

CPI-ENABLED LOGISTICS INFRASTRUCTURE INTEGRATION FRAMEWORK IN THE GREATER BAY AREA

Hang Wu^a, Linhao Huang^b, Ming Li^{a,c,*}, George Q. Huang^{a,c}

^a. *Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, China*

^b. *School of Management, Jinan University, Guangzhou, Guangdong, China*

^c. *Research Institute for Advanced Manufacturing, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, China*

* Corresponding author: Ming Li at Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, P. R. China, ming.li@polyu.edu.hk, Tel: +852 27664596, fax: +852 27664596

ABSTRACT

The emerging digital economy has facilitated the digital transformation of the logistics industry toward a standard and collaborative state. The logistics infrastructure has accordingly attracted much attention as the basis for supporting digital economy implementation. Moreover, the initiatives of regional collaborative development, such as in the Greater Bay Area (GBA) of China, have accelerated the integration of logistics infrastructure and operation-level collaboration. Fortunately, the cyber-physical internet (CPI) has created new opportunities for promoting infrastructure integration through the cyber, physical and internet dimensions. However, regional infrastructure integration in the GBA faces various challenges associated with CPI network operations. Thus, a bottom-up approach is proposed to facilitate logistics infrastructure integration in the CPI environment based on the conceptual fusion of computer networks. Considering the fundamental difference in network fusion and operation mechanisms between CPI and computer networks, a CPI- LI^2 (logistics infrastructure integration) framework is designed as an integrated solution. Finally, a simulation study is conducted using real-life medical logistics data from the GBA to verify the

effectiveness of the CPI-LI² framework. The experiments show that the framework provides a high infrastructure utilization rate and excellent load balancing ability.

Keywords: Cyber-physical internet; Infrastructure integration; Routing; Transportation

1. INTRODUCTION

A wave of digitalization is sweeping through all industries and the era of digital economy is coming with the well-digitalized infrastructure. The logistics industry has embraced this wave and constructed smart logistics infrastructure. In particular, the emerging cyber-physical internet (CPI) has promoted the integration of logistics infrastructure for regional collaborative development and created a new development direction in the logistics industry, especially in economic-centric areas worldwide, such as the Greater Bay Area (GBA). In recent years, the rapid development of e-commerce, especially cross-border e-commerce, has triggered changes in the manufacturing industry and greatly contributed to the growth of logistics demand in the GBA. The well-established logistics infrastructures in the GBA cannot meet high-growth logistics demands. Logistics infrastructure integration seems to be a better way to address this problem compared with building too many infrastructures for traditional logistics patterns because CPI can promote infrastructure resource sharing and load balancing through information flows configured at the cyber level [1]. The National 14th 5-Year Plan and the Outline Development Plan for the Guangdong-Hong Kong-Macao Greater Bay Area also note that logistics infrastructure integration is one of the most important steps in the integrative construction of the GBA. The application of CPI will greatly accelerate this process.

As the foundation of CPI, the physical internet (PI) provides an open logistics system based on shareable physical infrastructure, consensual protocols and digital interoperability [2]. PI sends and receives goods in the physical world as people send and receive messages on the internet. To better extend this concept to cyberspace, CPI strengthens the synthetic use of cyber capacity to optimize the operation and execution of PI. In this way, the locations and states of PI containers in the physical world are collected, communicated, and synchronized in real time. Furthermore, the flows of PI

containers are configured, orchestrated, and supervised through the flows of information in cyberspace initially.

The relevant research on CPI can be divided into PI-level research and cyber-physical-level (CP-level) research. Most of the existing CPI research has been at the PI level and has mainly been conducted in Europe. Preliminary results have been achieved in research on PI, with a focus on the PI backbone network, such as the standardization of PI nodes [3], [4], the mechanisms and optimizations of the operations and services of logistics networks [5], [6], and the expansion and maintenance of PI logistics networks [7]. Over the past few years, many scholars have combined and enhanced information and communication technologies with applications in CP-level research (e.g., cloud computing and digital twins), such as in cutting-edge applications of CPS in the logistics sector [8], [9], the cyber layer optimization of CPI [10], and the data sharing or interoperability services of CPI [11]. However, the existing research on CPI infrastructure integrative operation still has some limitations, which are summarized as follows.

(1) The existing CPI research has been generally conducted in a top-down manner, from CPI concept design to proof-of-concept to program implementation, for the overall control and management of CPI networks. These studies typically depend on the introduction of new infrastructure rather than enhancing the use of existing infrastructure. This paper adopts a bottom-up perspective to fuse infrastructure and establish an integrated logistics infrastructure network based on the existing logistics infrastructure in the GBA.

(2) Most studies have designed the operational mechanisms of large-scale networks from the perspective of operation research using combinational optimization methods. However, these methods cannot provide exact solutions in a limited time for large-scale network integration problems because large networks face more uncertainties and disturbances during problem abstraction and mathematical modeling than do small networks, and they require a comparatively enormous solution space, which makes it difficult to solve the algorithms and obtain an exact solution in a limited time window. Computer networks can handle massive sets of real-time data with a simple routing process.

Introducing the concept of computer networks may enhance the real-time performance of integrated infrastructure networks.

Since computer networks have faced similar integration problems, this paper explores a feasible migration approach for general conceptual methodologies to CPI by addressing the two research questions:

(1) How can the general features of computer network fusion be migrated to CPI? There is a fundamental difference between a completely virtual computer environment and a CPI-based virtual-real environment. Some factors hinder infrastructure integration in CPI, such as the connectivity among infrastructures. Thus, the infrastructure integration method for computer networks cannot be applied directly to CPI infrastructure networks.

(2) How can a novel routing mechanism be designed for the integrated CPI with the motivation of data package routing in computer networks? Path evaluation is an initial step in routing. The types and characteristics of paths are quite different between CPI and computer networks. Therefore, an innovative path evaluation mechanism for CPI should be investigated, and specific routing mechanisms should be designed to enable flexible PI routing in CPI.

To address the above problems, a CPI-LI² (logistics infrastructure integration) framework is proposed for the integrated infrastructure network in the GBA. The framework includes a two-layer structure with a link layer and a transport layer. For the fusion of the link layer, a logistics elemental area fusion (LEAF) approach is proposed in accordance with the openness and synergy of the infrastructure. The LEAF method is used to establish connections between different areas by electing CPI core autonomous areas and regional link centers to form an integrated CPI network. For the fusion of the transport layer, a distance vector master paths (DVMP) routing mechanism and a link state branch paths (LSBP) routing mechanism with different path evaluation mechanisms are developed based on the associated guide prices in the integrated CPI network. Finally, a simulation experiment is conducted to integrate the pharmaceutical logistics network in the GBA and verify the effectiveness and efficiency of the proposed framework from multiple dimensions.

The rest of this study is organized as follows. In Section 2, the literature on CPI and related topics is reviewed. In Section 3, the CPI- LI^2 framework is introduced. In Sections 4 and 5, the specific mechanisms for the link layer and transport layer are presented, respectively. Section 6 illustrates the simulation experiments, which involve the application of the proposed framework to the pharmaceutical logistics network in the GBA. Section 7 concludes the work and gives the major contributions of this paper and future research topics.

2. LITERATURE REVIEW

2.1. Physical Internet

The term ‘Physical Internet’ first appeared in *The Economist* in 2006 [12] as a novel concept for building large-scale complex logistics and transportation networks in the physical world. In 2012, Eric et al. provided a fundamental definition of PI. They considered the PI a multilevel network collaboration model based on global supply chains and used it to cope with sustainable demand in supply chains [13]. Montreuil et al. provided a more systematic definition of PI in accordance with existing theories. They considered the PI an open global logistics system based on the encapsulation and digital and operational interconnection of physical, interfaces and protocols, creating an evolving system driven by technology, infrastructure, and business innovation [2]. In recent years, Ray Y. Zhong et al. noted that PI has become a new logistics paradigm with the integration of cloud computing and big data analysis [14]. Additionally, Eric K.H. Leung et al. claimed that PI has enhanced the visibility and traceability of logistics processes through the application of digital twins. They suggested that PI will be the most popular logistics model in the future [15].

2.2. PI Framework

The PI framework primarily involves the interactions among different layers and different functions. In research on the interactions among different layers, Wooyeon et al. proposed a PI framework based on several Ethernet protocols, and it improved the real-time services between

physical and network layers [16]. Lee et al. built a 5C framework and clarified the key role of digital interoperability in establishing an interconnection between the PI network layer and the physical layer [17]. In addition, Fahim et al. proposed a novel PI framework based on DSR theory for vehicle routing in PI ports, which achieved the interactions and connections between the network layer and various physical modules in PI [18].

In terms of the functionality of the PI framework, Amitangshu et al. proposed a fresh food PI framework (FFPI or F^2). This framework reinforced collaboration and information sharing among different PI infrastructures [19]. Thibaut et al. focused on improving PI intelligence in decision-making in an unstable environment. They established a PI atomic framework to enhance the resilience and robustness of the airline supply chain [20]. Moreover, Mansour et al. constructed a PI cloud framework to enhance the network edge capacity. This framework displayed good performance for predicting disruptions in logistics processes [21].

2.3. PI Routing

Urban routing is the most basic routing in PI logistics networks. Pach et al. employed a predictive and reactive routing mechanism to solve the perturbation problem in urban routing and prevent routing deadlock based on a logical cross-docking strategy [22]. Eric K.H. Leung et al. used a mathematical model and synchronization strategy to satisfy the demand for continuous flows in urban logistics, and duplication and crossover were reduced among urban transportation routes [15].

Intercity routing is associated with a higher level of service scope than city routing. Gontara et al. developed a PI-BGP protocol to explore the optimal routing between intercity PI nodes. The results of this study indicated that the speed of PI container flows in intercity areas increased over time [23]. Pan et al. employed a collaborative vehicle routing perspective. The CVRP model was adopted to solve the intercity routing problem considering the fragmentation of supplier-customer connections [24]. Yoon et al. used a machine-learning approach to explore the correlation between PI system performance and routing constraints. Moreover, the results of their study indicated that increasing the

on-demand mobility of PI containers would reduce the uncertainty of dynamic routing [25]. Anirut et al. proposed a combination of MILP and heuristic models for vehicle routing problems to establish feasible routes, effectively avoid empty vehicle trips and reduce logistics costs [26].

The largest range of services is provided by international routing. Matray et al. developed a routing method for problems characterized by curvature and asymmetry. They embedded IP address services into the PI physical space to increase the stability of international routing [27]. Sarraj et al. defined a freight protocol for merging container flows, which effectively reduced carbon emissions. In addition, the protocol of their study can be adapted to applications in the automotive industry [28]. Steffen et al. designed a routing pattern based on protocols in a digital network. This research solved the problem of identifying the next-most-appropriate PI nodes for onward transport based on real-time traffic data [29].

3. GENERAL FRAMEWORK

In this study, a CPI- LI^2 framework is proposed for the problem of logistics infrastructure integration in the GBA, as presented in Fig. 1. The framework includes a two-layer network structure comprising a CPI link layer and a CPI transport layer. The CPI link layer is based on a logistics infrastructure integration network, which is used to achieve integrated and persistent management. The CPI transport layer is used to define the pattern of PI container flows and information flows in the link layer (i.e., the routing pattern) and achieve overall integration.

