所述第二指示資訊用於指示所述第二設備提供的所述第二算力資源，其中，所述第一設備向所述第一終端設備發送算力配置資訊，所述算力配置資訊包括所述第二指示資訊、或第一指示資訊和所述第二指示資訊，所述第一指示資訊用於指示所述第一設備提供所述第一算力資源，

所述第一設備與所述第二設備不相同。

【請求項25】 如請求項24所述的第二設備，其中

所述第一接收模組，用於接收所述第一設備通過 X2 口信令發送的所述算力轉移請求消息。

【請求項26】 如請求項24所述的第二設備，其中

所述第一接收模組，用於接收所述第一設備通過 N2 口信令發送的所述算力轉移請求消息。

【請求項27】 一種第一終端設備，包括：

記憶體，用於儲存電腦可讀指令；

還包括，與所述記憶體耦合的處理器，用於執行所述記憶體中的所述電腦可讀指令從而執行如請求項 1 至 4 中任一項所述的方法。

【請求項28】 一種第一設備，包括：

記憶體，用於儲存電腦可讀指令；

還包括，與所述記憶體耦合的處理器，用於執行所述記憶體中的所述電腦可讀指令從而執行如請求項 5 至 10 中任一項所述的方法。

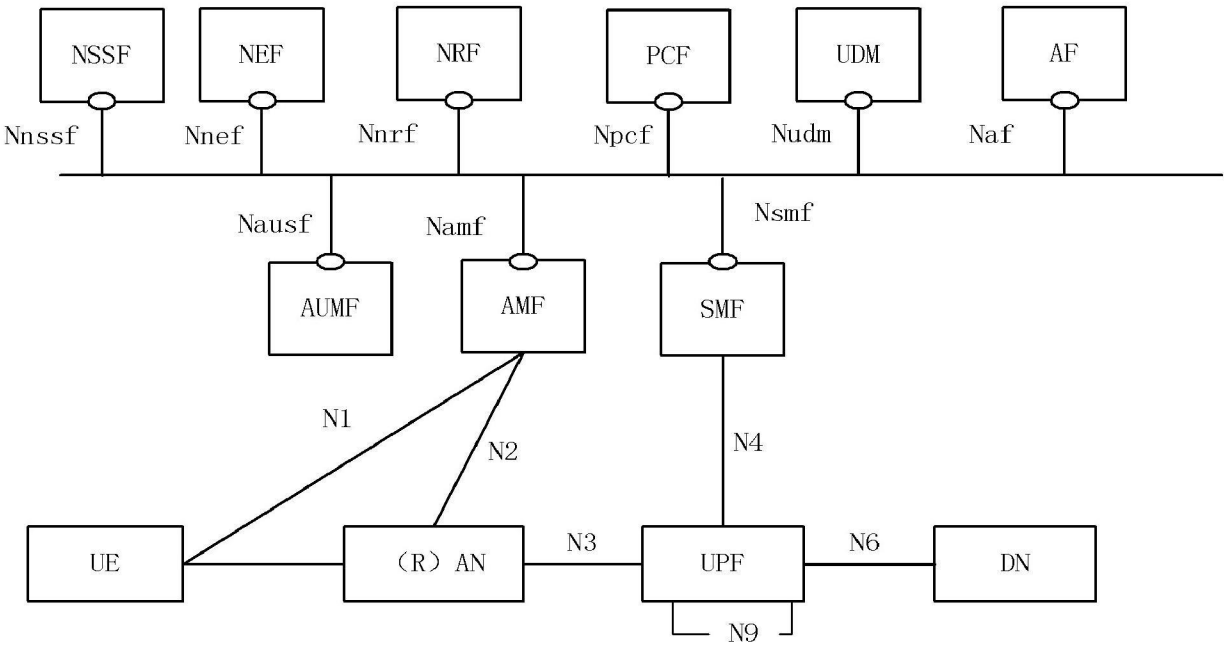
【請求項29】 一種第二設備，包括：

記憶體，用於儲存電腦可讀指令；

還包括，與所述記憶體耦合的處理器，用於執行所述記憶體中的所述電腦可讀指令從而執行如請求項 11 至 13 中任一項所述的方法。

