


RESEARCH

Open Access

Survivability-aware routing restoration mechanism for smart grid communication network in large-scale failures



Bao Ju Liu^{1,2} , Peng Yu^{1*}, Qiu Xue-song¹ and Lei Shi³

Abstract

Natural disasters such as earthquakes have consecutive impacts on the smart grid because of aftershock activities. To guarantee service requirements and smart grid stable operations, it is a challenge to design a fast and survivable rerouting mechanism. There are few studies that consider concurrent rerouting aiming at multiple services in smart grid communication network, however. Firstly, we formulate the node survivability, link survivability, and path survivability model in terms of the distance from the epicenter to the node and the link of the network. Meanwhile, we introduce the indicator of site difference level which is unique in the smart grid to further restrict the service path. Secondly, to improve the algorithm efficiency and reduce rerouting time, the deep first search algorithm is utilized to obtain the available rerouting set, and then the I-DQN based on the framework of reinforcement learning is proposed to achieve concurrent rerouting for multiple services. The experimental results show that our approach has a better convergence performance and higher survivability as well as the approximate latency in comparison with other approaches.

Keywords: Smart grid communication, Large-scale failures, Rerouting, Survivability, Deep Q network

1 Introduction

Smart grid (SG) is a new bulk power system-integrated automation technology and information technology (e.g., the advanced remote sensing technology, communication technology, information technology, and control technology) with the physical network. It offers consumers a reliable, economical, clean, and interactive power supply and various additional services by leveraging powerful demand side management and real-time pricing mechanism, which plays a significant role in the construction of smart cities [1]. Furthermore, SG is a typical cyber-physical system, which consists of two distinguishable and complex networks: the physical power system and the communication network. The former is in charge of the energy transmission, and the latter provides the necessary scheduling and control functions in terms of the carried services with the purpose of smooth operations in SG. Messages are obtained either by

sensing physical power system states or collecting from the intelligent terminal devices (e.g., PMU, IEDs, distributed generation resources, and sensors) and then transmitted and processed in the manner of wire or wireless to achieve multi-level and multi-dimensional situational awareness as well as the optimal control of the physical grid. Therefore, SG is considered one of the most significant applications of IoT [2].

With the accelerated interactions of source-grid-load and the employment of various power applications, services in smart grid communication network (SGCN) featured with low latency and high reliability are increasing in a geometric progression [3]. Taking point-to-point relay protection as an example, the unidirectional latency cannot exceed 10 ms, or else cascading failures are prone to happen. Furthermore, 5G-enabling technologies of huge capacity, low latency, ubiquity, less energy consumption, and faster deployment provide reliable transmission for services in SGCN. In addition, as the key technology of 5G, SDN (software-defined networking) is an emerging network paradigm. The separation of the data plane and the control plane, the global view of the network and the

* Correspondence: yupeng@bupt.edu.cn

¹State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, China
Full list of author information is available at the end of the article

programmability make it possible for real-time path recovery in SGCN [4]. Nowadays, the combination of 5G, SDN, and SG to achieve faster service deployment has become a serious concern for SGCN reliability [5–7].

Different from the general communication network, SGCN is a private network characterized by high-speed, integration, two-way communication. Moreover, the optical fibers in SGCN are always laid along the powerlines. Moreover, nodes referred to the communication devices deployed in power plants and substations (e.g., routers, switches) and links (e.g., ADSS, OPGW) are mostly exposed to a harsh environment, and thus they have a high probability to be damaged resulting from human and natural disaster factors, such as intentional attacks, destruction, hurricane, frost, earthquake, and tsunami. As SG heavily relies on the communication system supporting, the disruptive delivery easily leads to cascading failures in the physical power system to such as power flow transfer, load loss, and blackouts.

It is noted that natural disasters are always evitable, regional, and unpredictable, and they always cause destructive damage to people's lives and infrastructures. We think that the research about large-scale failures in SGCN with low frequent occurrence is equally or even more important than the minor failures. Taking an earthquake as an example, its occurrence is accompanied with the release of a large amount of destructive energy, and thus many electricity devices or secondary devices in terms of the distance from the disaster center, namely the epicenter, are damaged in different level. More severely, it possibly causes the whole network to paralyze. For instance, the Sichuan Province Changning county of China earthquake in June 2019 resulted in the shutdown of four 35-kV substations, four 35-kV powerlines, and seven 10-kV lines. About 38,000 consumers suffered from power outages [8]. Furthermore, the earthquake has a durative impact from a few seconds or minutes to several decades due to aftershocks [9].

Generally, the nodes along with links in the communication networks which are closer to the epicenter have a higher possibility to be damaged again in the subsequent aftershock activity, and vice versa [10]. In order to minimize the impact of natural disasters on the performance of communication networks and to ensure consecutive and real-time transmission for various types of services in the smart grid, it is necessary to provide effective rerouting solutions to enhance network resilience, especially in large-scale failure scenarios. To guarantee the smooth operation of SG in the event of aftershocks, we focus on establishing a faster and reliable rerouting mechanism for multiple interrupted services simultaneously in this research.

However, the existing network management system based on devices in SG confronts the issues of decentralized

control, human intervention, poor interoperability, lack of flexibility and scalability. Hence, it is inefficient in the implementation of intelligent operations and service diversification as well as the requirement of discrepancies QoS [11]. In multiple protocol label switching (MPLS), changing labels can be employed to control data flow direction and improve the controllability of MPLS. Nevertheless, routers need to be reconfigured for per new application, which increases the restoration time for services and restricts its application in SGCN. The IP-based rerouting strategy has a slow convergence speed in the event of large-scale failures. It is difficult to satisfy the rigorous latency requirement for services in SGCN. Due to the benefits of decoupling of software and hardware, flexible and efficient network management and control, SDN paradigms provide a new solution to improve the existing network architecture [12, 13].

Nevertheless, the traditional rerouting schemes usually formulate the requirements as a single objective optimization with respect to specific services, such as end-to-end latency, cost, vulnerability, network risk balance, or the weighting values for the mentioned metrics to restore services one by one [14–16]. There are few studies for simultaneous rerouting for multiple services. Besides, the conventional routing approaches may not adapt to the network dynamic changes; therefore, they could hardly provide a desirable performance in the multi-service rerouting scheme for large-scale failures.

To obtain the appropriate solution, it is of vital importance to formulate a solvable mathematical model in accord with practical requirements. However, it is hard to formulate an accurate routing model especially for the occurrence of natural disasters scenario, which refers to many interrupted links (nodes), services, and differentiated QoS demands and specific constraints of smart grid communication network. Moreover, it is difficult to obtain the optimal solution because of the complex system model or too many constraints during routing computation. As a consequence, it leads to the inconsistency between the optimal solution and the real solution if we ignore some constraints or make some ideal assumptions [17, 18].

As a model-free solution in reinforcement learning (RL), the DQN-based algorithms overcome the shortcomings of the traditional model and have been widely used in traffic control, resource management, and routing recovery [19–21]. Therefore, it motivates us to explore novel solutions under the framework of RL to solve the simultaneous rerouting for the entire services in large-scale failures.

In order to avoid service disrupted again due to aftershocks in SGCN, establishing a flexible and high survivability rerouting mechanism in the framework of deep reinforcement learning (DRL) for interrupted services is our main objective. To the best of our knowledge, this is the first research that solves the rerouting mechanism

under the framework of DRL. The key contributions of our work can be summarized as follows:

1. We first calculate the service recovery level in terms of service requirements of the end-to-end latency, importance, and the voltage level. After that, we build models of node survivability, link survivability, and path survivability according to the distance between nodes (links) and the epicenter. Besides, the end-to-end latency and site difference levels are also considered to be in accordance with the operation of SGCN.
2. We propose a recovery scheme for the entire services in the SGCN under the framework of deep reinforcement learning. To this end, we calculate the routing path set for disrupted services and then formulate the corresponding state, action, and the reward function of DRL.
3. We provide an improved deep Q network (I-DQN) based on priority resampling for faster convergence. The experiment results show that it has better convergence and better survivability compared with other reinforcement learning and other routing strategies.