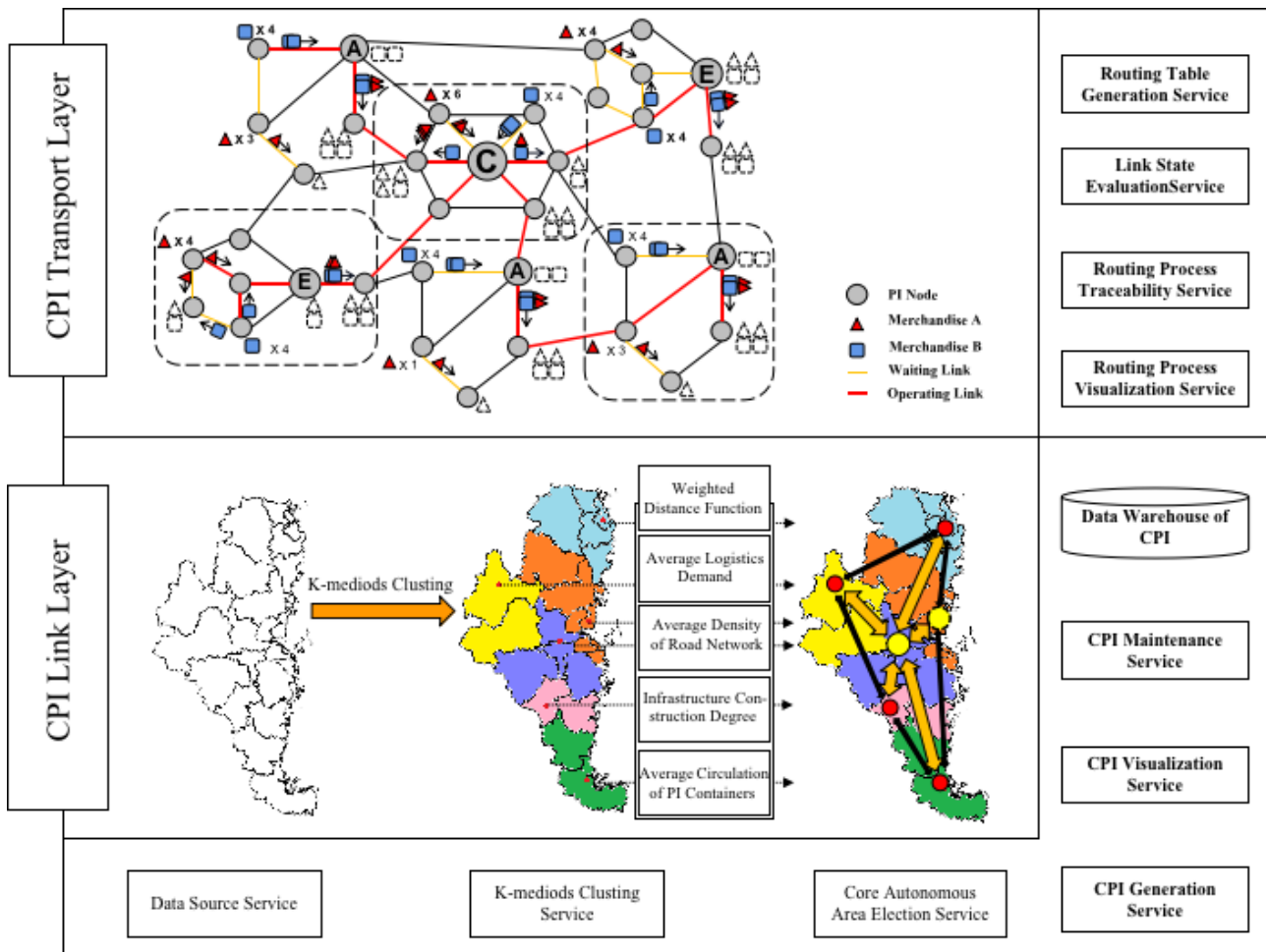


Fig. 1 CPI-LI² infrastructure integration framework

3.1. CPI Link Layer

The CPI link layer is divided into three main phases: link initialization phase, link segmentation phase, and link fusion phase. In the link initialization phase, data are collected and processed, and this step usually involves different subjects or information systems. The role of this phase is to provide the foundation for link-generation work. The link segmentation phase uses the K-medoids clustering algorithm to divide different autonomous areas based on the data provided in the link initialization phase. In the link fusion phase, the road network density, PI container demand scale and other factors are considered when selecting the core autonomous area and the supporting autonomous area from the autonomous areas identified in the link segmentation phase. A weighted distance function is employed to adjust the CPI regional link centers in different autonomous areas. The connections

within and between autonomous areas are established considering the core autonomous area and the CPI regional link centers.

In general, CPI link layer fusion is classified into two types of services as follows. One type of service supports the three phases of link layer generation (e.g., data source, clustering, core autonomous area election and other CPI link generation services). The other refers to maintaining link layer stability after link generation (e.g., CPI link generation, CPI visualization, CPI maintenance and data warehouse services).

The data source service is primarily used to extract the required data from a wide variety of sources and information systems for the generation of a link layer. The data sources largely include information systems and statistics publicly available on official websites, which include basic data and database middleware. Basic data can satisfy the needs of various generation tasks and constitute the foundation of the link layer. Next, database middleware can ensure the real-time updating of data and achieve efficient off-site data synchronization. In general, data source services are achieved by web crawlers and downloading public statistics from official websites.

The clustering service relies on the K-medoids clustering algorithm to partition the link layer into different autonomous areas. The K-medoids algorithm is adopted to divide the CPI link layer into different autonomous areas according to the capacity and distribution of the infrastructure. The division of autonomous areas reduces the pressure on the infrastructure and increases the local fusion speed. This approach avoids shocks to the whole network caused by a change in a small part of the infrastructure.

The core autonomous area election service elects the core autonomous area based on the existing autonomous areas. Adjacency relationships between core autonomous areas and other autonomous areas are used in the regional link layer to form a mesh link structure. The supporting autonomous network accounts for link load balancing and provides disaster recovery support for the core autonomous area. After the two types of autonomous areas are elected, the adjacency relationships between core autonomous and other autonomous areas are prioritized to complete the link fusion

process. In addition, the clustering center, as the CPI regional link center, is adjusted by considering the road network density, infrastructure soundness and PI container demand and scale in autonomous areas. On that basis, considerable duplicate information in the network can be filtered and removed, the database size can be reduced, and the loads on other infrastructure components in the autonomous areas can be reduced. This step plays an important role in shortening the demand response cycle and improving the service quality of the CPI logistics network.

CPI link layer fusion requires corresponding maintenance and management, which is primarily divided into CPI generation, CPI visualization, CPI maintenance and PI data warehouse services. The CPI generation service coordinates all the work in the three phases of network layer generation. The development, completion and articulation of the respective work are strictly controlled by the CPI generation service. The CPI generation service supervises the completion of the respective work to ensure the smooth connection among tasks and finally generates the PI link layer.

The CPI visualization service is used to visualize the operational status of the link layer and provides early warnings regarding abnormal conditions. The visualization service clearly depicts the adjacency and link relationships among autonomous areas, thus enabling the real-time monitoring of the link status in the CPI link layer (e.g., link quality and link abnormalities and failures).

In general, the CPI maintenance service involves the maintenance of link layer data, link status maintenance, and the handling of special situations. If a core autonomous area fails or is overloaded, the CPI maintenance service is responsible for updating the supporting autonomous areas to share resources and keep the link loads balanced. When the failure is overcome, the CPI maintenance service re-establishes the original scheme for the core autonomous area. After service docking between the supporting autonomous areas and the core autonomous area, the supporting autonomous areas return to the initial state.

The CPI data warehouse service provides data extraction, validation, storage, and retrieval functions. The data warehouse extraction function is capable of extracting the basic data and database middleware from the data source service while linking layer operation data (e.g., link state data and

autonomous area management and maintenance data). The above data are sent to the data warehouse in a structured form after data processing to effectively support the protocol operation of the link layer.

3.2. CPI Transport Layer

The CPI transport layer is used to determine the adjacency relationships among the respective PI nodes through the links generated by the link layer. The LSBP routing mechanism for link-state paths and the DVMP routing mechanism for distance vector paths are designed based on the characteristics of different roads in the link layer. The routing mechanism of the transport layer is formed by measuring the guide price based on the quotes of carriers. As a result, PI containers can be transported at high speeds in the transport layer network with numerous nodes and intricate road networks.

Transport layer operation services primarily include routing table generation services, link state evaluation services, routing process traceability services, and routing process visualization services. The routing table generation service is used to determine the neighbors of a respective PI node and establish the adjacency relationship to generate the routing table for the respective node. The generation of the routing table is the primary step in the stable operation of the transport layer. Because the routing table of each node records the paths to the neighboring nodes, path selection should be dependent on the routing table. In a transport network, even if two nodes are not adjacent to each other, they can still be connected through an adjacency relationship. The purpose of doing so is to avoid large node spacing between the origin and destination, thus leading to distortion of information in the transmission process.

The link state evaluation service is based on the path information recorded in the routing table. The guidance price for link state paths is based on the ladder offers of different carriers and the LSBP routing mechanism. The guidance price for distance vector paths is based on the unit distance offers of carriers using the DVMP routing method. Both routing methods calculate the path overhead based on the guide price because this price reflects the carrier's assessment of the path state and the

endogenous and exogenous factors affecting the path state. When a path is unavailable, the link state evaluation service sets the path overhead to infinity, such that the path cannot be selected as the next-hop path.

The routing process traceability service is used to perform real-time tracking in the PI container transportation process through CPI infrastructure. The service satisfies all kinds of management requirements of the CPI transportation process and forms a whole process information chain with human-machine related information coupling. Specifically, the people involved are order initiators, receivers, and handlers. Machines involve PI containers, robotic hands, identification readers, etc. In addition, the focus of transportation process traceability is the PI container. The type, quantity and quality of PI containers are monitored in real time using sensing technology and identification technology.

The routing process visualization service involves link state visualization, PI node visualization, vehicle visualization and PI container visualization. Link status visualization indicates the available status of the current path, which is illustrated based on the color of the link. For instance, red represents the link that performs transportation tasks, and yellow represents the current next-hop path to perform transportation tasks. PI node visualization can clearly distinguish different functional PI nodes (C nodes, E nodes, and A nodes in Fig.1) and general PI nodes based on the size of the graphic. Vehicle visualization provides feedback regarding the real-time position of vehicles, and it can indicate the maximum carrying capacity of a vehicle and the capacity of the current use. PI container visualization provides real-time feedback on the status and storage environment of goods (e.g., the location and quantity of the goods and the temperature and humidity of the PI container).