【請求項30】 一種電腦可讀儲存媒體，其中當指令在電腦上運行時，使得所述電腦執行如請求項1至4、或5至10、或11至13中任一項所述的方法。

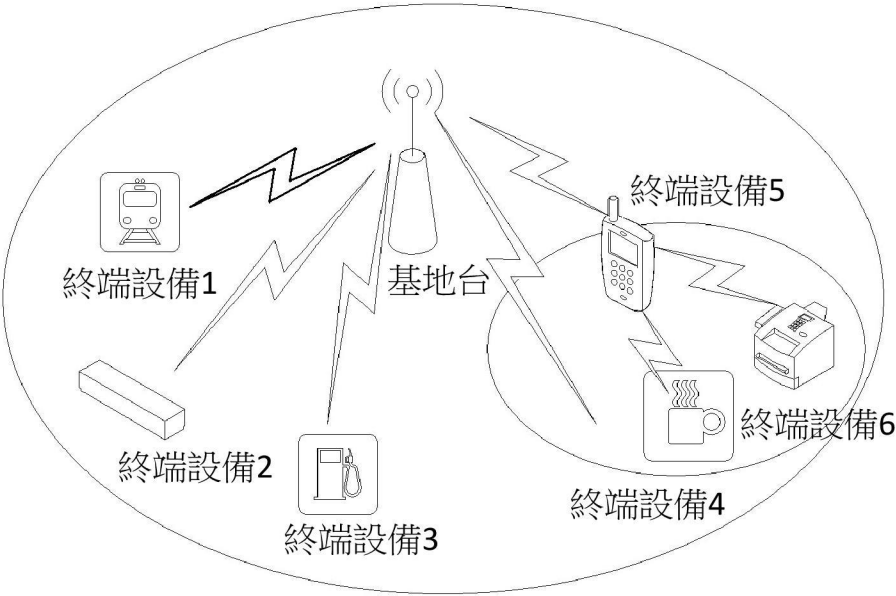
【發明圖式】



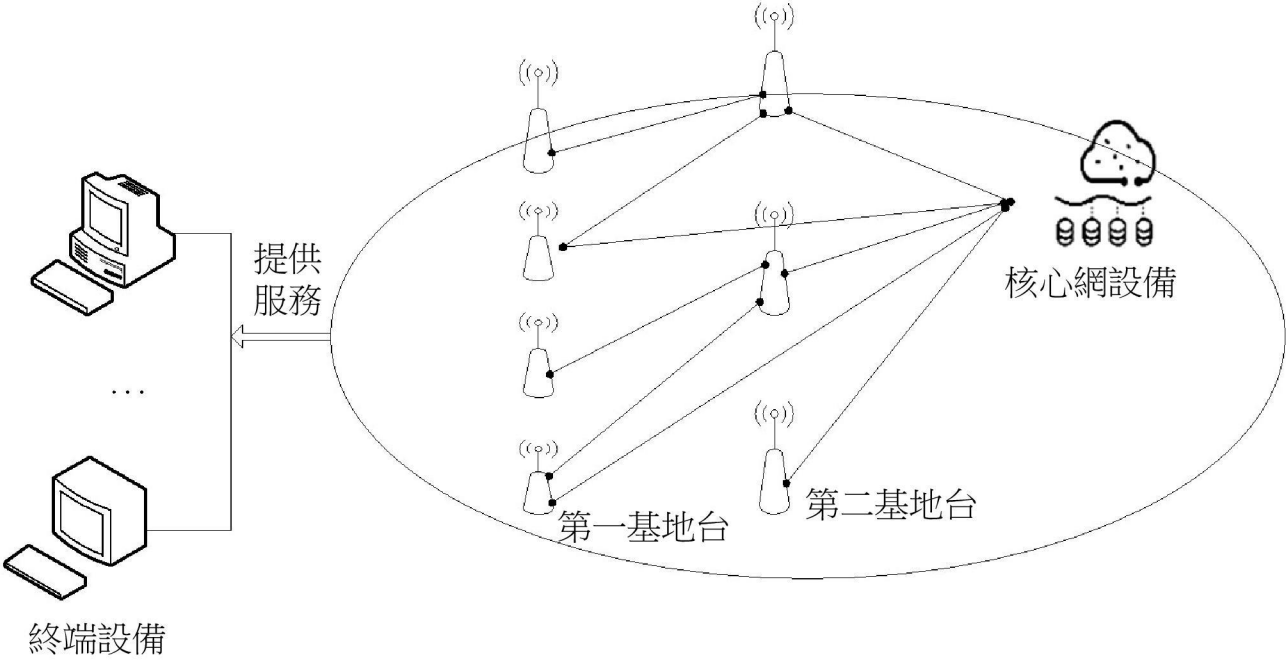
【圖1a】