The remainder of this paper is organized as follows. Section 2 investigates various fault recovery approaches with different network technology. Section 3 describes the concurrent rerouting mechanism for multiple services in SGCN. Section 4 analyzes the factors that affected path reliability in case of aftershocks occurrence. Section 5 provides a brief introduction of RL and the details of I-DQN. Section 6 presents the simulation results and Section 7 concludes the paper.

2 Related work

2.1 Single-failure restoration

Preventive protection and dynamic restoration are two kinds of restoration mechanisms in communication networks. In the former, the interrupted service is rerouted to a disjoint alternate path that is established in advance when the working path fails. Here, the working path and the alternate path are the optimal and sub-optimal paths in the network computed by some routing algorithms in terms of the network resource utilization and network topology. Similar protection methods are adopted in different networking technologies, for instance, automatic protection switching (APS) and self-healing rings (SHR) in SDH, intra-domain protection, inter-domain protection in ASON, linear protection, ring protection, and subnet protection in PTN and ASON protection in OTN. Various protection technologies can guarantee fast restoration. However, it is incapable of coping with double failures or multi-failure scenarios. Particularly, in

large-scale failures, both the working path and the alternate path are likely to be destroyed simultaneously, and thus the protection scheme is basically invalid.

In the optical network, various approaches based on P-cycle are proposed to achieve joint protection regarding on-cycle links and straddling links. In case of a single link failure, an automatic switching between the working path and the alternate path on the p-cycle can effectively guarantee consecutive delivery of service requests. Furthermore, such protection enables fast recovery especially in a ring-like network and in the scenario of multiple failures. Zhang et al. [22] proposed the PSPP algorithm through disjoint p-cycle configuration for each link in the working path to enhance the network survivability.

2.2 Multiple-failure restoration

However, SGCN is a complex time-variant system, and the protection scheme-reserved network resources for all possible failure situations results in inefficient resource utilization and limited recovery ability.

The restoration method refers to dynamically establishing rerouting for services according to real-time network topology and network resources. There are roughly two restoration approaches aiming at different networking technologies: the dynamic rerouting and pre-establishing rerouting. Different rerouting algorithms based on node-based rerouting, link-based rerouting, and end-to-end rerouting are widely studied. Compared to the protection method, the restoration scheme is more applicable to the large-scale failure scenario due to its flexibility and high resource utilization.

Note that it takes milliseconds or tens of seconds to converge again in the IP network in case of failures, and it worsens network performance because of packet loss or discarding during this period. Therefore, IETF developed an IP-based rerouting framework (IPFRR). This greatly attracts many researchers to launch the research in regard to the fast reroute. Elhourani et al. [23] proposed an IP fast reroute method by means of constructing a rooted arc-disjoint spanning tree in k -edge-connected network in case of $k-1$ link failures. Similarly, in the case of multi-link failures [24], a disjoint spanning tree based on edge cut was employed to reduce packet loss ratio, balance network load, and lower network recover delay. In order to improve the algorithm efficiency and performance as well as implement complete protection against link failures in the network [25], an approach of TOD (tunneling on demand) was addressed. An interface-specific routing model (ISR) is capable of handling link failures in most situations, and a tunneling mechanism will be activated if it fails to be backhauled by ISRs. At the same time, TOD is proved to enable protection against single-link failure and double-link failures

effectively and achieve complete link protection with minimal tunnel overhead when multiple link failures occur.

Ji et al. [14] analyzed the influence of link failures in SGCN combing communication service attributes with graph theory. To achieve the minimal end-to-end latency and uniform distribution of communication services, a rerouting algorithm combined the k -shortest path algorithm with genetic algorithm which was designed to reduce network operation risk. On this basis [15], the interdependence relation between the cyber network and physical network in CPPS (Cyber-physical power system) was explored and the impact of link failures on the physical network and communication services was described. According to this, the cross-space risk mapping from the cyber network to the physical network was established, and then, the genetic algorithm aiming to make sure networking risk balancing was adopted.

2.3 Large-scale-failure restoration

In large-scale network failure scenarios, both the affected scope of disasters and the number of damaged equipment have a direct effect on network performance. The reserved network resources are projected at a fivefold growth in case of a triple-failure scenario or more in comparison with a single failure. Furthermore, this demand grows rapidly with the increase of the number of failures in the network. Thus, the pre-planned routing method is hard to satisfy specific communication requirements of network operations in case of large-scale failures [26]. To improve network-restoring ability, the shared physical links in the network are described as Shared Risk Link Groups (SRLG) to form resource bundles. Kiese et al. [27] proposed a multi-invulnerability protection scheme aiming to minimize the number of shared SRLGs between the working path and the alternate path.

To verify the recovery feasibility in large-scale failure scenarios [28], an epicenter as the center point and the impact range as the radius of a circle were modeled and then changes in terms of network capacity requirements and fault notification time were analyzed. Authors presented that only a small amount of spare capacity (bandwidth resources) is needed in the mesh network and obtains a higher service recovery ratio. Neumayer et al. [29] performed a network vulnerability assessment under the influence of natural disasters. Authors modeled the disaster impact as a line segment or a circular cut based on bipartite graphs and then calculating a worst-case line segment or circular cuts.

In order to reduce disruption time [10], a preventive protection scheme in case of large-scale failures was proposed. The authors established a dynamic, deterministic, and probabilistic failure model for the nodes and links according to the characteristics of seismic wave

propagation and the regional failures evolved with time and then calculated each path failure probability based on k shortest path algorithm. Each path is examined whether it belongs to the safe zone or not in terms of the preset threshold and then the available rerouting set for services is acquired. The first route is selected as the working path and one of the routes in the available rerouting set can be regarded as the alternate path only if satisfying survivability requirements and available network resources.

However, it is inapplicable to SGCN due to a lack of consideration for specific service requirements.

2.4 Resilience in SDN

To make a trade-off between the recovery time and forwarding rule occupation resulting from frequent interactions between SDN controller with switches in case of the single-link failure [30], a flow aggregation strategy with the objective of minimizing the number of reconfiguration rules was proposed and was formulated as an integer linear programming problem. In the SDN-based mesh network [17], a fault-tolerant routing scheme was presented and converted into two sub-problems: the routing tree construction and the SDN controller deployment. The authors leveraged a pruning method to improve the survivability of the routing tree. This method is superior to the traditional methods with respect to the number of protected nodes in the network as well as reducing network vulnerability.

2.5 Machine learning application

Apart from various heuristic algorithms and intelligent algorithms, various machine learning algorithms have been applied to solve routing problems as well. Deep learning (DL), reinforcement learning (RL), and deep reinforcement learning (DRL) as the important branches of ML are gaining considerable attention from academia and industry [31]. RL is an AI (artificial intelligence)-based algorithm by trial and error to obtain rewards from the environment and learns the optimal strategy according to the maximal cumulative expected rewards. As a classical algorithm of RL, Q-learning enables to handle the problem of discrete state and action space. However, when the state-action space is enormous, it may not be feasible. DRL can be considered an improved version of RL that exploits the deep neural network (DDN) to approximate the value function or the policy.

In order to reduce flooding routing overhead and the impact of channel dynamic availability [32], a clustering routing algorithm based on spectrum sensing for cognitive radio networks was developed and solved in the framework of RL. Du et al. [21] proposed a joint routing and spectrum allocation mechanism for cognitive radio multi-hop networks. To guarantee power efficiency, it

aims to minimize transmission delay and exploits the improved DQN to solve it. The results prove that it has some advantages over packet loss ratio, throughput, etc. Ding et al. [33] presented an adaptive routing strategy based on the quality of service with the goal of minimizing the transmission delay and then leveraged the Q-learning algorithm to acquire better convergence.