Table 1 Notations

Symbol	Description
D	Dataset obtained from data source service
K	The larger number of autonomous areas
$m_k(D, k)$	Function that calculates the median of the k th autonomous area
p_k	An array that contains all data points in k th autonomous area
$S(m, p)$	Function that calculates the sum of squares of the error between the median of the k th autonomous domain and the other data points in the autonomous area
$SS(k, a)$	Function that calculates the accumulative sum of squares of the error for k autonomous areas
p_o	An array that contains data points other than those contained in the k th autonomous area
$ad(m, n)$	Function that calculates the sum of the average distance between the data point corresponding to the median and all the data points in the array
$Sil(a, b)$	Function that calculates the contour coefficient with a, b as parameters
f	The excess capacity cost
v	The ratio of excess capacity when carriers operate in an individual mode
α	The average burden degree
β	The coefficient of excess capacity when carriers operate in an individual mode
L	The original ladder quotes
\bar{F}	The average capacity corresponding to capacity range
v_d	The daily capacity when carriers operate in an individual mode
\bar{v}_r	The average nominal capacity of road r
v_{rm}	The maximum nominal capacity of road r
n	The number of carriers that operates in an individual mode
v_{cm}	The maximum capacity when carriers operate in an individual mode
v'	The ratio of excess capacity when carriers operate in an alliance mode
β'	The coefficient of excess capacity when carriers operate in an alliance mode
v_d'	The daily capacity when carriers operate in an alliance mode
v_{cm}'	The maximum capacity when carriers operate in an alliance mode
$node_list$	An array that contains the coordinates of the PI nodes in the CPI logistics infrastructure integration network with values in latitude and longitude
$path_list$	An array that contains the path ID, path type and the length of the path in the CPI logistics infrastructure integration network
$carrier_list$	An array that contains the carrier ID and the carrier name for each PI node
$carrier_quotation$	An array that contains the carrier quote for the corresponding path in $path_list$
$carrier_capacity$	An array that contains the capacity provided by the carrier for the corresponding path in $path_list$
UG	The uniform guide price gradient * = i : carriers are operating in individual mode, and the excess capacity cost is calculated using Equation (3)
f_*	* = u : carriers are operating in an alliance mode, and the excess capacity cost is calculated using Equation (8)
$Calcost(a, b, c, d)$	Function that calculates the sum of $carrier_quotation$ and f_* . Carriers are operating in union mode when parameter d is zero
$Calcap(a, b, c)$	Function that calculates the transportation capacity provided by the carrier. Carriers are operating in union mode when parameter c is zero
$Acc(a)$	Function that sequentially returns the cumulative value of the elements in a . For example, if $a = (1,1,1,1)$, the result of $Acc(a) = (1,2,3,4)$.
$Fit(a, b)$	Regression function with a and b as parameters

<i>Cap_interval</i>	An array containing the preset capacity interval
<i>Calgra(a, Cap_interval)</i>	Function that returns the price gradient according to <i>Cap_interval</i>
<i>CU</i>	The capacity utilization rate of the logistics network
q_i	The capacity allocated to path i
d_i	The length of path i
C_i	The maximum transportation capacity of path i
<i>LB</i>	The load balance degree of the logistics network
\bar{c}	The average load balance degree of the logistics network
c_i	The load balance degree of path i
<i>raw_data</i>	The raw data are stored as an array, which includes the names of the logistics nodes, the infrastructure ID number, and the customer number and rank
<i>Lng_border_up</i>	The longitudinal upper boundary of autonomous areas
<i>Lng_border_down</i>	The longitudinal lower boundary of autonomous areas
<i>Lat_border_up</i>	The latitudinal upper boundary of autonomous areas
<i>Lat_border_down</i>	The latitudinal lower boundary of autonomous areas
<i>Crawling(A)</i>	Function that crawls the longitude and latitude based on the input A
<i>median(A)</i>	Function that calculates the median based on A
<i>Ffill(A)</i>	Function that uses the previous nonnull value to replace the current null value in A
<i>Score(A)</i>	Function that returns the score based on the rank attribute of A
<i>Correlation(A)</i>	Function that returns the correlation of the attributes of A

4. CPI LINK LAYER FUSION MECHANISM

CPI link layer fusion can be divided into three main processes. First, data are obtained from different information systems and official websites through data source services, and the data are preprocessed to form a usable data point set. Then, based on the data point set in two-dimensional space, the elbow method, silhouette coefficient and Calinski-Harabaz (CH) index are used to obtain the K value, i.e., the number of autonomous areas, and the K-medoids clustering algorithm is used to divide the CPI link autonomous areas. Finally, based on the actual logistics scenario, location factors and other realistic constraints, the core autonomous area is selected, and a distance weighting function is adopted to generate CPI regional link centers.

4.1. CPI Autonomous Area Division

In a real logistics network, the division of the area should consider the effects of many realistic factors, such as the density of PI demand nodes, regional road network density, infrastructure load capacity and other factors. This requires the actual data for various factors obtained from the data source service and the determination of the number of autonomous areas.

In this study, we use the K-medoids clustering algorithm [30] to divide autonomous areas in the CPI network. K-medoids is a classical clustering method that can be used to solve NP-hard optimization problems[31]. Although the increased volume of data will lead to limitations in some aspects of K-medoids, it is still effective for the scale of data corresponding to the experimental environment in this study. It outperformed other clustering methods, such as K-means, CLARANS and DBSCAN, in terms of the silhouette coefficient. We find that when the number of clusters is 6, silhouette coefficients for each clustering method are 0.5383 for K-means, 0.5625 for CLARANS, 0.5717 for DBSCAN and 0.6557 for K-medoids, which indicates that K-medoids are more applicable at the current data scale. Similar to K-means, K-medoids requires the value of K to be set for the number of autonomous areas to be divided. The larger the K value is, the larger the number of autonomous areas and the more difficult the infrastructure integration and management. However, the smaller the K value is, the larger the database size and infrastructure load, the higher the performance requirements of the infrastructure and the lower the security of the network. Given the importance of the number of autonomous areas K in an integrated infrastructure network, the appropriate number of autonomous areas K is determined based on the elbow method [32], silhouette coefficient [33] and CH index [34].

The K values of the autonomous areas determined by the elbow method, the silhouette coefficients and the CH index cannot be equal. Therefore, it is necessary to select a satisfactory K value within the interval of K values obtained using the three methods. When K values obtained using the three methods are different, the value for which the total sum of the squared error is minimized and the silhouette coefficient and CH index are maximized is selected as the K value for autonomous area division.

4.2. K-Medoids Clustering and CPI Region Link Center Generation

The operation process of the K-medoids clustering algorithm is shown in Fig. 2.

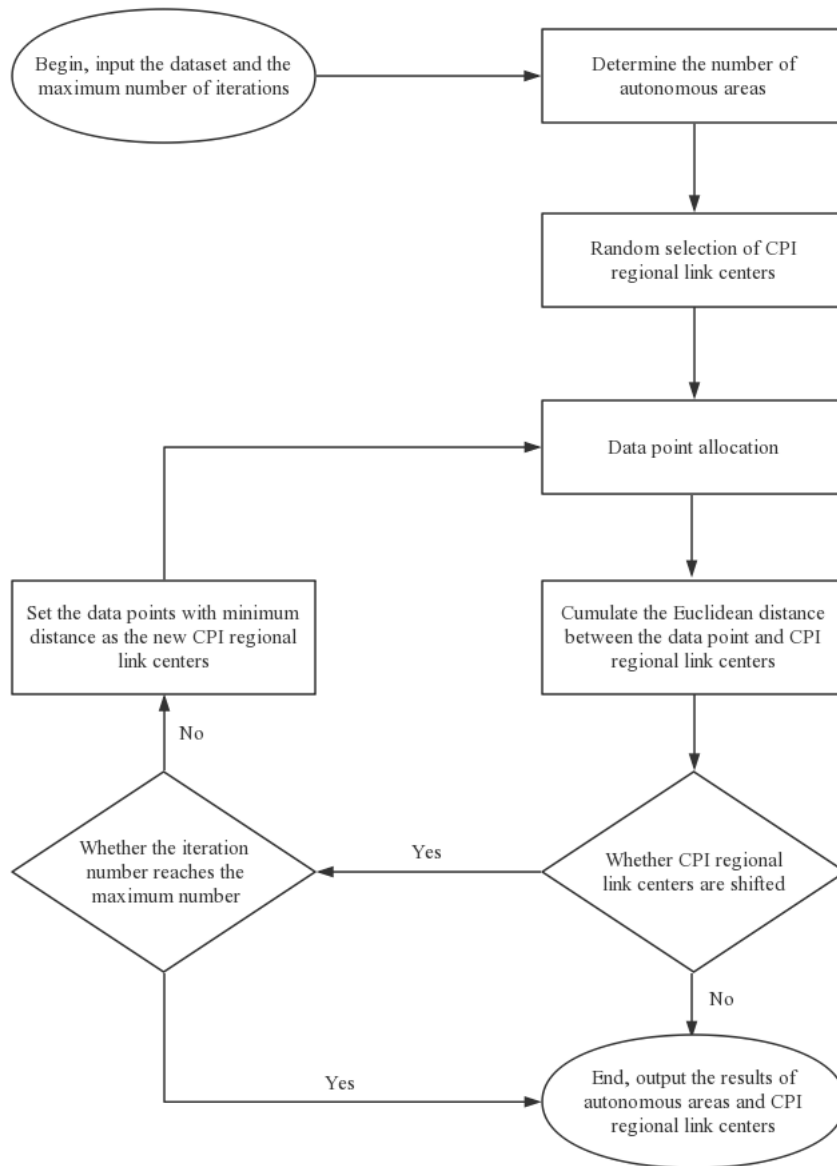


Fig. 2 K-medoids autonomous area division flowchart

In the K-medoids method, the minimum sum of distances between regional centers and other data points is calculated with Equation (1):

$$l_m = \arg \min_{x_m \in C_i} \sum_{x_j \in C_i} \|x_m - x_j\|^2 \quad (i = 1, 2, \dots, k) \quad (1)$$

l_m in the formula denotes the sum of the minimum distance between the central point x_m of the region in cluster C_i and the other data points, such as x_j , within the cluster.

In this study, PI customers are divided into three categories according to the scale of the logistics demand of different CPI customers. The first category is demand-intensive customers, such as some

large e-commerce enterprises. This kind of user has an extremely high logistics demand at almost every moment, and such demand is often diversified and changeable. The second type of customer is characterized by smooth demand, such as some manufacturing enterprises. The third category includes customers with fragmented demand, such as retail stores. This type of customer demand is usually related to the sales volume, and the demand is lower than that for the first two types of CPI customers; moreover, the demand is generally more fragmented and uncertain.

In the logistics process, demand-intensive customer needs are usually satisfied first. Next, the needs of the demand-stable customers are satisfied. Finally, the needs of the fragmented customers are met. In the integrated CPI infrastructure logistics network, demand-intensive customers have the most rigorous requirements regarding the response speed of the network, and a high response speed and long-distance transport are associated with a high-cost investment. As a result, the use of regional link centers can reduce costs, provide faster access to demand-related information, and provide an increased response speed when close to demand-intensive users. Accordingly, the distance formula is applied, and weights are set when calculating the distance between the CPI regional link center and various CPI customers. Specifically, the weights are set as shown in Table 2.