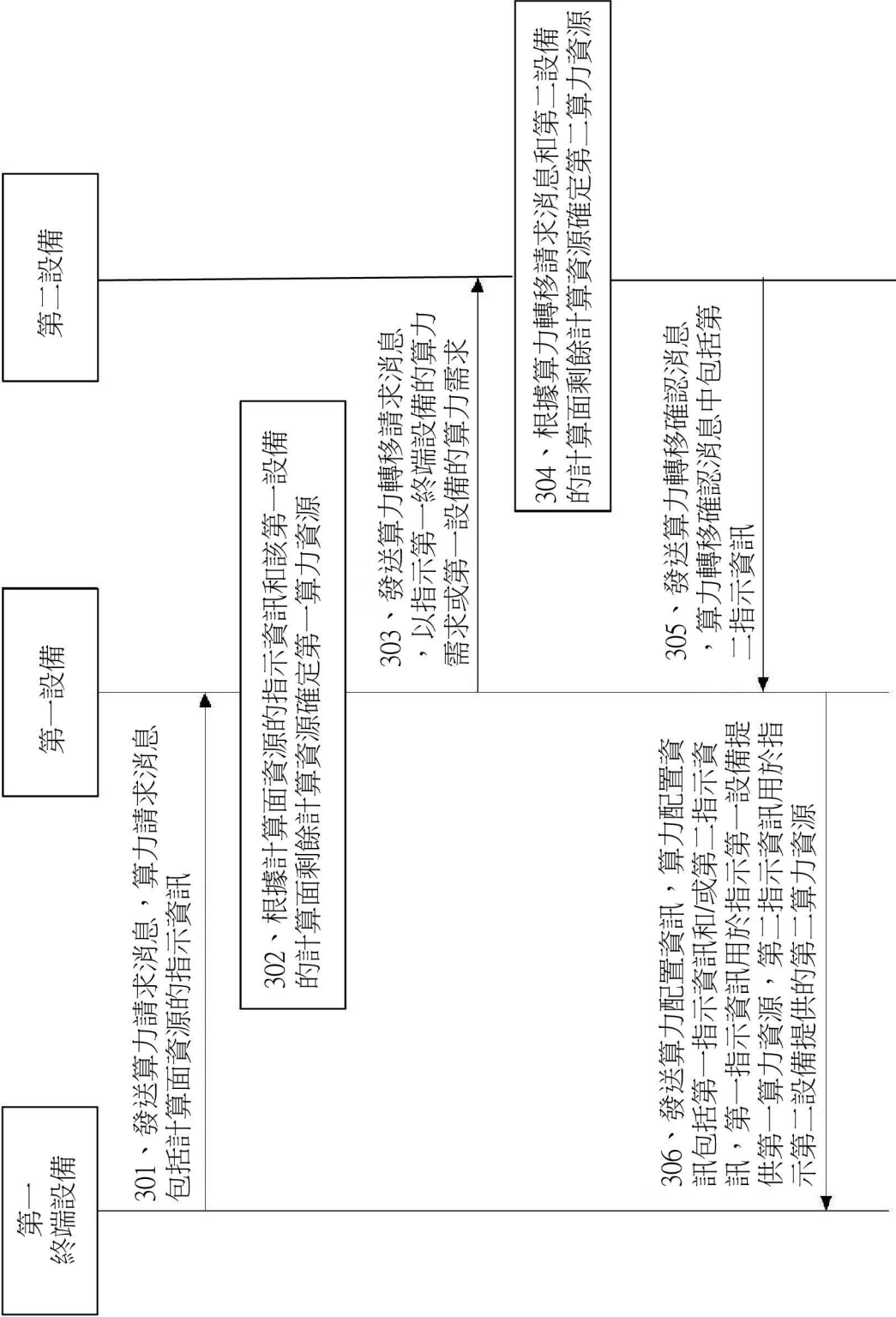
【圖1b】



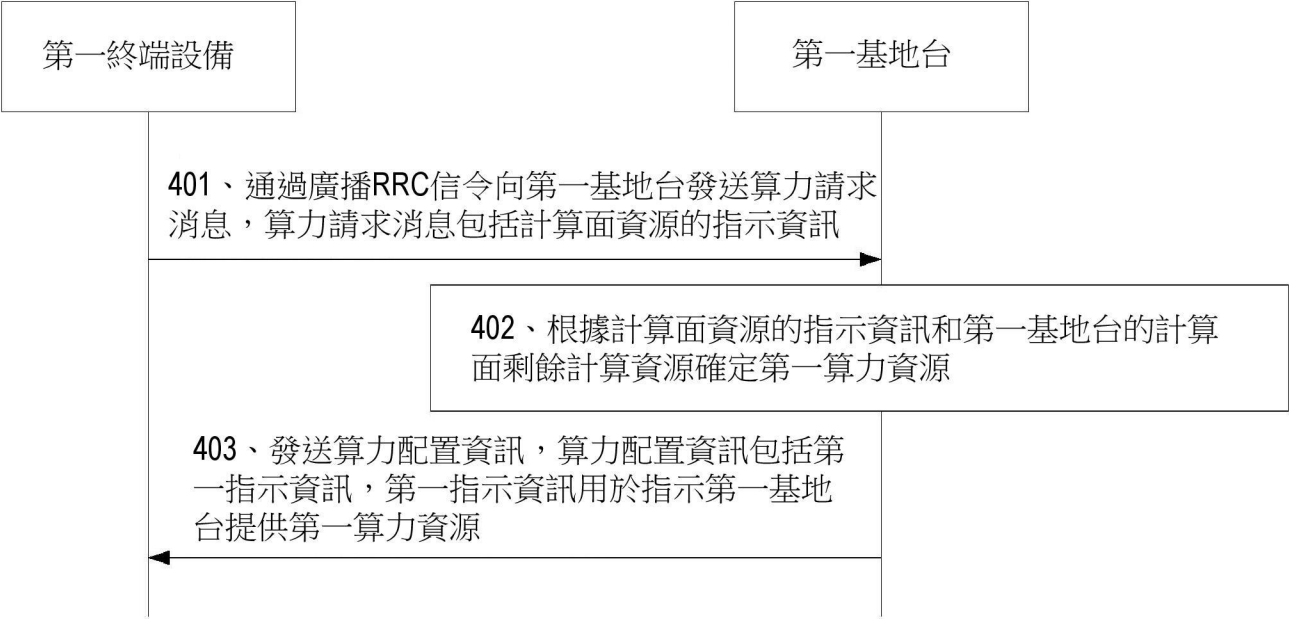
【圖2a】



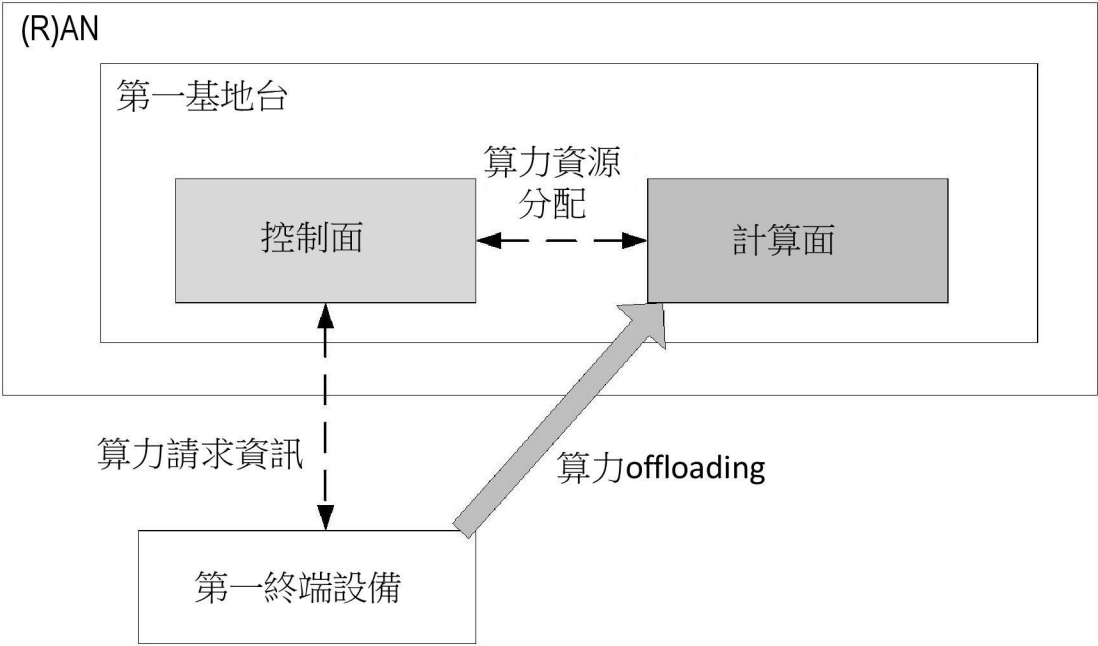
【圖2b】



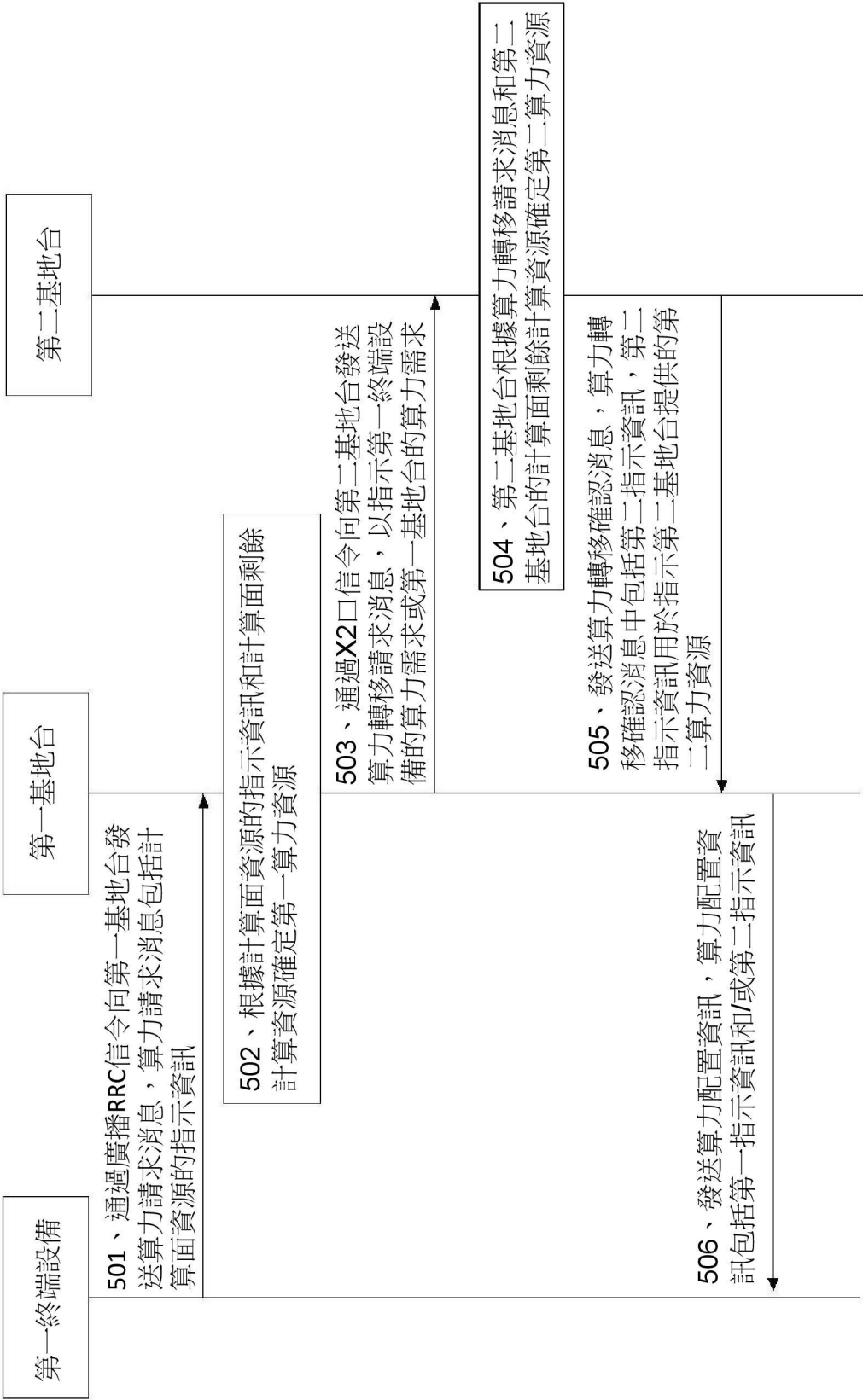
【圖3】