However, there are few studies combining machine learning with routing survivability in large-scale failures of SGCN. Inspired by [10], we propose a concurrent rerouting approach for multiple services of SGCN in case of large-scale failures, which is solved under the framework of DRL.

3 Rerouting mechanism in large-scale failures

3.1 Communication architecture

SDN paradigm is characterized by centralized control and management, programmability, independent protocol, granularity, etc. Since the smart grid heavily relies on communication networking assistance, SDN-based communication architecture can be used to monitor and manage communication entities and improve network

efficiency and resiliency potentially. SDN-based SGCN has been applied to load balancing, dynamic service routing adjusting, fast fault detection, and so on [17]. In this part, we still adopt the proposed architecture for SGCN which has been presented in our previous research shown in Fig. 1 [13].

A heterogeneous/hybrid communication network for SG is built integrated with various demand and supply at the data plane, and fine-grained forwarding rules are installed in the switch flow table through the standard interface to perform packet deliver. At control plane, SDN controller is in charge of monitoring and managing communication networks of SG as well as computing paths for services and managing the defined forwarding rules. For instance, it determines the paths that services take through the network. At application plane, various typical service systems are involved, including wide-area measurement system (WAMS), distribution management system (DMS), meter data management system (MDMS), etc. Notifications about control services, protections services, and various measurements from backbone SDN switches are by intelligent terminals.

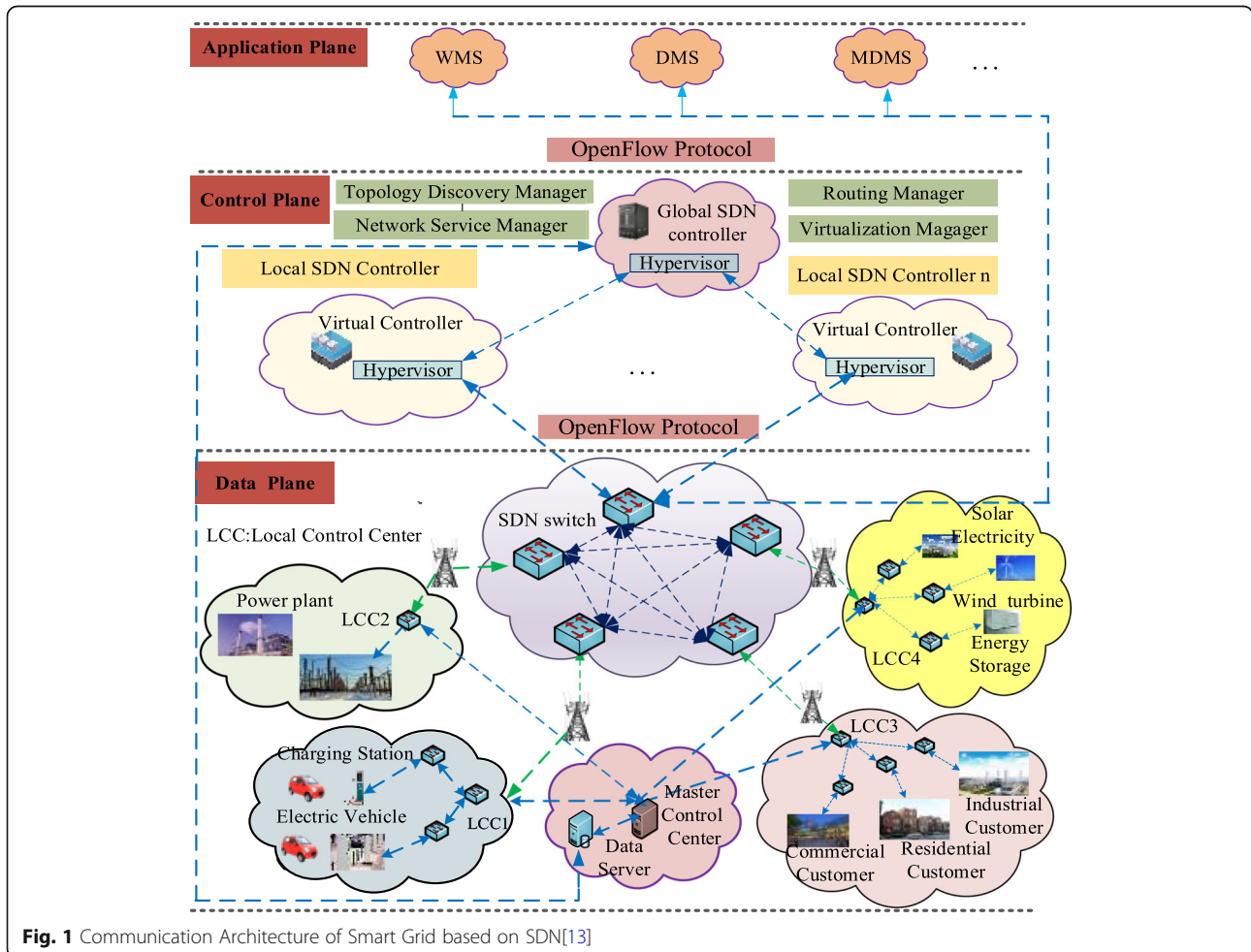


Fig. 1 Communication Architecture of Smart Grid based on SDN[13]

SDN-based SGCN provides a centralized view of the complicate network to applications. An integration and interaction of three planes guarantee SDN controller with the global view of the network and make it possible to make wiser and flexible routing decisions in a distributed network. It can better meet the needs of rigorous recovery latency and high reliability of services in SGCN in case of large-scale failures.

3.2 Rerouting framework for SGCN

We design the recovery mechanism for SGCN based on the following assumptions: (1) Node failure results in all connected link failures. (2) If the source or destination for any service is damaged, this service is unrecoverable. Given the $N-1$ principle for the smart grid, there are two pre-planned routes for critical services, one is the primary route and the other is the alternate route. Services can be rerouted to the alternate route in case of primary route failure. However, it is possible that both the primary route and the alternate route are damaged simultaneously, and thus a reliable rerouting strategy to restore services is necessary for such situation. In general, the recovery process can be divided into two phases: candidate routing set calculation and routing decision for service. In phase one, recovery levels for disrupted services are determined, and then we obtain the available paths for each service. In phase two, the approach of DQN-based is exploited to make routing decisions for services. The rerouting process in large-scale failure scenarios can be shown in Fig. 2.

3.3 Calculation of service recovery level

Besides the traditional voice, data, and video services, there are some other important services such as relay protection and stability control which are crucial to stable operations of the SG. However, the occurrences of disasters often cause many services to be interrupted or even worse paralyze the whole system. Upon detecting the interruptive service notification, the SDN controller needs to adopt a fast rerouting scheme and carry out the optimal path calculation to restore the interrupted services. To guarantee service performance, it is of vital importance to determine service superiority in the case of many interrupted services.

According to the State Grid Corporation's 13th Five-Year of Communication Network Plan [34], communication services are classified into the following five groups in terms of service importance: (1) relay protection of 500 kV/220 kV; (2) stability control; (3) wide-area phasor measurement, dispatching automation, telephone dispatching, and electric energy metering; (4) substation video monitoring, television consultation, and protection information management; and (5) office automation (OA), administrative telephone, and cloud terminal

application. Generally, protection and control services are more crucial to SG stable operation in comparison with management information service and thus have a higher recovery level. Moreover, different service categories have obvious QoS discrepancies. For example, the end-to-end latency for dispatching automation services that are responsible for scheduling, monitoring, and analysis and calculation for the smart grid is assumed to be no more than 100 ms, while the latency requirement for administrative telephone service is less than 250 ms. Hence, QoS standards of services have an effect on the service recovery level. In addition, as to the point-to-point relay protection, relay protection of substations with higher voltage due to a vaster control area has a greater effect on SG than those with lower voltage.