Table 2 Table of customer types and weight

Customer Category	Institution	Weights
Demand-intensive customers	Large e-commerce companies	0.5
Demand-smooth customers	Manufacturing companies	0.75
Demand-scattered customers	Retail stores	1

There are two main purposes for setting weights. One is to make CPI regional link centers as close as possible to PI customers with high demand. Thus, the real-time demands of CPI customers can be met, and the service satisfaction of core customers can be measured. Additionally, large enterprises usually have optimized infrastructure and transportation conditions. The above conveniences provide conditions for CPI regional link centers to maintain certain regional link statuses. The adjusted summed minimum distance is expressed as Equation (2):

$$l_m = \arg \min_{x_m \in C_i} \sum_{x_j \in C_i} \|x_m - x_j\|^2 \cdot \omega \quad (i = 1, 2, \dots, k) \quad (2)$$

5. CPI TRANSPORT LAYER FUSION MECHANISM

5.1. Transport Layer Routing Process

The PI node types in the link layer are classified into functional PI nodes and ordinary PI nodes. Functional nodes are superior to ordinary nodes and are comparatively less abundant. In addition, they are responsible for updating the routing table of the ordinary PI nodes while completing logistics tasks. Functional PI nodes are divided into core PI nodes (C nodes in Fig. 1), edge PI nodes (E nodes in Fig. 1) and auxiliary PI nodes (A nodes in Fig. 1). The core PI node acquires the path information from other autonomous areas through the auxiliary PI node to form the routing table for the whole link layer. Other nodes can be informed of the same path information by accessing the core PI node to support the routing process. There is no auxiliary PI node if the core PI node is included in the autonomous area. Edge PI nodes exist in different autonomous areas and perform logistics tasks between different autonomous areas. The paths connected by edge nodes involve two path types, namely, link state paths and distance vector paths, and edge nodes are included in the link state routing table and distance vector routing table. The link state routing table records the path information for other PI nodes in the domain, and the distance vector routing table records the path information for other connected edge nodes. The edge nodes between autonomous areas are connected by distance vector paths. All nodes within autonomous areas, except those connected by distance vector paths, are connected by link state paths.

The abstract network topology diagram in Fig. 3 is adopted as an example to introduce the routing of PI containers over a wide variety of paths and nodes to effectively illustrate the routing process in the CPI transport layer. Two CPI autonomous areas are included in the figure. TH C Node 1 represents the core PI node, NS E Node 4, HP E Node 5, and GM E Node 6 are edge PI nodes, and the remaining nodes are ordinary PI nodes. Fig. 3 depicts the routing process for a PI container from one autonomous area to another. The PI container is input at YS PI Node 2. YS PI Node 2 queries the real-time routing table (the blue routing table in the figure) obtained from the core PI node TH C

Node 1. Then, YS PI Node 2 learns that the PI container should be transported to GM E Node 6 and finally transported to the destination via GM E Node 6. YS PI Node 2 employs the information in the routing table to select link state path ③ as the optimal next-hop path to transport the PI container to TH C Node 1. TH C Node 1, after receiving the PI container, updates the path information and repeats the path rotation process of YS PI Node 2. Subsequently, link state path ④ is selected as the optimal next-hop path, and the PI container is transported to HP E Node 5. HP E Node 5 transports the PI container to GM E Node 6 through distance vector path ⑦ connected to GM E Node 6. Distance vector path ⑦ has a low path overhead. Furthermore, GM E Node 6 transports the PI container in the same way until the PI container arrives at the destination, and the information regarding the completion of the logistics task is provided to the other PI nodes through TH C Node 1. If link state path ④ becomes insufficient during the transportation of PI containers, the overhead of this path will be infinite, thus causing TH C Node 1 to select another path, such as link state path ⑤.

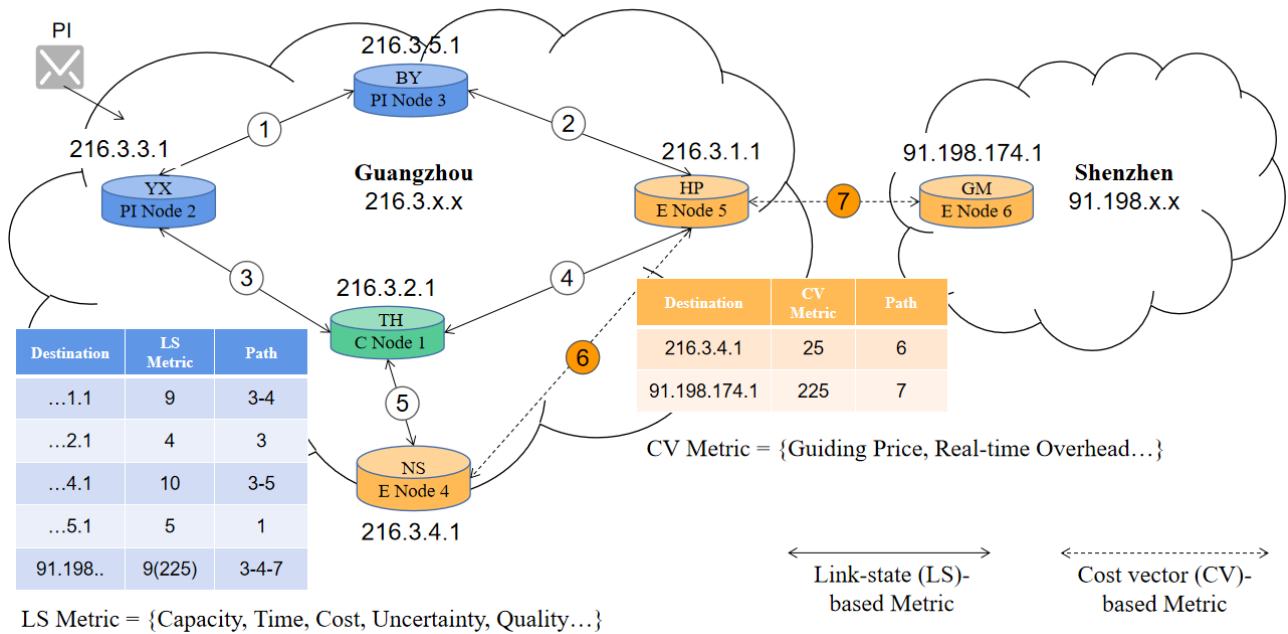


Fig. 3 Example of an abstract network topology

5.2. Transport Layer Routing Mechanism

Accordingly, the guide price of a path can be calculated based on the carrier quote considering path overhead. However, there are different characteristics for paths of different types, and it is necessary to design a suitable guide price calculation method according to the path type. The DVMP routing mechanism and LSBP routing mechanisms are developed for the distance vector paths and link state paths, respectively. The following assumptions are proposed before establishing the transport layer routing mechanism.

(1) The PI containers flowing in the CPI logistics network are all standard boxes with the same weight, volume and type.

(2) Distance vector links can support a sufficiently large capacity, while link state paths have limited carrying capacity.

(3) Each respective path is associated with a carrier quote. The quotes for distance vector links are based on cost per unit distance, and those for link state paths are ladder quotes based on the number of PI containers.

(4) Carriers are quoted only once for the same path, and the carrier quotes remain unchanged during a logistics process.

Fig. 4 presents the flow chart of the transport layer routing mechanism. First, the volume demand and transport endpoint of the PI container should be clarified, as well as the number of paths connected by nodes. Then, the health status of the path is used to determine whether the path can be the next-hop path. For example, road repairs, capacity maximization or road closure would cause the path to be classified as 'unhealthy'. The overhead of the path is then set to an infinite value. Next, a different routing method is applied based on the path type. The node selects the path with the smallest path overhead as the next-hop path. In addition, the path overhead is added to the total path overhead. Moreover, if the next hop is the end of the transportation route, the result, distance and total overhead of this path will be output. Otherwise, the current PI node will be replaced with the PI node of the next hop, and the incoming path will be eliminated. The process is re-executed until the end of

transportation routing is reached. The DVMP routing mechanism and LSBP routing mechanism are described in detail in subsequent sections.

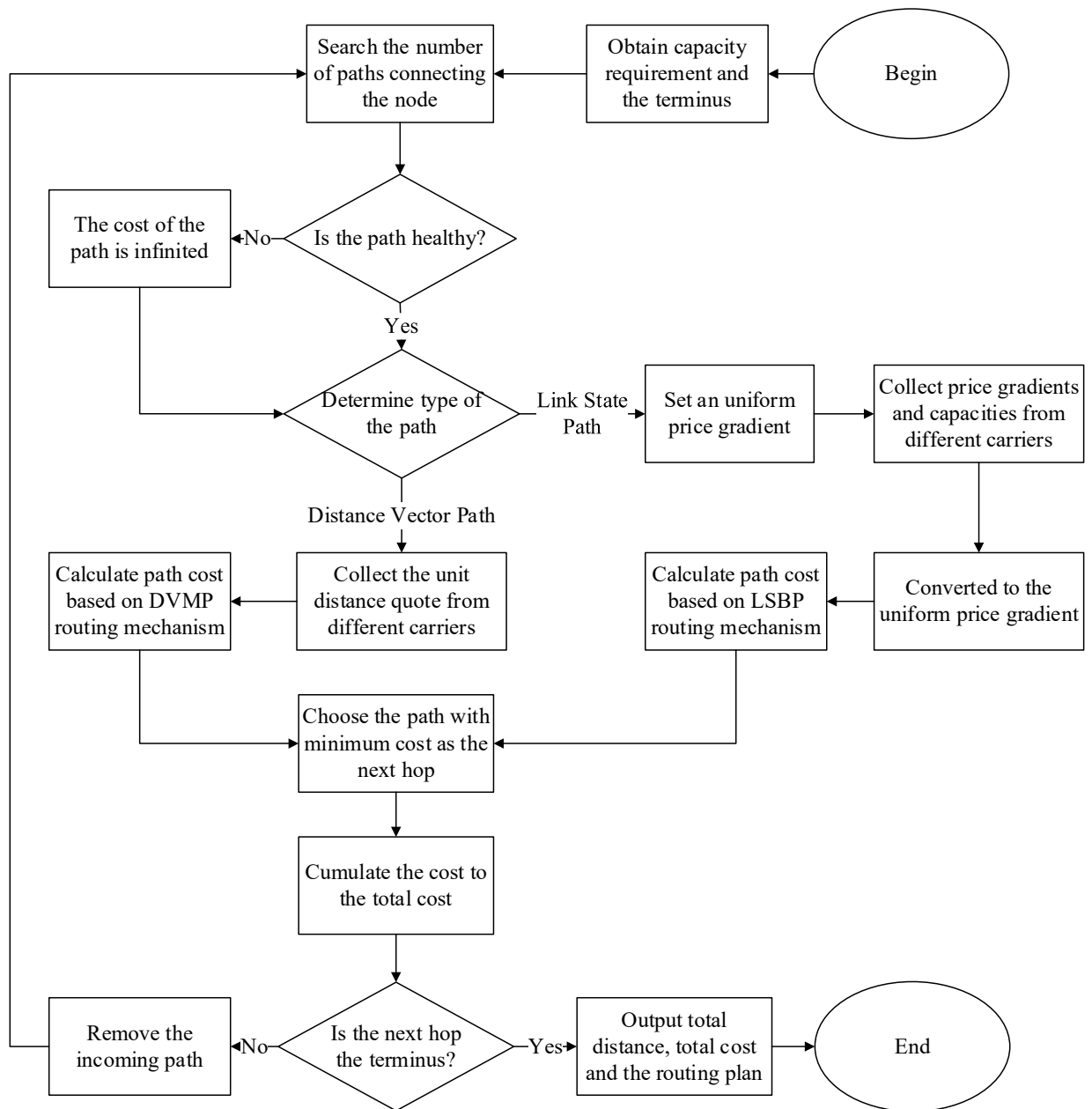


Fig. 4 Flow chart of the routing mechanism

5.2.1 DVMP Routing Mechanism

The DVMP routing mechanism for distance vector paths is simple. Because it is almost a perfect competitive market for distance vector paths. The capacity that distance vector paths can carry is sufficiently large, and each respective path has multiple carriers to provide competitive quotes. Neither carriers nor other logistics participants can affect the quotes provided for distance vector paths

by monopolizing the path capacity. The participants in the corresponding logistics process are all price focused, and the quote only depends on the supply and demand. In these cases, the unit distance is the main factor influencing distance vector path overhead. Carriers use unit distance quotes for distance vector paths and compete for transportation services. In distance vector path routing, first, the unit distance quotes of carriers should be obtained for different paths. Then, they are multiplied by the path distance and the number of PI containers to obtain the guide price for each path. The path with the smallest path overhead is selected as the optimal path, and the lowest guide price in the distance vector path is the corresponding path overhead and added to the total path overhead.