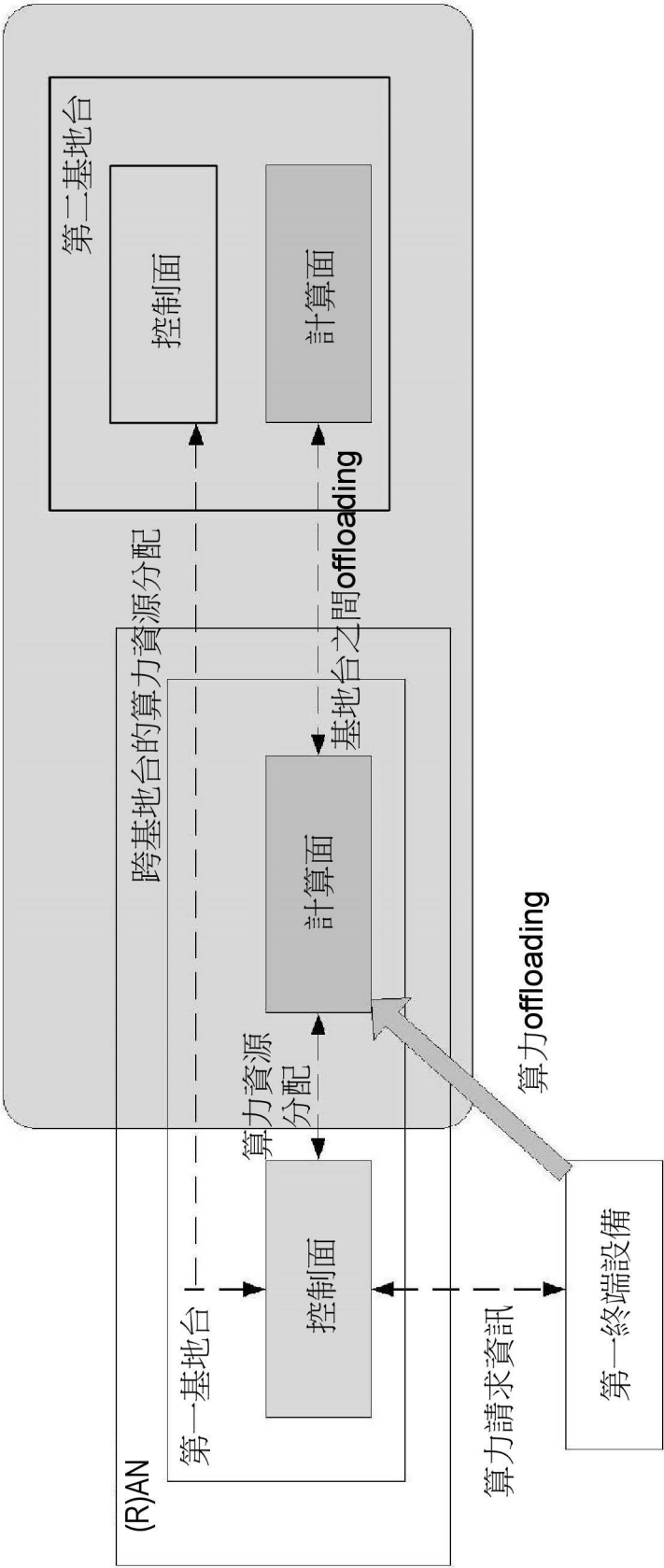
【圖4a】



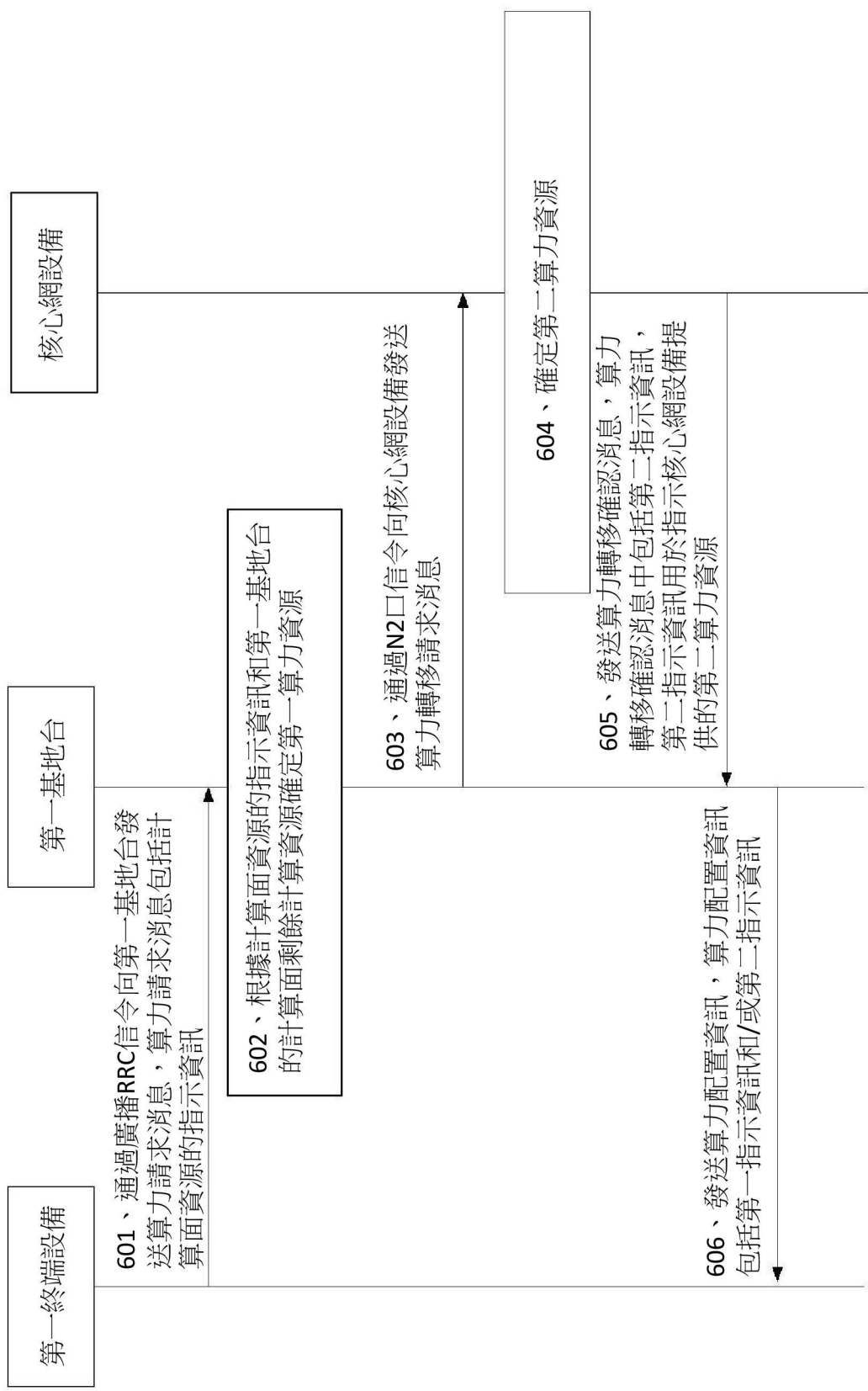
【圖4b】



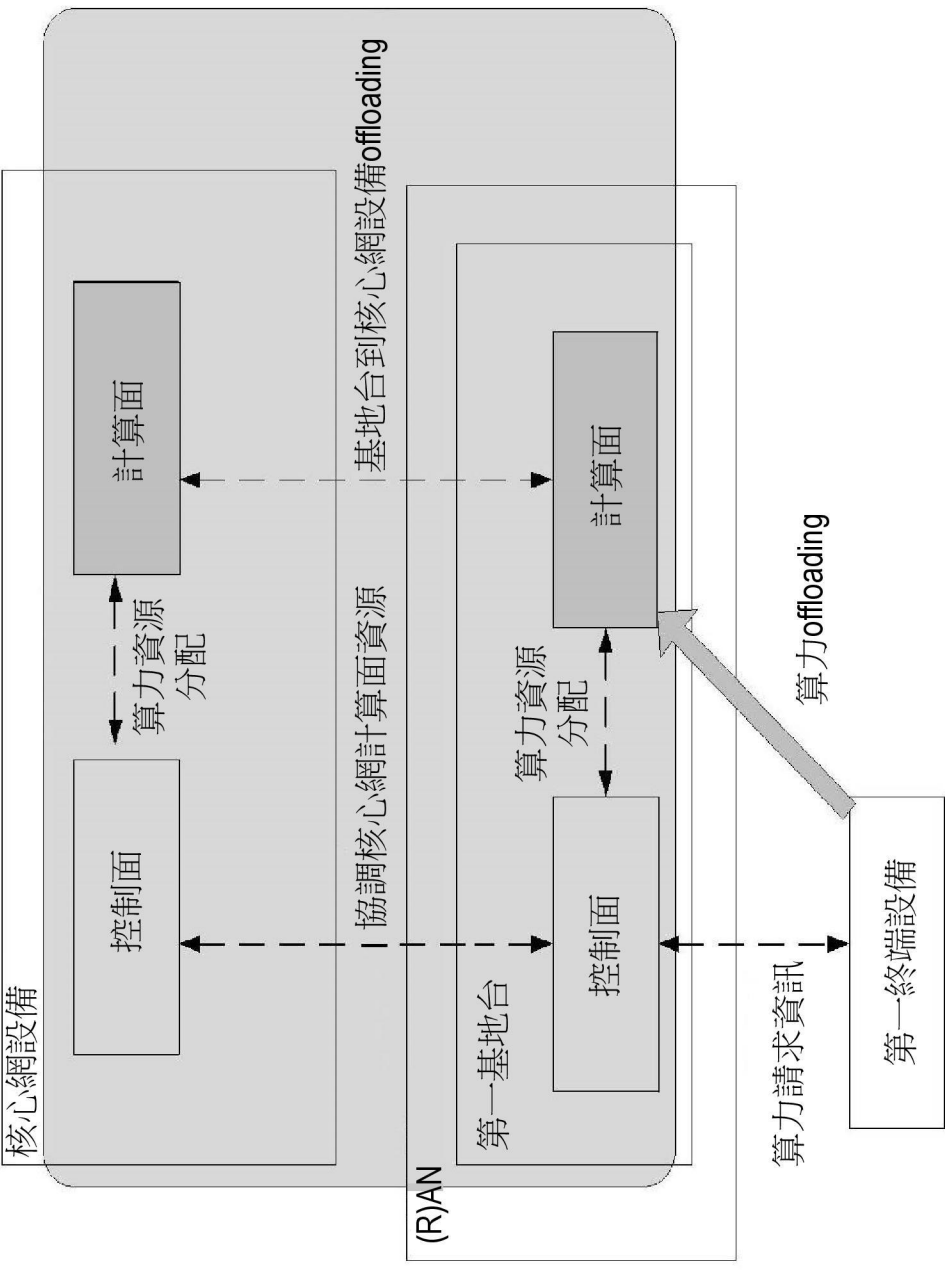
【圖5a】



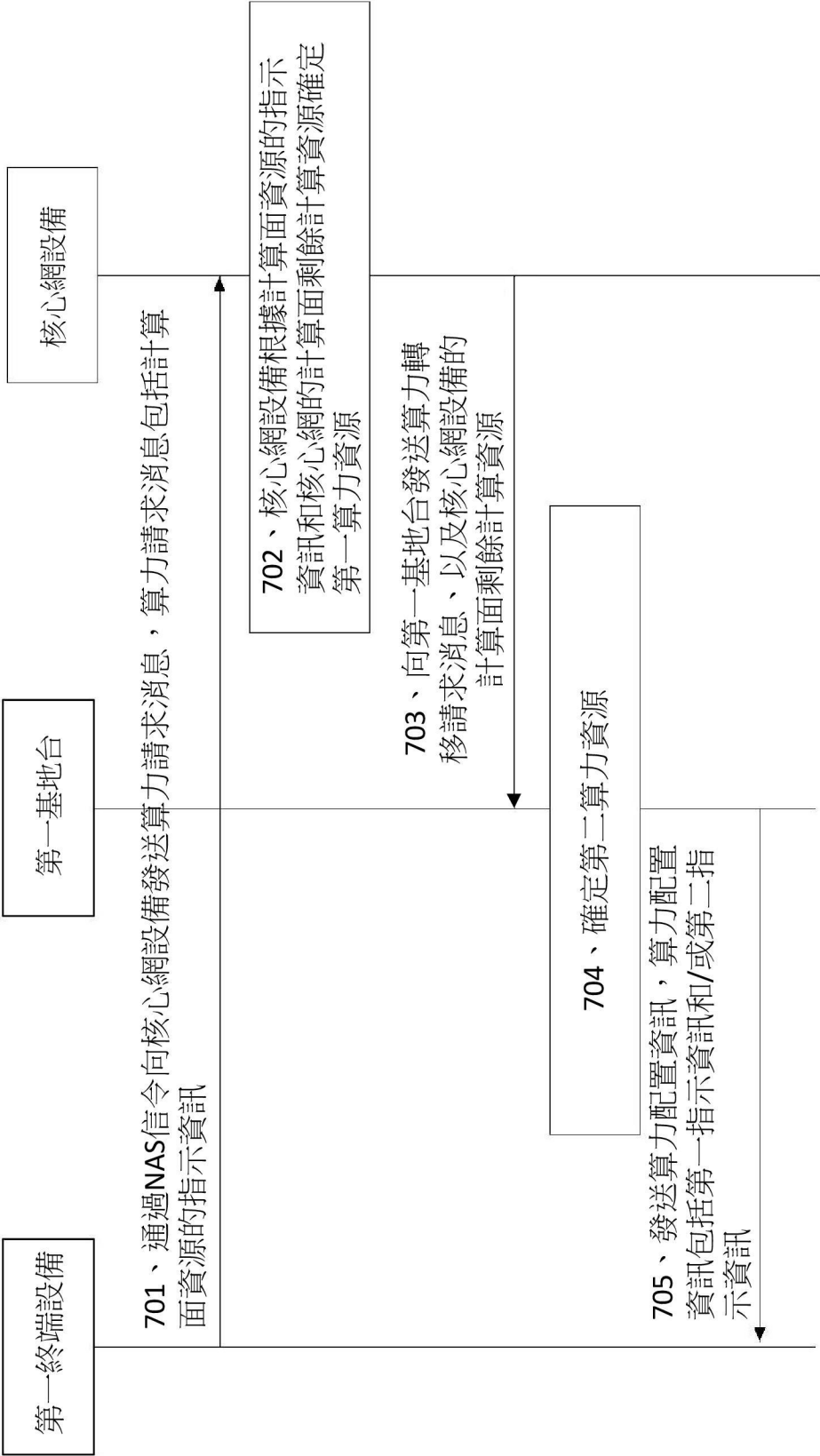
【圖5b】