Hence, we calculate the service recovery level from the following three dimensions: end-to-end latency τ_k' , service importance α_k' , and voltage grade β_k' . Considering the typical characteristics of SG communication services [35], the five categories of services are distinguished by an integer in the range from 1 to 5. Similarly, the voltage requirement is represented as an integer from 1 to 3. That is, non-relay protection service is 1, and relay protection with 220 kV and relay protection with 500 kV are denoted as 2 and 3, respectively. Therefore, service recovery level for any b_k is computed below:

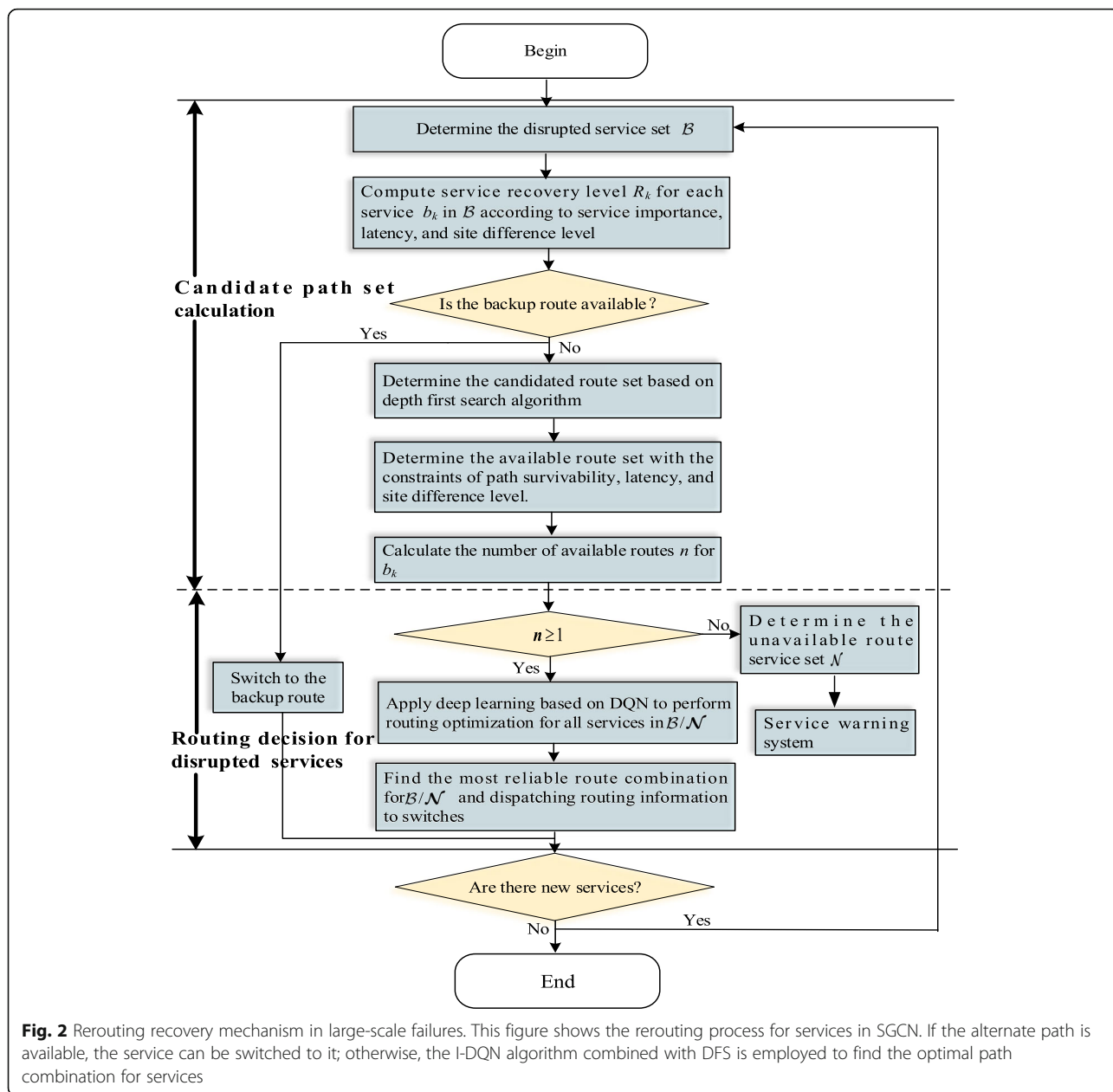
$$R_k = \vartheta \tau_k'' + \psi \alpha_k'' + \nu \beta_k'' = \vartheta \left(\frac{\tau_k' - \tau_{\min}'}{\tau_{\max}' - \tau_{\min}'} \right) + \psi \left(\frac{1}{\alpha_k'} \right) + \nu \left(\frac{\beta_k' - 1}{\beta_{\max}'} \right). \quad (1)$$

where R_k represents service recovery level for b_k , and here, ϑ , ψ , and ν are coefficients. They could be assigned a certain value in terms of specific requirements, and $\vartheta + \psi + \nu = 1$. τ_k'' , α_k'' , and β_k'' are normalizations of the above three indicators.

4 Methodology

We firstly model SGCN as an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{L}, \mathcal{B})$. Here, the switch deployed in the substation or the control center is abstracted into node $v_i, v_i \in \mathcal{V}$, and \mathcal{V} is the node set. The communication link from v_i to v_j is denoted as an edge $e_{ij}, e_{ij} \in \mathcal{L}$, and \mathcal{L} is the edge set. $\mathcal{B} = \{b_1, b_2, \dots, b_k, \dots\}$ is the affected service set, where b_k is the k th service in \mathcal{B} . A service requirement is represented as a quintuple $(v_f(k), v_t(k), T_k, B_k, R_k)$, where $v_f(k), v_t(k) \in \mathcal{V}$ are the source and the destination of b_k . T_k, B_k , and R_k indicate service requirements of latency, bandwidth, and service recovery level, respectively. p_k is the route for b_k , $p_k \in \mathcal{Z}$, and \mathcal{Z} is the service rerouting set.

To avoid the sequential effects of aftershocks, we make a comprehensive analysis with respect to survivability, end-to-end latency, and site difference level, which is relevant to routing reliability in SGCN.



4.1 Survivability

It refers to an ability to provide continuous transmission even if in case of network failures. Distinct from other natural disasters, earthquakes often have persistence in the time scale. Thus, finding high survivability paths is more practical for communication services in SGCN, especially in the frequent-earthquake districts. In a large-scale failure scenario due to an earthquake, it is vital to find a high-survivable path as far as possible from the epicenter for all disruptive services to provide consecutive delivery and guarantee network performance. Since an end-to-end routing path consists of a series of ordered nodes and edges, node survivability, link

survivability, and path survivability are defined in the following section.

4.2 Node survivability

To acquire node survivability, failure probability needs to be computed. For simplicity, different seismic wave propagation patterns are negligible in this research. In occurrence of an earthquake, a large amount of destructive energy can be released and spreads in a circle at a constant rate from the source. The epicenter is the location of the earthquake source on the ground. Due to the medium impact, the destructive effect is changed with the Euclidean distance between nodes or links and the

epicenter. Assume it conforms to the exponential distribution [10]. Let r be the impact radius. If the node is out of the impact radius, the failure probability is 0. λ is an attenuation factor, which describes the variation of earthquake attenuation. Other factors, e.g., network component aging, categories, techniques, and man-made destruction, are omitted in this research. Hence, the node failure probability can be represented as follows.

$$\mathcal{P}_i^v = \begin{cases} e^{-\lambda l_i^v(C)} & l_i^v(C) \leq r \\ 0 & l_i^v(C) > r \end{cases}, \quad \forall v_i \in \mathcal{V}, \quad (2)$$

where $l_i^v(C)$ is the distance between v_i and the epicenter C . Accordingly, the survivability of v_i is represented as the following formula.