5.2.2 LSBP Routing Mechanism

The LSBP routing mechanism for link state paths is as follows. First, since the price ladders quoted by different link state path carriers are probabilistically different, we convert the ladder quotes of different carriers into a uniform price gradient to facilitate subsequent comparison of guide prices. The algorithm for calculating the uniform price gradient is shown in Algorithm 1. After a uniform price gradient is obtained, the excess capacity costs are calculated considering the operating status of the relevant carriers based on the volume demand. The price gradients are compared, and the lowest price gradient is selected in the respective interval to form the final guide price gradient. The capacity demand of each link state path is multiplied by the corresponding price gradient and path distance to obtain the guide price of the path. The path with the lowest guide price is selected as the optimal path for the next hop, and the guide price is added to the total path overhead as the overhead of the path.

Each link state path has a path capacity limit, and if the capacity of the path exceeds a certain value, the probability of affecting the quality of transport along the path will be high. Thus, the guide price calculation method should be designed according to the path capacity in the LSBP routing mechanism. The guide price of the link state path contains two parts: the quote of the carrier and the excess capacity charge. The excess capacity charge is defined as follows: when the number of PI containers is less than the path capacity, carriers are not required to pay extra fees for excess capacity.

However, when the number of PI containers is greater than the path capacity, carriers must pay extra capacity fees for the excess capacity. The LSBP routing mechanism is adapted to convert the price gradient for cases with individual and multiple carriers, which is convenient for the calculation of the guide price. When a carrier operates alone, the average burden degree is used to convert the ladder quotes of different carriers into the guide price of the path. However, if multiple carriers are used, the transportation alliance is considered one carrier, and the quotes of carriers in the alliance are consistent. Furthermore, excess capacity charges form part of the guide price.

The calculation of the excess capacity cost paid by carriers operating in individual mode is as follows:

$$f = v \cdot \alpha \cdot \beta \cdot L \cdot \bar{F} \quad (3)$$

$$v = \frac{v_d - \bar{v}_r}{v_{rm} - \bar{v}_r} \quad (4)$$

$$\bar{v}_r = \frac{v_{rm}}{n} \quad (5)$$

$$\alpha = \frac{1}{n} \quad (6)$$

$$\beta = \frac{1}{v_{cm} - \bar{v}_r} \quad (7)$$

The calculation of the excess capacity cost paid by carriers operating in an alliance mode is:

$$f = v' \cdot \beta' \cdot L \cdot \bar{F} \quad (8)$$

$$v' = \frac{v'_d - \bar{v}_r}{v'_{rm} - \bar{v}_r} \quad (9)$$

$$\beta' = \frac{1}{v'_{cm} - \bar{v}_r} \quad (10)$$

For carriers operating individually, the excess capacity ratio describes the correlation between the capacity provided by the carrier and the carrying capacity of the path. The average burden degree is closely related to the number of carriers using the path. When the number of carriers tends to infinity, the excess capacity cost tends to zero. The market as a whole tends to be an optimally competitive market, and in this case, a path can be considered a distance vector link. The excess capacity coefficient reflects the correlation between the capacity provided by the carrier and the excess capacity cost. With other parameters unchanged, if the maximum capacity that the carrier can provide is large, the excess capacity cost that the carrier must pay is small. The carrier transports most

of the available PI containers and may be the leading enterprise in the region. These carriers are likely to be supported by various subsidies and business scaling effects. As a result, the extra cost associated with the excess capacity is reduced, consistent with real logistical situations. While carriers are operating in an alliance mode, the maximum capacity of the link state path is equal to the sum of the maximum capacity that all carriers can provide. Since there is only one carrier, the calculation of the average burden degree is omitted. This study employs the average freight rates of the carriers corresponding to the price gradient multiplied by various parameters. The resulting excess capacity costs reflect the average level of link capacity costs.

Algorithm 1: Uniform guide price gradient calculation

INPUT: *node_list, path_list, carrier_list, carrier_quotation, carrier_capacity, Cap_interval*

OUTPUT: *UG*

For a node in *node_list*:{

 set $cost_i \leftarrow \text{Calcost}(\text{path_list}, \text{carrier_quotation}, f_i, \text{carrier_list})$

 set $cap_i \leftarrow \text{Calcap}(\text{path_list}, \text{carrier_capacity}, \text{carrier_list})$

 set $acc_cost_i \leftarrow \text{Acc}(cost_i)$

 set $acc_cap_i \leftarrow \text{Acc}(cap_i)$

 set $fit_i \leftarrow \text{Fit}(acc_cap_i, acc_cost_i)$

 set $UG_i \leftarrow \text{Calgra}(fit_i, \text{Cap_interval})$

 set $cost_u \leftarrow \text{Calcost}(\text{path_list}, \text{carrier_quotation}, f_u)$

 set $cap_u \leftarrow \text{Calcap}(\text{path_list}, \text{carrier_capacity})$

 set $acc_cost_u \leftarrow \text{Acc}(cost_u)$

 set $acc_cap_u \leftarrow \text{Acc}(cap_u)$

 set $fit_u \leftarrow \text{Fit}(acc_cap_u, acc_cost_u)$

 set $UG_u \leftarrow \text{Calgra}(fit_u, \text{gradient})$

if $UG_i < UG_u$ **then** $UG = UG_i$

else $UG = UG_u$ }

Return *UG*

6. CASE STUDY

6.1. Greater Bay Area CPI Medical Logistics Network with Link Layer Fusion

Various basic data regarding medical logistics were obtained from the official website and database of the medical providers through a data source service. Each dataset includes information such as the medical institution name, latitude and longitude data, and CPI customer type. The data structure is shown in Table 3.

Table 3 Data structure table

Name	Field Name	TYPE
Area ID	ID	Int
Medical institution	Name	String
Longitude	Lng	Float
Latitude	Lat	Float
Infrastructure number	Infra_num	Int
Customer number	Customer-num	Int
Rank	Rank	String

Excluding the ID and Name data dimensions, there are five data dimensions: Longitude, Latitude, Infrastructure Number, Customer Number and Rank. Infrastructure Number is used as the dependent variable, and the remaining data dimensions are used as independent variables for sensitivity analysis. The results shown in Table 4-6 indicate that a change in Customer Number causes a large change in Infrastructure Number, followed by Rank (hospital rank in the simulation), and that changes in Latitude and Longitude have little effect on Infrastructure Number.

Table 4 Total Sensitivity and Confidence

	ST	ST_confidence(95%)
lng	0.003125	0.000245
lat	0.000630	0.000060
customer_num	0.866178	0.051622
rank	0.132102	0.011361

Table 5 First-order Sensitivity and Confidence

	S1	S1_confidence(95%)
lng	0.002915	0.004984
lat	0.000554	0.002417
customer_num	0.865089	0.064248
rank	0.129958	0.033722

Table 6 Second-order Sensitivity and Confidence

	S2	S2 confidence(95%)
(lng, lat)	0.000165	0.006220
(lng, customer_num)	0.000385	0.009538
(lng, rank)	0.000192	0.006712
(lat, customer_num)	0.000203	0.004655
(lat, rank)	-0.000062	0.003495
(customer_num, rank)	0.001743	0.096916

Based on the information obtained from the data source service, we obtained the raw data from the official website of the Guangdong Health Care Commission and finished the data preprocessing steps, as shown in Algorithm 2. A heatmap (Fig. 5) is adopted to show the correlation between each attribute. Latitude and longitude data are used as the main indicators for experimental calculations, and the orders are used as weighted indicators for CPI regional link centers. We used the K-medoids clustering method for autonomous area division and three evaluation methods, including the elbow method, silhouette coefficient and CH index, for measuring clustering performance. We input the data into a Python program combining these three methods to obtain the appropriate K value for the number of autonomous areas, and the output of the program is shown in Fig. 6.

Algorithm 2: Data preprocessing

Input: raw_data, Lng_border_up, Lng_border_down, Lat_border_up, Lat_border_down
Output: new_data, Infra_null, customer_null, rank_null, corr
 Infra_null = 0, customer_null = 0, rank_null = 0
 Lng, Lat \leftarrow Crawling(raw_data)
 Median_Infra \leftarrow median(raw_data[i][3])
 Median_customer \leftarrow median(raw_data[i][4])
 For i in range (len (raw_data)) {
 raw_data[i][1] \leftarrow Lng
 raw_data[i][2] \leftarrow Lat
 If (raw_data[i][1] < Lng_border_up) and (raw_data[i][1] > Lng_border_down) **then**
 If (raw_data[i][2] < Lat_border_up) and (raw_data[i][2] > Lat_border_down) **then**
 new_data[i] \leftarrow raw_data[i]
 End if
 End if
 If (new_data[i][3] == null) **then**
 Infra_null \leftarrow Infra_null + 1
 new_data[i][3] \leftarrow Median_Infra
 End if
 If (new_data[i][4] == null) **then**
 customer_null \leftarrow customer_null + 1
 new_data[i][4] \leftarrow Median_customer
 End if
 If (new_data[i][5] == null) **then**
 rank_null \leftarrow rank_null + 1
 new_data[i][5] \leftarrow Ffill(new_data)
 End if
 new_data[i][5] \leftarrow Score(new_data[i][5]) }
 corr \leftarrow Correlation(new_data)
Return new_data, Infra_null, customer_null, rank_null, corr