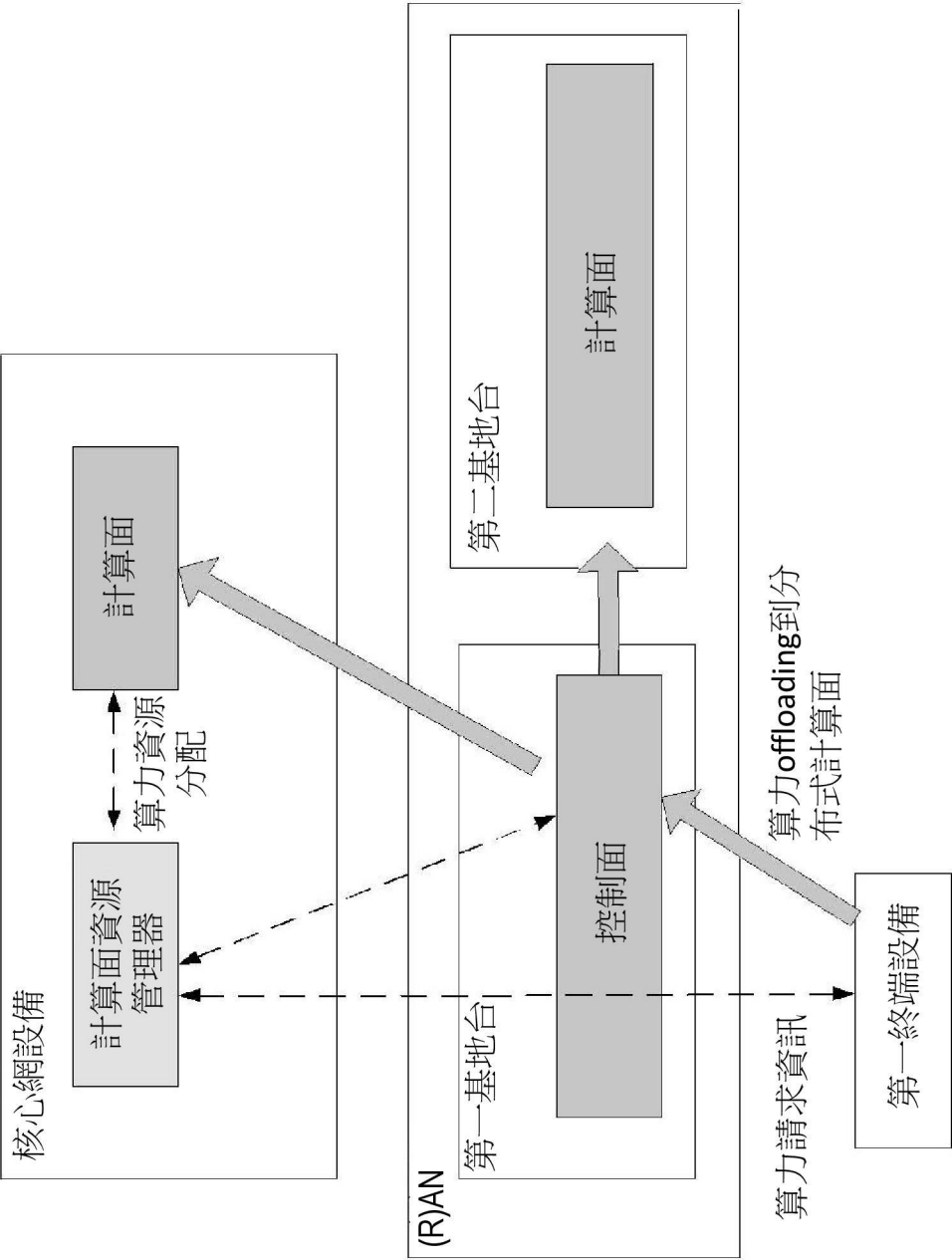
【圖6a】



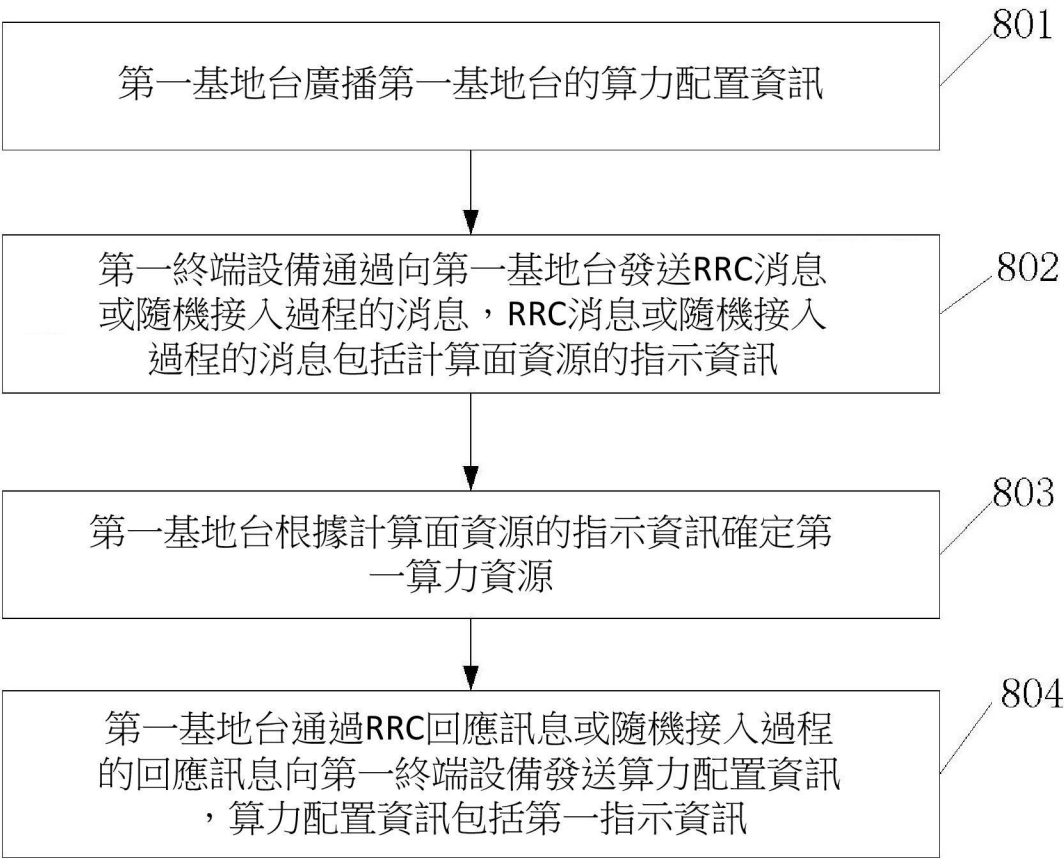
【圖6b】



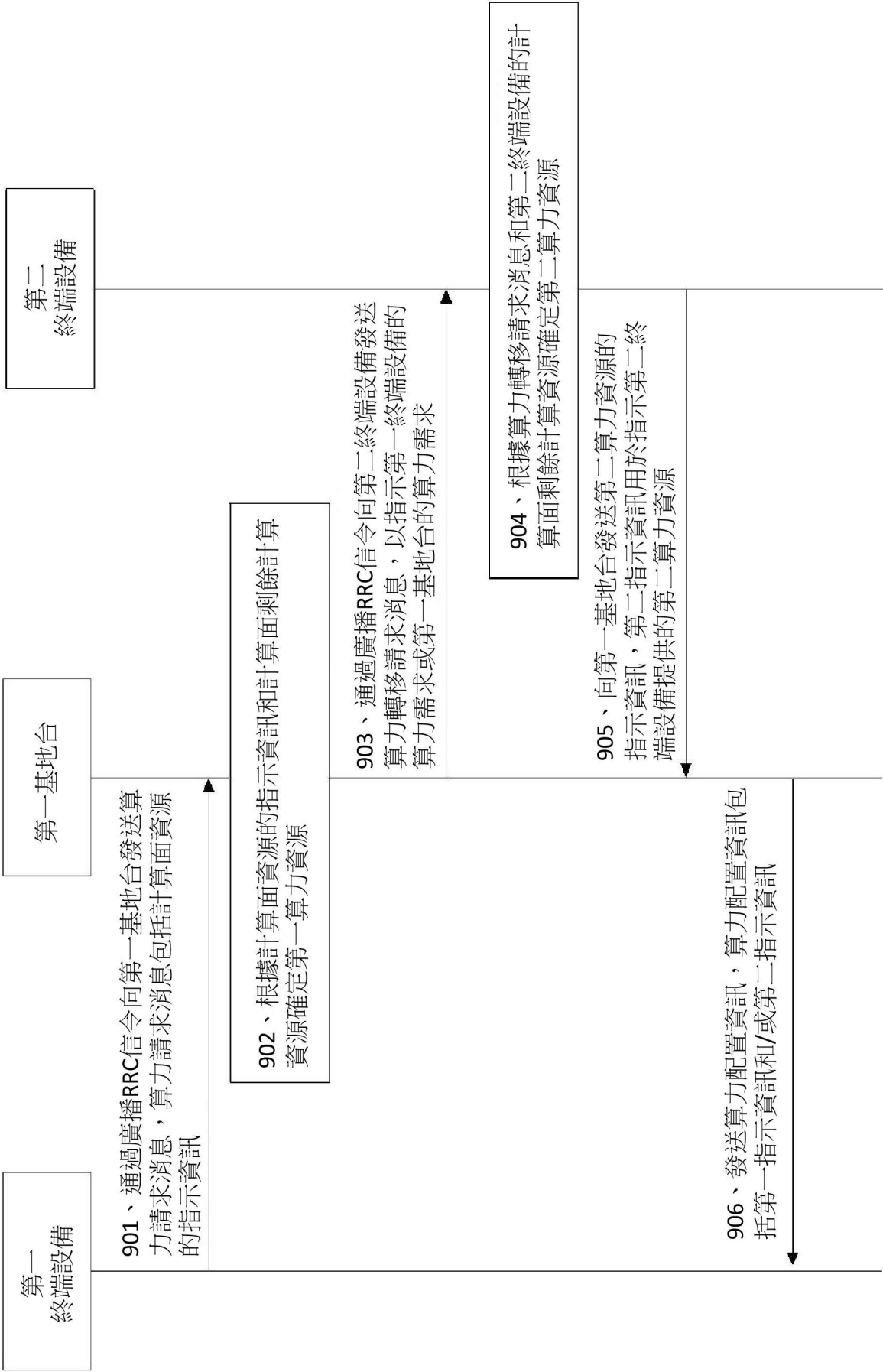
【圖7a】



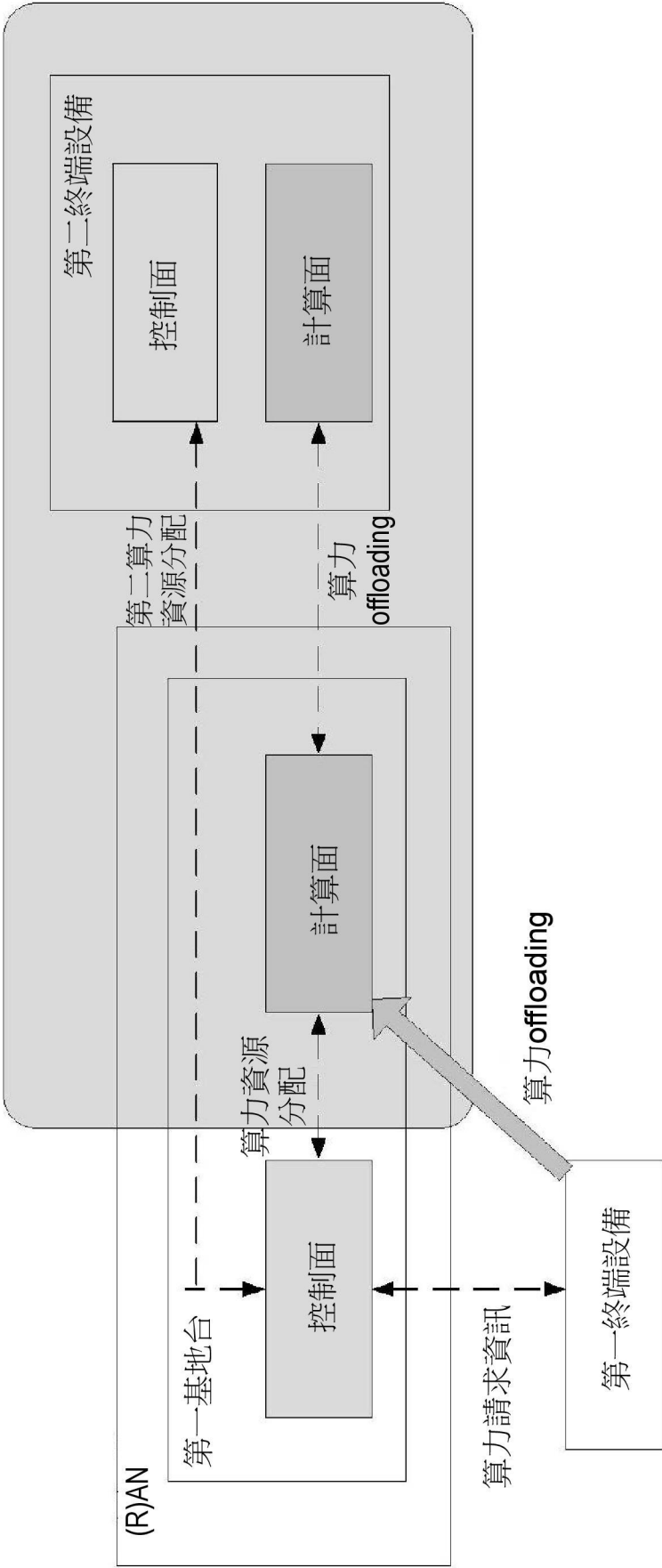
【圖7b】