$$S_i^v = 1 - \mathcal{P}_i^v \quad (3)$$

Similarly, link survivability describes the uninterrupted probability of a link in the event of failures. The link failure probability depends on the distance of the epicenter to the nearest point of the link. The smaller the distance is, the closer to the epicenter is, and the higher the link failure probability has, and vice versa. Figure 3 shows a simple example on how to compute the Euclidean distance for links in different locations and the epicenter

with different impact radius. The following formula denotes the link failure probability:

$$\mathcal{P}_{ij}^e = \begin{cases} e^{-\lambda \min(l_{ij}^e(C))} & l_{ij}^e(C) \leq r \\ 0 & l_{ij}^e(C) > r \end{cases}, \quad \forall e_{ij} \in \mathcal{L}, \quad (4)$$

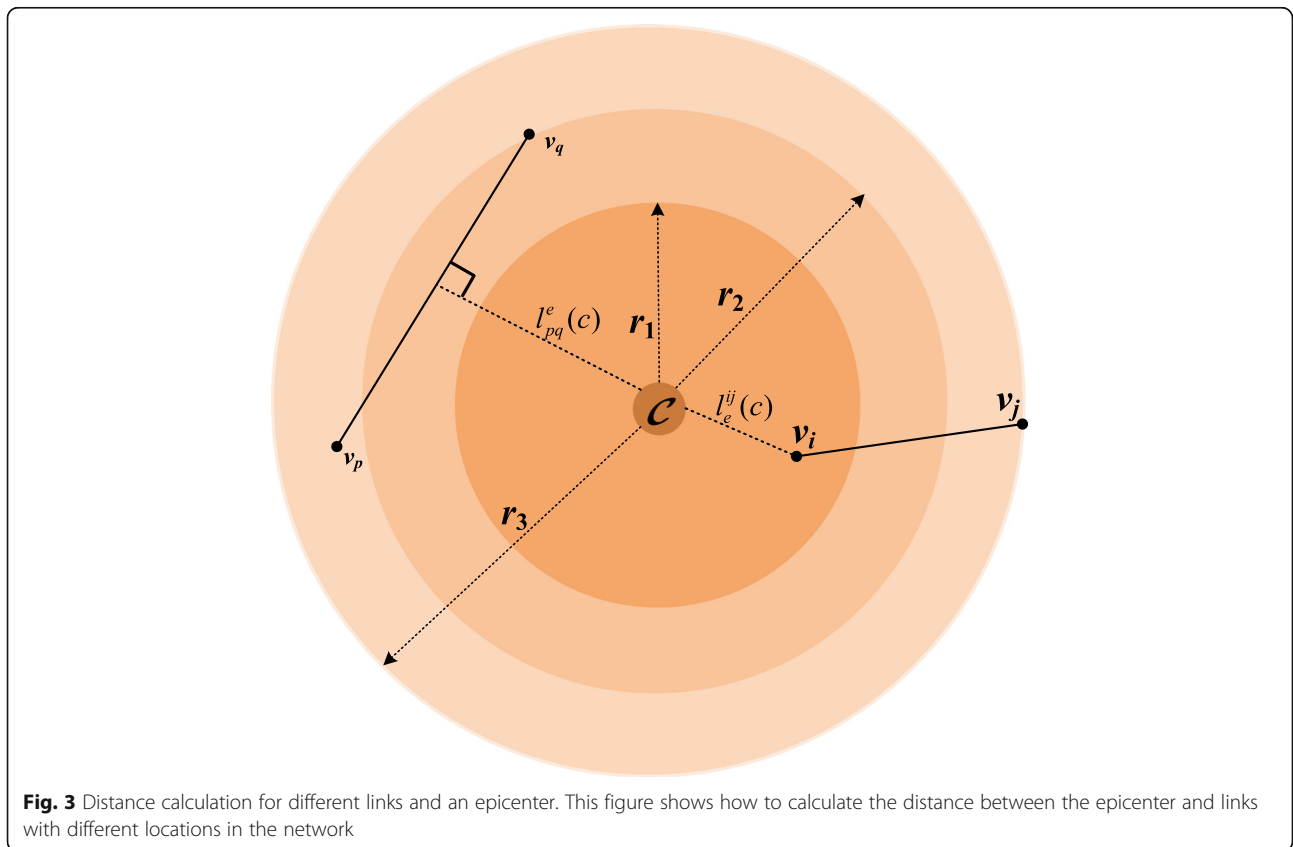
Accordingly, the link survivability is computed as follows:

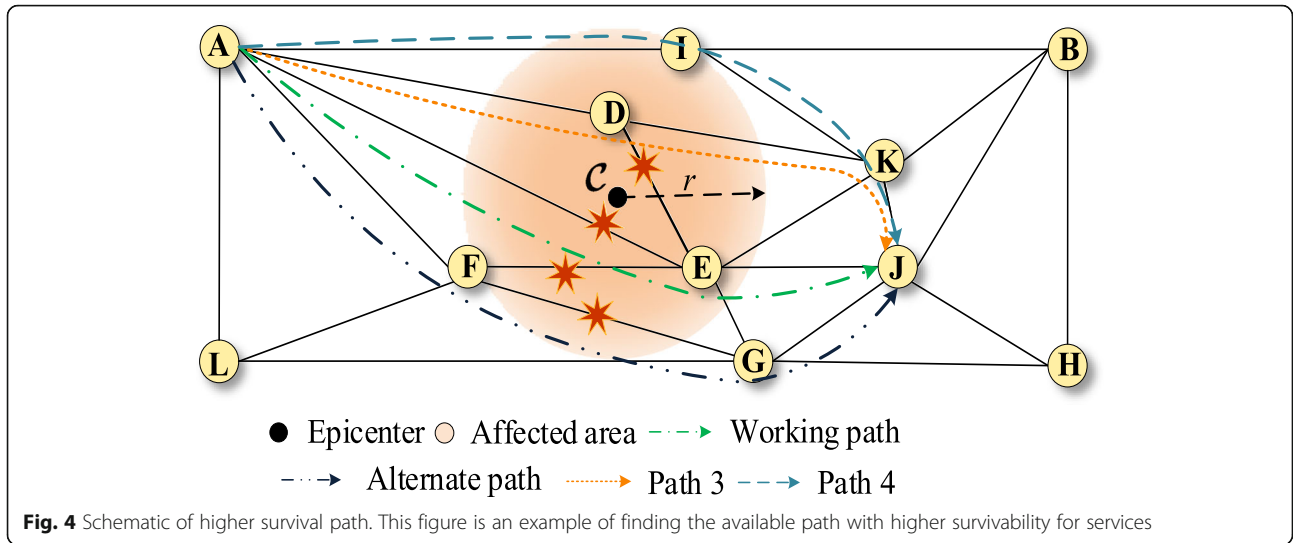
$$S_{ij}^e = 1 - \mathcal{P}_{ij}^e \quad (5)$$

We assume that the node failures and link failures are independent. Hence, path survivability is represented as the product of node survivability and link survivability.

$$S_{p_k} = \prod_{e_{ij} \in p_k} (1 - \mathcal{P}_{ij}^e) \prod_{v_i \in p_k} (1 - \mathcal{P}_i^v), \quad p_k \in \mathcal{Z}. \quad (6)$$

Figure 4 shows a simple example under path survivability decision-making procedure owing to large-scale failures. Take the service from A to J as an example, all available paths that meet service requirements in the current network status have been highlighted in different colors. It is assumed that the pre-planned path 1 and 2 are the working path and the alternate path, respectively. Path 1 and 2 are unavailable due to simultaneous failures of A-E, F-E, F-G. Since the node D and the link D-K in





path 3 are closer to the epicenter, and path 3 has higher failure probability in comparison with path 4. Hence, path 4 is more reliable for services to resist the influence of aftershocks.