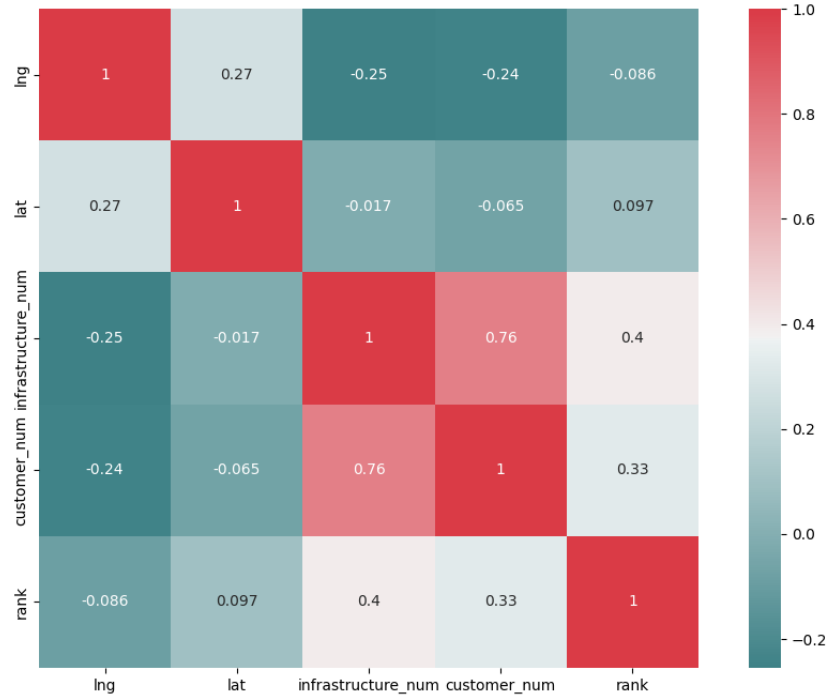


Fig. 5 Data attribute correlation heatmap

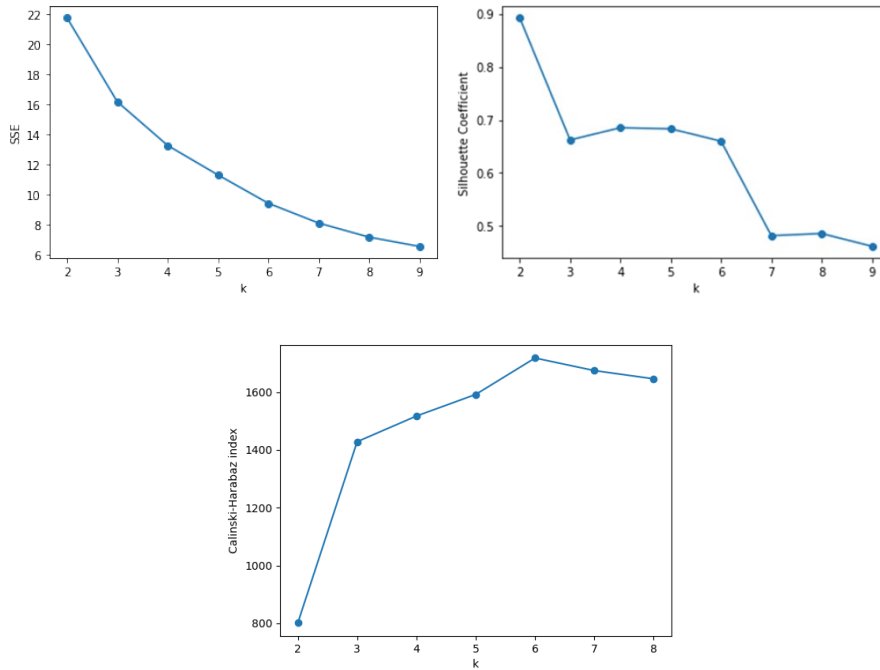


Fig. 6 The results of the elbow method, silhouette coefficient and CH index

The elbow method describes the closeness of the internal structure of clusters based on the sum of square error (SSE). The silhouette coefficient method reflects the rationality of the number of clusters determined based on the silhouette coefficient. The CH Index considers both inter-cluster

dispersion and intra-cluster dispersion. The output of the elbow method reveals that the SSE tends to decrease with an increasing value of K . However, the plot does not display an obvious turning point. The output of the silhouette coefficient method shows that the silhouette coefficient reaches a maximum when the value of K is 2, whereas the SSE also reaches a maximum and the CH index reaches a minimum when K is equal to 2, such a scheme with $K=2$ is discarded. When the value of K is in the interval of $[3, 6]$, the silhouette coefficients are similar. If the value of K is increased further, the silhouette coefficient decreases substantially, and the CH index starts to decrease. Based on the output results of the three indicators, when $K=6$, the SSE is small, and the silhouette coefficient and the CH index are large, indicating a tight structure inside autonomous areas and a reasonable number of autonomous area divisions. Accordingly, $K=6$ is selected as the number of autonomous area divisions.

After the autonomous area division, the core autonomous area is determined through the election. Subsequently, the weighted distance function updated in 4.2 is adopted to generate CPI regional link centers. Fig. 7 presents the CPI-based GBA medical logistics network with link layer fusion.

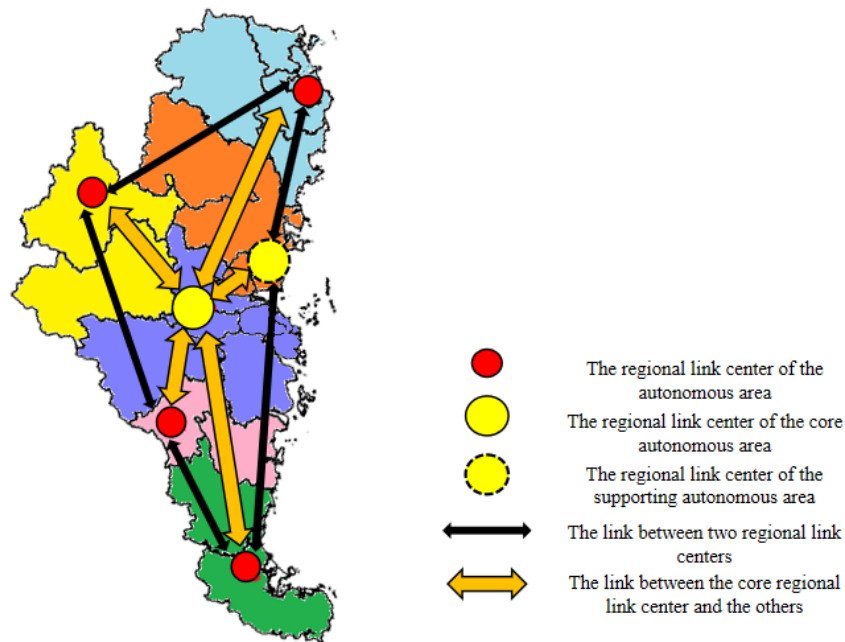


Fig. 7 Schematic diagram of the division of autonomous areas

The different autonomous areas are distinguished by different colors, and CPI regional link centers are marked in Fig. 7. The purple and orange areas represent the core autonomous area and the supporting autonomous area in the CPI link layer, respectively. The regional link center of the core autonomous area, represented by a yellow solid circle in the figure, connects to the other regional link center and assists the core autonomous area in maintaining connectivity with other autonomous areas. The regional link center of the core autonomous areas has better performance and better quality of communication links with other regional link centers. If the core autonomous area fails, the supporting autonomous area will replace the core autonomous area, and the regional link centers will be transferred. To illustrate the relationship between these two areas, we use a yellow dashed circle to represent the regional link center of the supporting autonomous area.

6.2. Greater Bay Area CPI Medical Logistics Network with Transport Layer Fusion

There are two roles of PI logistics nodes and PI demand nodes in the CPI medical logistics network, and only PI demand nodes have transportation demand. All the capacity in the network is provided by PI logistics nodes, and a respective PI logistics node has at most 20 PI transporters that can carry 50 capacity units. The maximum capacity along a link state path is 50 capacity units, and the maximum capacity of a distance vector path is infinite. PI transporters will return to PI logistics nodes after completing their tasks. When the capacity along a path exceeds half of the maximum carrying capacity, the excess capacity fee is added, and the network is updated once a day.

To analyze the operation of the routing mechanism for different network sizes, a portion of the data obtained from the data source service is randomly selected. Three groups of logistics scenarios of different scales are designed: one with 2 PI logistics nodes and 50 PI demand nodes, one with 2 PI logistics nodes and 100 PI demand nodes, and one with 4 PI logistics nodes and 100 PI demand nodes. The respective groups serve as control groups for each other. Ten randomized experiments were conducted for each scenario, and the average of the final statistical results data was used as the experimental result.

The ladder quotes of different carriers for link state paths and unit distance quotes for distance vector paths are simulated. The demands of different PI demand nodes are randomly set. A Python program is constructed to unify the ladder quotes with different gradients for carriers to obtain a consistent price gradient. Then, the excess capacity cost is calculated from the maximum capacities in the path. The final guide price is based on the consistent price gradient and the excess capacity cost. Moreover, to make the simulation network more realistic, the states of paths in the experiment are randomly assigned to healthy or unhealthy by a random function. A healthy path is in a stable state and can be accessed normally, while an unhealthy path is required to pay a larger price for its access or even cannot be accessed. In this experiment, distance vector paths between PI demand nodes and PI logistics nodes are set as healthy paths to ensure that at least one path accesses a PI demand node so that the demand of all PI demand nodes can be satisfied.

For three different scale logistics scenarios, two algorithms are used to solve the routing problem. Experiments 1-3 adopted a genetic algorithm (GA) to obtain a routing solution. With the objective of satisfying the demand of all PI demand nodes and the total path overhead as fitness, 100 iterations are performed to obtain the final logistics solution. GA assumes that no matter how far apart the nodes are, there is always a direct path between them. However, direct paths do not exist between all nodes in real networks. Therefore, a sufficiently large number is set as the path overhead for unhealthy paths in this experiment, meaning that a large fee needs to be paid when the logistics solution generated by the GA contains unhealthy paths. Experiments 4-6 adopted the CPI routing mechanism, which is based on combining DVMP and LSBP routing mechanisms. Each PI logistics node randomly selects a nearby PI demand node to serve. If a carrier has remaining capacity after satisfying the demand of a PI demand node, it will select another PI demand node from the routing table that matches the capacity balance and has a healthy path to be accessed as the next hop and will not return to the original PI logistics node until the capacity balance is insufficient. If no other PI demand node is available, the carrier will return to the original PI logistics node in advance. This process is repeated until all PI demand nodes have been satisfied or all PI logistics nodes have run out of capacity. Two

indicators (the logistics network capacity utilization rate and the load balance degree) are employed to evaluate the experimental results. The logistics network capacity utilization rate expresses the ratio of the total network capacity to the maximum network capacity. The larger the ratio is, the higher the utilization rate. The load balance degree expresses the degree of deviation of the path load from the average load. The smaller the value is, the closer the path load is to the average load. The calculation methods for the logistics network capacity utilization rate and load balance degree are illustrated in Equation (11) and Equations (12)-(13).