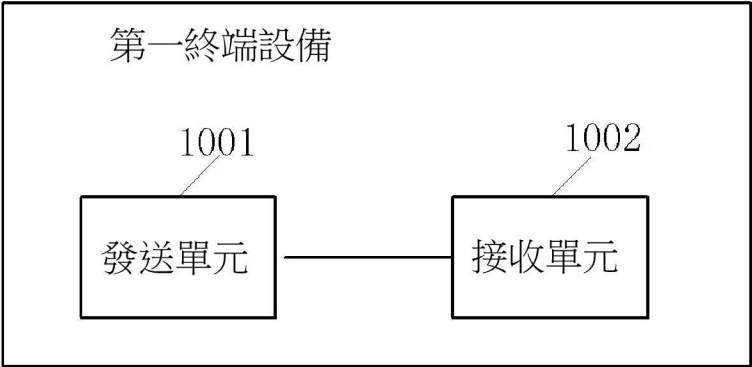
【圖8】



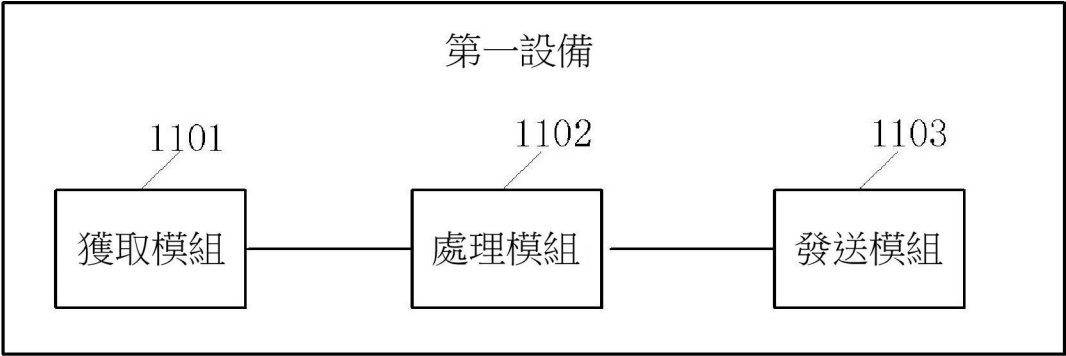
【圖9a】



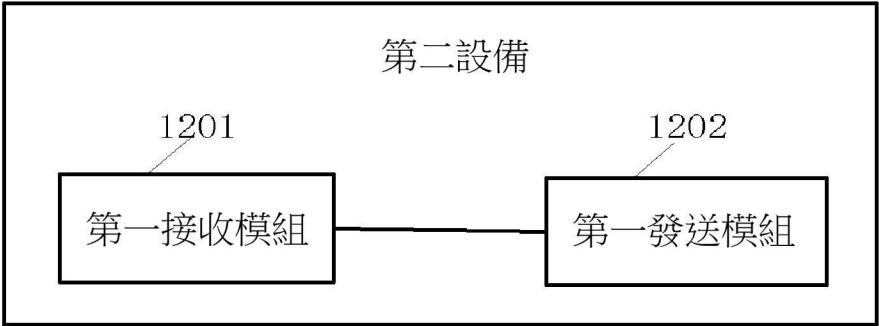
【圖9b】



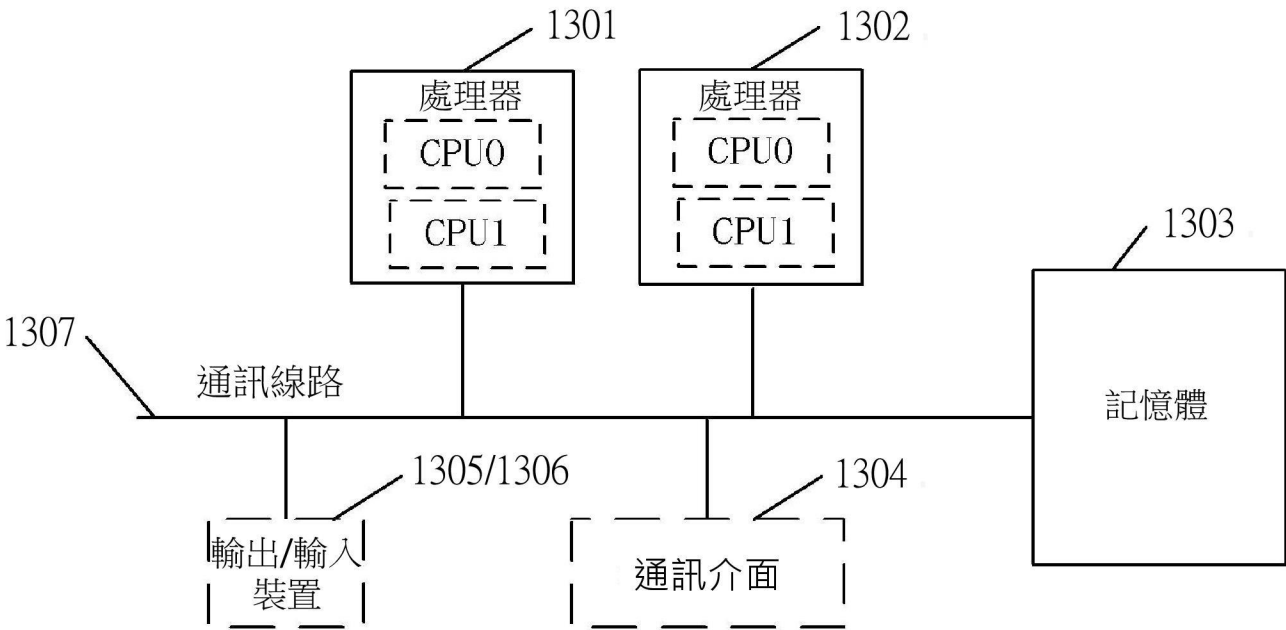
【圖10】



【圖11】



【圖12】



【圖13】