4.3 Recovery latency

Comparatively, the end-to-end latency rather than the bandwidth factor should be considered for SG communication services. Taking the control service as an example, the total processing time for control measures is limited to no more than 300 ms in case of disturbances, and the end-to-end latency is less than 50 ms according to the international SG regulations [33]. As to the rerouting process for services in case of large-scale failures, the recovery latency refers to the total time from services disruption to being completely restored.

According to the SDN-based fault management mechanism, the recovery latency includes the following parts: the notification time between the adjacent node and SDN controller, the processing time of SDN controller, and transmission time. Hence, the recovery latency is represented as follows:

$$T_{p_k} = T_0(k) + T_{p_k} + T_{fc}, \quad (7)$$

Where $T_0(k)$ is a constant that denotes propagation latency for b_k before the occurrence of failures. T_{p_k} describes the propagation time delay for service b_k in p_k . T_{fc} represents the fault processing time for SDN controller, and it is formulated as the following equation:

$$T_{fc} = 2l_{cont}^{swit}/c + T_s, \quad (8)$$

Where l_{cont}^{swit} describes the link length between the switch and the nearest SDN controller. c is the light speed in fibers. The first item denotes notification

exchanged time between the SDN controller and switches, and the second item is the algorithm execution time.

Assume the SDN switch has sufficient processing ability; therefore, the queue time delay can be omitted. The propagation latency for b_k in route p_k is given in the following formula:

$$T_{p_k} = \sum_{e^{ij} \in \mathcal{L}} \frac{x_{ij}^k \times l_{ij}}{c} + \eta \times T_q + T_t, \quad p_k \in \mathcal{Z}, \quad (9)$$

$$\eta = \sum_{e^{ij} \in \mathcal{L}} x_{ij}^k. \quad (10)$$

where η is the number of forwarding devices of p_k , namely, the number of nodes of p_k . T_q describes the queue time delay in a switch. T_t depicts the message transmission time. l_{ij} represents link length of e_{ij} , and x_{ij}^k is a binary variable such that

$$x_{ij}^k = \begin{cases} 1, & \text{if the link } e_{ij} \text{ is in the route of } p_k \\ 0, & \text{otherwise} \end{cases} \quad (11)$$

4.4 Site difference level

SG is a typical cyber-physical system (CPS) composed of the communication network and the electric network. Both of them have similar structures. Different from the common communication devices, switches deployed in different substations correspond to different levels.

Generally speaking, substation with a larger voltage grade due to a vaster area has a higher level, and vice versa. For the convenience of an illustration, an integer which ranges from 1 to 4 describes different substation level. For instance, the substation level with 500 kV is 3, 220 kV is 2, and the like. Additionally, services are usually delivered among the switches in substations with the

same or similar voltage level according to specifications of information exchange and region division in SGCN [36], and there is no cross-level transmission. However, with the implementation of ultra-high voltage (UHV) projects in China, to reduce transmission cost and electric energy loss, the service from one UHV substation to the other always takes a detour with the aid of the lower-voltage substations in the vicinity. As a result, this possibly results in obvious substation level variations without substations level restriction and subsequently increases service transmission risk. Hence, we introduce the metric of site difference level to further limit nodes of the service path as follows:

$$\Delta_{ij} = |\kappa_i - \kappa_j| \leq \zeta, \forall i, j \in p_k, \kappa_i, \kappa_j = 1, 2, 3, 4, \quad j = i + 1, \quad (12)$$

$$\omega_{p_k} = \sum_{ij \in p_k} \Delta_{ij}. \quad (13)$$

where v_i and v_j are neighboring nodes in p_k , and κ_i and κ_j describe the substation level of v_i and v_j , respectively. Δ_{ij} represents the difference of κ_i and κ_j . ζ is a preset threshold and $\zeta = 1$ in this paper. ω_{p_k} is the sum of site difference level for all nodes in p_k .

5 Rerouting scheme based on DQN

To have a better understanding of the following sections, a brief introduction about reinforcement learning and deep Q network is illustrated in Section 5.1. However, it is far from a comprehensive survey, and its objective is just to provide the basic knowledge for readers.

5.1 Reinforcement learning and deep Q network

The process of RL is usually formulated as a Markov Decision Process (MDP), where time is divided into a series of time steps $t = 1, 2, \dots$. And there are four variables in RL procedure: state space S , action space A , probability transition matrix for states P , and rewards R . The objective of RL is to make the optimal action policy according to the immediate reward from the feedback of interaction with the environment through exploration and exploitation. The state is a description of the environment that RL agent perceives, and correspondingly, state space is the set of states. The action space denotes the possible action set that the agent can choose at each time step t .

The mapping from the state space to the action space is defined as a strategy π . Specifically, at any time step t , the agent observes the present state $s_t = s \in S$ and takes an action $a_t = a \in A$ in terms of the strategy π . Then, the agent gets feedback from the environment, that is, an immediate reward. At the same time, there is a transition from s_t to a new state $s_{t+1} = s' \in S$ in accordance with $P_{ss'}(a) \in P$. The agent repeats this process by trial and error and learns the optimal strategy with the objective of maximizing the

accumulative rewards in the end. It is worth mentioning that both state transition and feedback procedures from the environment are not controlled by the agent in the learning phase. The agent just has an effect on the environment by actions choice and perceives the environment further according to the gained reward.

Deep reinforcement learning (DRL) can be recognized as an enhanced version of RL, which has attracted considerable concerns from academia and industry due to AlphaGo of the DeepMind team [31]. DRL has obvious advantages in decision-making with respect to the issues of high-dimension state in comparison with the traditional RL, and it is applied to deal with complicated control issues. The action value function-based DRL and deep deterministic policy gradient (DPG) are two basic DRL methods [37]. The deep neural network (DNN) is utilized to approximate the action value function in the former. However, DNN in DPG is used to near the policy on basis of the policy gradient to learn the optimal strategy. The distinction is that the former is appropriate to solve discrete problems and the latter is widely used for continuous problems.

The objective of RL is to evaluate the action value function and greedily select the state-action pair with the maximal Q value to maintain Q table; TD-learning and Q-learning are classical algorithms based on state value function and action value function. DQN is the integration of Q-learning and deep neural network, and it is a value iteration algorithm-based on the Q network.

Nevertheless, there are so many state-action pairs in practice, the tabular algorithm (e.g., Q-learning) may encounter with dimension curse in accompany with more memory occupation and computation. Moreover, sparse samples may lead to instability, slow or even no convergence of the algorithm. Therefore, three countermeasures are taken to improve DQN performance. Firstly, to solve the problem of dimension explosion, the deep convolutional neural network (CNN) is used to estimate the action value-integrated DNN with Q-learning. Secondly, to break the temporal correlations among samples, the experience replay method is exploited in DQN. Specifically, at each time step t , the i th transition $e^{(i)} = (s^{(i)}, a^{(i)}, r^{(i)}, s'^{(i)})$ is stored into the experience memory $\mathcal{D} = \{e_1, \dots, e_M\}$. At the training phase, the statistic gradient descent algorithm is employed to update the neural network parameters on the basis of a minibatch of state transitions from \mathcal{D} . Finally, the target network is applied to predict the target Q value and improve training stability. The target network is considered an earlier snapshot of the Q network. And thus, it has the identical neural structure to the Q network.

In the training process, the optimal objective of the Q network $r + \gamma \max_{a'} Q(s', a'; \theta^-)$ derives from the target

network, where s' is the next state, a' is the possible action, and θ_i^- are target network parameters at iteration i . The parameters θ_i^- are updated with θ_i per \mathcal{H} iterations and θ_i^- remain unchanged in the rest procedures.

The Q network is updated with the goal of minimizing the loss function $L_i(\theta_i)$ at each iteration i [38]:

$$L_i(\theta_i) = E_{s,a \sim \rho(\cdot)} \left[\left(r + \gamma \max_{a'} Q(s', a'; \theta_i^-) - Q(s, a; \theta_i) \right)^2 \right], \quad (14)$$

Where $Q(s, a; \theta_i)$ is the value produced via Q network, and $\rho(s, a)$ is the probability distribution of state-action pair (s, a) .