The logistics network capacity utilization rate is calculated as follows:

$$CU = \frac{\sum q_i \cdot d_i}{\sum C_i \cdot d_i} \quad (11)$$

The load balance degree is expressed as follows:

$$\bar{c} = \frac{\sum c_i}{\sum i} \quad (12)$$

$$LB = \frac{\sqrt{\frac{1}{\sum i} (c_i - \bar{c})^2}}{\bar{c}} \quad (13)$$

The results of the experiments are shown in Table 7 and Fig. 8:

Table 7 The results of the experiments

ID	Overall Distance	Total Cost	Number of Paths	Capacity Utilization	Load Balance Degree	Experimental Description
1	107.05	1598984.43	31	0.722	3.193	2 PI Logistics Nodes, 50 PI Demand Nodes
2	76.57	221255.78	31	0.727	3.240	4 PI Logistics Nodes, 50 PI Demand Nodes
3	143.78	1575243.42	46	0.838	3.501	4 PI Logistics Nodes, 100 PI Demand Nodes
4	145.33	28328.43	38	0.682	0.334	2 PI Logistics Nodes, 50 PI Demand Nodes
5	85.57	21223.72	33	0.687	0.269	4 PI Logistics Nodes, 50 PI Demand Nodes
6	183.74	35768.42	60	0.818	0.371	4 PI Logistics Nodes, 100 PI Demand Nodes

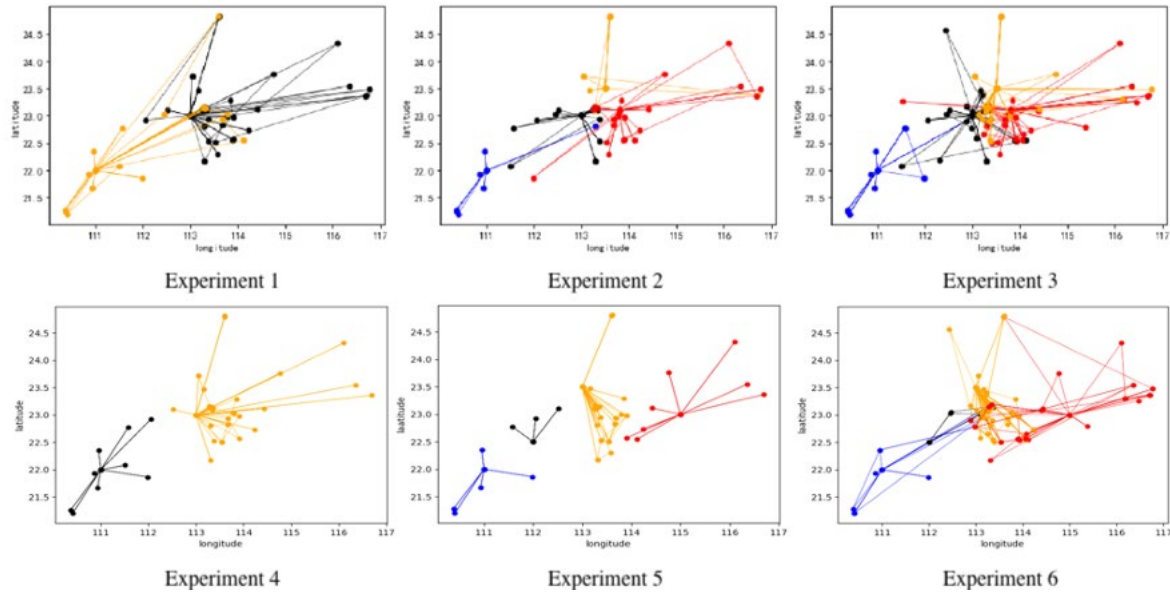


Fig. 8 The results of the experiments

In this study, three sets of comparative experiments are designed for each of the two routing algorithms. Ten random repetitions are conducted for each set of experiments, and the average of the output data is used as the result of the experiment to explore the performance of the two routing algorithms under network capacity increasing, logistics demand increasing and both capacity and demand increasing. The results of the experiments can be analyzed in two ways:

A comprehensive view of the results indicates that at the same logistics scale, the total cost, capacity utilization and load balance degree increase gradually as the overall distance decreases. Because both routing algorithms tend to select shorter paths, the network capacity is thus more concentrated. When the number of paths does not change too much, the path overhead gradually increases with a corresponding increase in capacity utilization and load balancing, which corresponds to the reality of logistics. Moreover, the results show that the number of PI logistics nodes has a greater impact on the overall distance and total cost of the network. With the unchanged number of PI demand nodes, the increase in PI logistics nodes has little impact on the number of network paths, capacity utilization and load balancing degree, but the overall distance and total cost of the network have a relatively large decrease. This is because the total demand of the logistics network fluctuates less, and the required capacity does not change much, so the changes in the number of network paths,

capacity utilization and load balancing degree are also small. However, since the increasing PI logistics nodes give PI demand nodes more options, PI demand nodes will preferentially accept services from the nearest PI logistics nodes. As a result of distant logistics routes in the original network being reduced, the overall distance and total cost of the network decrease. Meanwhile, it is common sense that an increasing number of PI demand nodes would lead to varying degrees of increase in the various outcomes. In addition, the comparison between the results of Experiment 1 and Experiment 3 shows that a multiplying increase in PI demand nodes and PI logistics nodes does not lead to an exponential increase in overall distance and total cost. Because the routing algorithms of the network allow the capacity of the paths to be fully utilized and the overall distance and total cost to be minimized, the network maintains a high utilization rate and low operating costs. This also illustrates the cost advantages of the *CPI-LI²* framework. The *CPI-LI²* framework is effectively applied to the pharmaceutical logistics network in the GBA, as experiments have proven that the *CPI* pharmaceutical logistics network has high capacity utilization and relatively stable load balancing.

From the comparison of the experimental results of both the GA and *CPI* routing mechanisms, the following conclusions can be drawn. First, for the same logistics scale scenarios (comparing Experiment 1 and Experiment 4, Experiment 2 and Experiment 5, and Experiment 3 and Experiment 6), the total cost for the *CPI* routing approach was significantly lower. This is because the logistics scenarios using GA add some unhealthy paths to the routing solution by crossover and mutation operations during the iterative process, affecting the total cost significantly. The *CPI* routing mechanism only finds healthy paths, so the total path overhead is lower. Second, the *CPI* routing mechanism always achieves a higher overall distance than the GA in the same logistics scenario. The main reason is that GA goes through 100 iterations, and the overall distance obtained is closer to the optimal solution, while the *CPI* routing mechanism only performs once without an iterative process; thus, the result obtained is most likely a local optimal solution. Therefore, the *CPI* routing mechanism performs weaker than the GA in terms of overall distance. Finally, since unhealthy paths can be adopted in the GA, the number of paths can be reduced by paying a high fee to make some

unreachable nodes accessible. This situation will be mitigated to a great extent by an increase in PI logistics nodes (comparing Experiment 1 and Experiment 2) since the number of paths increases as the PI logistics nodes increase and the likelihood of selecting an unhealthy path decreases. The total cost obtained by the CPI routing mechanism fluctuated less (comparing Experiments 1, 2, and 3 and Experiments 4, 5, and 6) because if the capacity balance is sufficient but the paths between nodes are all unhealthy, CPI carriers will give up selecting the unhealthy path and return to the original logistics node, and the demand of other PI demand nodes will be satisfied by new paths. Then, the number of paths under the CPI routing mechanism will be larger, which is more obvious with the increase in logistics nodes and demand nodes, such as Experiment 3 and Experiment 6. Another reason for this phenomenon is that the number of nodes increases and the demand increases exponentially. In the case that the proportion of unhealthy paths in the total paths remains unchanged, the CPI routing mechanism tends to increase the number of paths to cope with the surging logistics demand within the capacity range since it does not have an iterative process.

7. CONCLUSION

CPI is a novel paradigm for building global interconnected logistics networks in the future, with the goal of not only increasing the efficiency of logistics network operations by reconfiguring traditional networks but also supporting new forms of logistics management by reconfiguring the logistics network. CPI builds a cyber layer on top of PI, allowing PI containers to be configured and scheduled in cyberspace, thus enabling seamless flows of PI containers in the logistics network similar to packets that are transmitted in the computer network. The fusion of the logistics network is key to realizing this concept. The integration of logistics infrastructures and the operation of integrated infrastructures are essential to improve the intelligent service level of CPI integrated logistics networks. The CPI integrated logistics network reflects changes in data in real time, enabling real-time decisions based on real-time data and dynamically adjusting the operation mode of the network.

As one of the world's economic centers, the GBA has complete logistics infrastructures, yet the capacity of some logistics infrastructures has not been effectively developed and utilized. Logistics infrastructure integration is the key to promoting the GBA to achieve a coordinated and efficient freight integrative network. In this study, a CPI-LI² framework with a link layer and transport layer is proposed to address the problem of CPI logistics infrastructure integration in the GBA. A LEAF method is adopted in the link layer to establish connections between logistics infrastructure and form autonomous areas and CPI regional link centers, improving the interoperability of logistics infrastructures and achieving the integration of logistics infrastructures. The DVMP routing mechanism and LSBP routing mechanism are proposed based on the distance vector characteristics of the path and the path state characteristics, respectively. Combining these two routing mechanisms, a CPI operating mechanism of the logistics infrastructure integrated network is formed based on the guide price. At the transport layer, the corresponding operation mode according to the characteristics of logistics networks at different levels is designed, and the evaluation indicators for path overhead are formed. By calculating the path overhead of the logistics network and planning an effective routing solution, PI containers can be transferred more conveniently in the CPI logistics network, balancing the operational efficiency and cost advantages of the logistics network.

Through the simulation experiment with the background of the GBA medical logistics network, this study can conclude that the CPI-LI² framework effectively improves the operational efficiency and cost advantages of logistics networks. This framework improves infrastructure interoperability through logistics infrastructure integration and enables real-time monitoring, analysis and decision-making in CPI logistics networks through flexible operational mechanisms. The CPI-LI² framework has potential and can promote CPI development, which is shown as follows.

(1) This paper proposed the LEAF method to integrate logistics infrastructures. This method introduces the concept of computer networks and uses clustering methods to divide the logistics network into different network layers, similar to the autonomous areas in computer networks. In this case, failures in the local network do not have a huge impact on the whole network, and the division

of areas reduces the size of the database and lowers the requirements for infrastructure capacity. Moreover, the LEAF method improves network response efficiency and disaster tolerance through a weighted distance function and core autonomous area election. The LEAF method is a combination of several general methods. It considers the influence of realistic factors on the division of physical network areas and adopts a suitable method to construct a reliable physical network link layer, which provides a solution to the problem of large-scale logistics infrastructure integration that is influenced by multiple factors.

(2) This paper developed a DVMP routing mechanism for distance vector paths and an LSBP routing mechanism for link state paths at the transport layer and adopted the guide price as the path overhead, forming a CPI routing mechanism that can flexibly select the optimal next hop. This mechanism ensures that the network has higher utilization and more stable load balancing, improving the quality of service and cost-effectiveness of the network. In addition, this mechanism has portability. Since the guide price is based on carriers' quotes, which widely exist in logistics networks for different types and different scales, the CPI routing mechanism can be applied in other transportation scenarios (e.g., air transportation and shipping) and even be further developed into a multimodal transport model.

Notably, some specific topics are worthy of further exploration. First, this study uses the Euclidean distance rather than the actual logistical distance as the distance between two locations. The main reason is that distance does not have a large impact on the results, and most studies used the Euclidean distance to design routing mechanisms. If the distance between two locations is given or can be calculated, the Euclidean distance can be replaced by the actual distance. For another reason, the reality is highly complicated, while the experimental approach in this paper is a simulation based on the principle of simplicity; thus, it is unnecessary to use a complex model consisting of actual paths for research. Next, the process of designing the CPI-LI² framework considers mostly physical factors and neglects human factors. When the regional link centers are elected based on a distance weighting function, the weights are roughly estimated based on experience. The high levels of

standardization and intelligence of CPI infrastructure significantly reduce the manpower required in the logistics process. In addition, labor costs are only a small part of the operating cost in the logistics process. Finally, there is no comparison between the CPI routing mechanism and traditional navigation. On the one hand, the CPI routing mechanism proposed in this paper is based on a highly shared intelligent logistics network; however, the current logistics infrastructure and logistics information from different logistics stakeholders are not shared at a high level, and there is still a gap compared to an ideal CPI logistics scenario, so this paper cannot force the level of infrastructure sharing. On the other hand, the limited scale of data collected in this paper is not sufficient to support the comparison of CPI routing mechanisms with traditional navigation, which is also one of the limitations of this paper. This part of the work will be supplemented if more comprehensive data can be obtained subsequently. Although the current research on CPI is at the stage of theoretical accumulation and validation, the experiments in this paper can reflect the advancement and feasibility of the CPI concept to some extent.