After that, taking the derivative of the loss function with respect to θ_i , the corresponding gradient could be cast as

$$\nabla_{\theta_i} L_i(\theta_i) = E_{s,a \sim \rho(\cdot)} \left[\left(r + \gamma \max_{a'} Q(s', a'; \theta_i^-) - Q(s, a; \theta_i) \right) \nabla_{\theta_i} Q(s, a; \theta_i) \right], \quad (15)$$

$$Q^*(s, a) = E[r + \gamma \max_{a'} Q^*(s', a') | s, a], \quad (16)$$

The optimal strategy can be acquired in terms of the Bellman equation:

$$\pi^*(s) = \operatorname{argmax}_{a \in A} Q^*(s, a). \quad (17)$$

5.2 Problem formulation

In order to achieve fast rerouting for services in large-scale failures, the rerouting recovery problem is modeled as a model-free strategy learning process in this section. Figure 5 shows the schematic DQN execution process under SDN communication architecture. And here, the SDN controller can be regarded as the agent, the data plane consists of switches deployed in substations and the control center. The upper control plane is composed of TED (traffic engineering databases) and PCE (Path Computation Elements). The TED module is responsible for network topology and network connections. The PCE module supports the manner of centralized routing computation. Aiming to the multi-service rerouting problem in the large-scale failure scenario, it is extremely important to precisely identify the three metrics: the environment and state, the action, and the corresponding reward. We define them as follows.

5.2.1 Environment and state

The agent is responsible for making intelligent decisions and policy deployment in DRL. As can be seen from Fig. 5, in the SDN-enabled SGCN, the SDN controller with the global view of the network topology is able to accomplish the collection of networking parameters and service information, path computation and routing

deployment, traffic management, etc. And thus, it is regarded as the agent of deep reinforcement learning. The observed environment includes the current network topology and the interrupted services due to the damaged nodes or links.

In addition, in light of the requirement of low latency for services in SGCN, it is imperative that each path should satisfy the service latency requirement, or else the consideration about the path survivability and site difference level is meaningless for rerouting. Hence, to reduce computation complexity and guarantee service performance, our solution is to find all the possible paths for services based on depth first search algorithm (DFS) and search the available service rerouting set. After that, the DNN is introduced to learn the optimal path combination for all services under the framework of DRL. Therefore, the state is defined as follows:

$$s_t = \{[B_1(t), I_1(t)], [B_2(t), I_2(t)], \dots, [B_k(t), I_k(t)]\}. \quad (18)$$

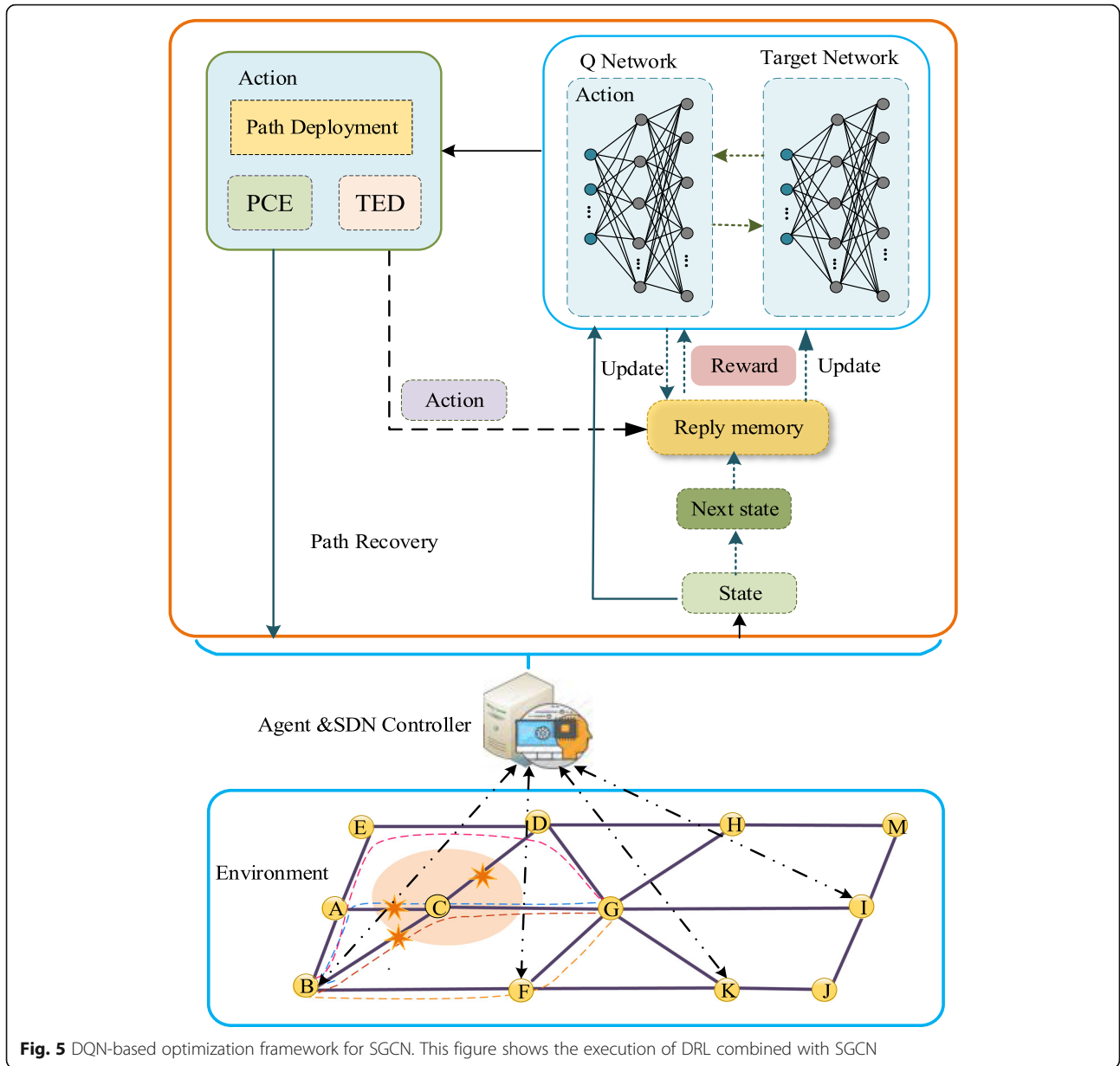
Where $B_k(t)$, $I_k(t)$ represent bandwidth requirement and the index of paths in available rerouting set, respectively. To improve the learning efficiency and get a better state representation, it is important to note that the metric of the bandwidth requirement should be normalized in advance in practice.

In fact, many nodes or links are damaged in large-scale failures due to earthquake occurrence, and many services are interrupted subsequently. Therefore, it leads to a larger state space in DRL. However, the appropriate rerouting set for services is enumerable, and the improved deep Q networks (I-DQN) algorithm for rerouting is designed based on action value function in this paper.

5.2.2 Action

As to the concurrent rerouting for all interrupted services, the action space is all the path combinations in terms of specific objectives; therefore, it is a discrete problem. If the action is designed to sequentially select paths from the available set, the action space is $A = \{u^k\}$, where u is the number of paths for b_k in the available rerouting set. Note that the action space size is varied with the number of services exponentially. Since there were a large number of services interrupted in large-scale failures, the action space is correspondingly very huge.