However, the $CPI-LI^2$ logistics infrastructure integration framework developed in this paper is based on the land transportation scenario in the GBA. The framework has great potential to be improved since the current routing mechanism at the transport layer cannot support multimodal transport scenarios. Future research will focus on exploring the CPI logistics network, considering the route transfer and transport mode conversion problems under multimodal transport scenarios, and improving the transport layer routing mechanism, enabling the $CPI-LI^2$ framework to be adapted to the multimodal transport scenario of the CPI logistics network, thus enabling the CPI logistics network to operate in integrated infrastructures under the multimodal transport scenario.

ACKNOWLEDGMENTS

This research is supported by the RGC Theme-based Research Scheme (No. T32-707/22-N) and the PolyU DGRF of ISE (No. P0045748).

REFERENCES

- [1] M. Fazili, U. Venkatadri, P. Cyrus, and M. Tajbakhsh, “Physical Internet, conventional and hybrid logistic systems: a routing optimisation-based comparison using the Eastern Canada road network case study,” *Int. J. Prod. Res.*, vol. 55, no. 9, pp. 2703–2730, May 2017, doi: 10.1080/00207543.2017.1285075.
- [2] B. Montreuil, R. D. Meller, and E. Ballot, “Physical Internet Foundations,” in *Service Orientation in Holonic and Multi Agent Manufacturing and Robotics*, T. Borangiu, A. Thomas, and D. Trentesaux, Eds., in *Studies in Computational Intelligence*. Berlin, Heidelberg: Springer, 2013, pp. 151–166. doi: 10.1007/978-3-642-35852-4_10.
- [3] M. Rajahonka, A. Bask, S. A. Yawar, and M. Tinnilä, “The physical internet as enabler of new business models enhancing greener transports and the circular economy,” *Sustain. Effic. Transp.*, pp. 70–94, Aug. 2019.
- [4] P. Pagano *et al.*, “RTPORT: the 5G-based Model-Driven real Time Module for General Cargo Management,” p. 17.
- [5] Y. Sallez, S. Pan, B. Montreuil, T. Berger, and E. Ballot, “On the activeness of intelligent Physical Internet containers,” *Comput. Ind.*, vol. 81, pp. 96–104, Sep. 2016, doi: 10.1016/j.compind.2015.12.006.
- [6] T. Ambra, A. Caris, and C. Macharis, “Towards freight transport system unification: reviewing and combining the advancements in the physical internet and synchromodal transport research,” *Int. J. Prod. Res.*, vol. 57, no. 6, pp. 1606–1623, Mar. 2019, doi: 10.1080/00207543.2018.1494392.
- [7] H. Tran-Dang and D.-S. Kim, “The Physical Internet in the Era of Digital Transformation: Perspectives and Open Issues,” *IEEE Access*, vol. 9, pp. 164613–164631, 2021, doi: 10.1109/ACCESS.2021.3131562.

- [8] X. T. R. Kong, K. Kang, R. Y. Zhong, H. Luo, and S. X. Xu, "Cyber physical system-enabled on-demand logistics trading," *Int. J. Prod. Econ.*, vol. 233, p. 108005, Mar. 2021, doi: 10.1016/j.ijpe.2020.108005.
- [9] X. T. R. Kong, X. Yang, K. L. Peng, and C. Z. Li, "Cyber physical system-enabled synchronization mechanism for pick-and-sort ecommerce order fulfilment," *Comput. Ind.*, vol. 118, p. 103220, Jun. 2020, doi: 10.1016/j.compind.2020.103220.
- [10] Z. Zhao, R. Y. Zhong, Y.-H. Kuo, Y. Fu, and G. Q. Huang, "Cyber-physical spatial temporal analytics for digital twin-enabled smart contact tracing," *Ind. Manag. Data Syst.*, vol. 121, no. 5, pp. 1082–1106, Jan. 2021, doi: 10.1108/IMDS-11-2020-0634.
- [11] Y. Jiang, X. Liu, K. Kang, Z. Wang, R. Y. Zhong, and G. Q. Huang, "Blockchain-enabled cyber-physical smart modular integrated construction," *Comput. Ind.*, vol. 133, p. 103553, Dec. 2021, doi: 10.1016/j.compind.2021.103553.
- [12] E. Ballot, B. Montreuil, and Z. G. Zacharia, "Physical Internet: First results and next challenges," *J. Bus. Logist.*, vol. 42, no. 1, pp. 101–107, 2021, doi: 10.1111/jbl.12268.
- [13] E. Ballot, B. Montreuil, and R. D. Meller, *The Physical Internet*. La Documentation Française, 2014. Accessed: Aug. 06, 2022. [Online]. Available: <https://hal-mines-paristech.archives-ouvertes.fr/hal-01113648>
- [14] R. Y. Zhong, C. Xu, C. Chen, and G. Q. Huang, "Big Data Analytics for Physical Internet-based intelligent manufacturing shop floors," *Int. J. Prod. Res.*, vol. 55, no. 9, pp. 2610–2621, May 2017, doi: 10.1080/00207543.2015.1086037.
- [15] E. K. H. Leung, C. K. H. Lee, and Z. Ouyang, "From traditional warehouses to Physical Internet hubs: A digital twin-based inbound synchronization framework for PI-order management," *Int. J. Prod. Econ.*, vol. 244, p. 108353, Feb. 2022, doi: 10.1016/j.ijpe.2021.108353.
- [16] W. Jo, S. Kim, H. Kim, Y. Shin, and T. Shon, "Automatic whitelist generation system for ethernet based in-vehicle network," *Comput. Ind.*, vol. 142, p. 103735, Nov. 2022, doi: 10.1016/j.compind.2022.103735.

- [17] J. Lee, B. Bagheri, and H.-A. Kao, “A Cyber-Physical Systems architecture for Industry 4.0-based manufacturing systems,” *Manuf. Lett.*, vol. 3, pp. 18–23, Jan. 2015, doi: 10.1016/j.mfglet.2014.12.001.
- [18] P. B. M. Fahim, R. An, J. Rezaei, Y. Pang, B. Montreuil, and L. Tavasszy, “An information architecture to enable track-and-trace capability in Physical Internet ports,” *Comput. Ind.*, vol. 129, p. 103443, Aug. 2021, doi: 10.1016/j.compind.2021.103443.
- [19] A. Pal and K. Kant, “F2 π : A Physical Internet Architecture for Fresh Food Distribution Networks,” p. 16.
- [20] Thibaut Cerabona, Frederick Benaben, Jean-Philippe Gitto, Matthieu Luras, and Benoit Montreuil, “Physical Internet inspired Atomic Modeling for Supply Chain Risk Management,” in *IPIC 2021 - 8th International Physical Internet Conference*, Online, Greece, 2021.
- [21] M. Mededjel, G. Belalem, and A. Neki, “A cloud-fog architecture for physical-internet-enabled supply chain,” *Supply Chain Forum Int. J.*, vol. 0, no. 0, pp. 1–16, Nov. 2021, doi: 10.1080/16258312.2021.1996861.
- [22] C. Pach, Y. Sallez, T. Berger, T. Bonte, D. Trentesaux, and B. Montreuil, “Routing Management in Physical Internet Crossdocking Hubs: Study of Grouping Strategies for Truck Loading,” in *Advances in Production Management Systems. Innovative and Knowledge-Based Production Management in a Global-Local World*, B. Grabot, B. Vallespir, S. Gomes, A. Bouras, and D. Kiritsis, Eds., in IFIP Advances in Information and Communication Technology. Berlin, Heidelberg: Springer, 2014, pp. 483–490. doi: 10.1007/978-3-662-44739-0_59.
- [23] S. Gontara, A. Boufaied, and O. Korbaa, “Routing the Pi-Containers in the Physical Internet using the PI-BGP Protocol,” in *2018 IEEE/ACS 15th International Conference on Computer Systems and Applications (AICCSA)*, 2018, pp. 1–8. doi: 10.1109/AICCSA.2018.8612885.
- [24] S. Pan, D. Trentesaux, E. Ballot, and G. Q. Huang, “Horizontal collaborative transport: survey of solutions and practical implementation issues,” *Int. J. Prod. Res.*, vol. 57, no. 15–16, pp. 5340–5361, Aug. 2019, doi: 10.1080/00207543.2019.1574040.

- [25] G. Yoon, J. Y. J. Chow, A. Dmitriyeva, and D. Fay, “Effect of Routing Constraints on Learning Efficiency of Destination Recommender Systems in Mobility-on-Demand Services,” *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 5, pp. 4021–4036, 2022, doi: 10.1109/TITS.2020.3038675.
- [26] A. Kantasa-ard, T. Chargui, A. Bekrar, A. A. El Cadi, and Y. Sallez, “Dynamic Multiple Depots Vehicle Routing in the Physical Internet context,” *IFAC-Pap.*, vol. 54, no. 1, pp. 92–97, Jan. 2021, doi: 10.1016/j.ifacol.2021.08.011.
- [27] P. Mátray, P. Hága, S. Laki, G. Vattay, and I. Csabai, “On the spatial properties of internet routes,” *Comput. Netw.*, vol. 56, no. 9, pp. 2237–2248, Jun. 2012, doi: 10.1016/j.comnet.2012.03.005.
- [28] R. Sarraj, E. Ballot, S. Pan, D. Hakimi, and B. Montreuil, “Interconnected logistic networks and protocols: simulation-based efficiency assessment,” *Int. J. Prod. Res.*, vol. 52, no. 11, pp. 3185–3208, Jun. 2014, doi: 10.1080/00207543.2013.865853.
- [29] S. Kaup, A. Ludwig, and B. Franczyk, “Framework Artifact for the Road-Based Physical Internet based on Internet Protocols.” arXiv, Jun. 15, 2021. doi: 10.48550/arXiv.2106.08286.
- [30] H.-S. Park and C.-H. Jun, “A simple and fast algorithm for K-medoids clustering,” *Expert Syst. Appl.*, vol. 36, no. 2, Part 2, pp. 3336–3341, Mar. 2009, doi: 10.1016/j.eswa.2008.01.039.
- [31] W. Sheng and X. Liu, “A hybrid algorithm for k-medoid clustering of large data sets,” in *Proceedings of the 2004 Congress on Evolutionary Computation (IEEE Cat. No.04TH8753)*, Jun. 2004, pp. 77-82 Vol.1. doi: 10.1109/CEC.2004.1330840.
- [32] F. Liu and Y. Deng, “Determine the Number of Unknown Targets in Open World Based on Elbow Method,” *IEEE Trans. Fuzzy Syst.*, vol. 29, no. 5, pp. 986–995, 2021, doi: 10.1109/TFUZZ.2020.2966182.
- [33] S. Aranganayagi and K. Thangavel, “Clustering Categorical Data Using Silhouette Coefficient as a Relocating Measure,” in *International Conference on Computational Intelligence and Multimedia Applications (ICCIMA 2007)*, 2007, pp. 13–17. doi: 10.1109/ICCIMA.2007.328.
- [34] C. T. Harabasz and M. Karoński, “A dendrite method for cluster analysis,” in *Communications in Statistics*, 1974, pp. 1–27.