The path combinations are sorted in ascending order in terms of latency. We assume l to be the index of path combination randomly initially. Consequently, the action space is divided into two parts by l : the upper part is $l'_p = \{a_p | a_p \in A, 0 \leq a_p < l, \tau_a^p > \tau^*\}$ and the lower part is $l'_b = \{a_b | a_b \in A, |u^k| \geq a_b > l, \tau_a^b < \tau^*\}$, $A = l'_p \cup \{l\} \cup l'_b$, where



τ^* is the average end-to-end latency which can be obtained from the historical data. τ_a^p and τ_a^b are the average latency of the path a_p and a_b , respectively. Particularly, if $l = 0$, then $a_p = l$, and if $l = \max(|u^k|)$, then $a_b = l$. Such design is based on the latency of current average path and the historical average path. If the two indicators are equal, the value remains unchanged. If the average communication latency of the present state is larger than τ^* , the paths with smaller end-to-end latency should be selected from the upper part l'_p . Otherwise, the path is selected from the lower part l'_b to avoid getting stuck on a local optimum. $a_t = \{a_p, a_b, l\} \in A$, A is the set of candidate actions at time

step t . Compared to the initial action space, the newly generated action space is reduced enormously.

5.2.3 Rewards

DQN is known to train the neural network under the guidance of the reward, and the agent obtains an immediate reward from the environment in case of choosing a_t in the state of s_t . As the goal of DQN is to maximize the accumulated rewards, our objective is to simultaneously maximize path survivability and minimize site difference levels in the rerouting mechanism of SGCN. Therefore, we define the reward function that the DRL agent obtained as follows:

$$r_t = \sum_{b_k \in \mathcal{B}} (\mathcal{S}_{p_k} - \ell \omega_{p_k}), \quad p_k \in \mathcal{Z}. \quad (19)$$

where \mathcal{S}_{p_k} , ω_{p_k} correspond to the path survivability and site difference level for p_k , respectively. \mathcal{S}_{p_k} ranges from 0 to 1 and the site difference level is a dimensionless non-negative integer. ℓ is the coefficient for balancing the two variables. Given the discrepancies of the above two indicators, to ensure learning efficiency, we normalize them to a united scope in advance.

5.3 Improved DQN based on prioritized resampling

The issues about which transitions to store and which experience to replay have become serious concerns in the approach of DQN and various improved versions [19, 39]. Our research focuses on the latter. Temporal-difference (TD) error is a basic conception in RL, which refers to the difference between the target value function and current value function. And here, the target value function is the sum of the immediate reward and the next state value function. In [40], Mnih proposed a natural-DQN method without considering the disparity of samplings, which randomly select samplings from the replay memory to update neural network parameters. In fact, samples with different magnitudes of TD error have disparate backpropagation impacts [41]. In other words, samples with higher absolute TD-errors correspond to more important backpropagation impacts due to the larger loss. Hence, they should be replayed more often than others, and vice versa. The SARSA and Q-learning algorithms have calculated the absolute values of TD-error $|\delta|$. Thus, samplings are restored with the probability, namely $|\delta|$, which is a normalized metric.

$$P(j) = \frac{p_j}{\sum_m p_m}. \quad (20)$$

where $p_j = |\delta_j|$. This prioritization resampling mechanism ensures the higher magnitude TD-error samples to be stored and the lower samples to be erased at the same time, which further prevents the model degradation. Nevertheless, it results in limited samples and insufficient training as well. To enhance utilization efficiency and improve sample diversity [39], designed a prioritized replay sampling mechanism. The sampling probability is proportional to the sample storage priority. The storage priority is determined by $|\delta|$ which is derived from the last trained sample and avoids the possible bias in the update process. However, it leads to extra time complexity owing to the storage structure based on the binary heap with priority.

To make sure DQN efficiency and higher superiority samples update in a higher probability during the training phase, the coefficients α and the bias β are introduced in the calculation of sampling probability. The sampling probability of the learning experience j is defined as follows:

$$P(j) = \alpha \times \frac{p_j^\sigma}{\max_m (p_m^\sigma)} + \beta. \quad (21)$$

To acquire p_j , the method which is proportional to the magnitude of TD-error is adopted, that is $p_j = |\delta_j| + \epsilon$, where ϵ is a very small positive constant. Its role is to overcome the sampling weight approximate to 0 once their error is 0. The exponent $\sigma \in [0, 1]$ is exploited to describe the utilization degree of superiority. Especially $\sigma = 0$ corresponds to uniform sampling. Compared to the previous random sampling, the resampling method enables sample diversity further.

The rerouting algorithm (I-DQN) for the entire services based on the framework of DQN can be divided into two phases: the available rerouting set and various path metrics for services such as path survivability and site difference level are firstly calculated in phase one; after that, DQN is adopted to achieve the optimal path combination for services in phase two. The details of the algorithm are provided as follows:

6 Results and discussion

6.1 Parameter settings and service deployment

In this section, we assess the performance and efficiency of concurrent rerouting scheme for multiple services based on DQN in case of large-scale failures in SGCN. The experiment was conducted in tensorflow1.13.1,

Phase one: The algorithm of path parameters calculation

- 01: Input: Networking topology G , service set \mathcal{B} , and the disaster center and impact radius r
 - 02: Output: Candidate rerouting set and end-to-end latency, survivability and site difference level for services
 - 03: for each b_k in \mathcal{B} do
 - 04: if the alternate path p_k of b_k is available then
 switch b_k to the p_k
 - 05: else
 - 06: calculate the candidate path for services according to BFS and saved in \mathcal{Z}
 - 07: endif
 - 08: end for
 - 09: for each b_k in \mathcal{B} do
 - 10: Calculate all the possible paths for b_k with the constraint of latency based on DFS
 - 11: Calculate end-to-end time delay, survivability, site difference level of each path for b_k
 - 12: end for
-

Phase two: I-DQN based service path combination algorithm

```

01: Input : memory size  $N$ ,  $K$ , steps  $T$ , minibatch  $m$  and
         $\alpha, \beta, \sigma, \gamma$ , update frequency  $C$  of  $\tilde{Q}$ 
02: Initialize replay memory  $\mathcal{D} = \phi$ 
03: Initialize parameters of the Q network and the target
        network  $\tilde{Q}; \theta$  and  $\theta^-$ 
04 : for episode = 1,  $K$  do
05:   Initialize state  $s_1$ 
06:   for  $t = 1, T$  do
07:     Select action  $a_t$  with probability  $\epsilon$  randomly
08:     Otherwise select  $a_t = \operatorname{argmax}_a Q(s_t, a; \theta)$ 
09:     Observe the reward  $r_t$  corresponding to  $a_t$  and  $s_{t+1}$ 
10:     Compute TD-error of the sampling transition
         $j: \delta_j = r_j + \gamma Q(s_j, \operatorname{argmax}_a \tilde{Q}(s'_j, a; \theta^-)) - Q(s_{j-1}, a_{j-1}; \theta)$ 
11:     Compute the sampling weight of transition  $j: p_j = |\delta_j| + \epsilon$ 
12:     Compute sampling probability of transition  $j$  according
        to (21)
13:     Store  $(s_j, a_j, r_j, s_{j+1})$  into  $\mathcal{D}$ 
14:     Sample a minibatch of transitions  $(s_i, a_i, r_i, s_{i+1})$  from  $\mathcal{D}$ 
        in terms of  $P(j)$ 
15:     Compute the present objectives
        
$$y_i = \begin{cases} r_i, & \text{if episode ends up with } i+1 \\ r_i + \gamma \max_a \tilde{Q}(s_{i+1}, a; \theta^-), & \text{Otherwise} \end{cases}$$

16:     Carry out gradient descent on  $(y_j - Q(s_j, a_j; \theta))^2$  with
        respect to  $\theta$ 
17:   Set  $\tilde{Q} = Q$  every  $C$  steps
18:   end for
19: end for

```

Python3.7. To verify the algorithm performance, the proposed method is compared with two baseline solutions: Q-Learning and Natural-DQN [40]. Meanwhile, to validate the effectiveness of different rerouting schemes, the proposed rerouting strategy is compared with the shortest path (SP) and risk balancing routing algorithm (RBRA) as well [13].

The network topology in [42] has been utilized to verify the effectiveness of the proposed approach as shown in Fig. 6. It consists of 29 nodes and 47 links. Here, 5, 20, and 29 are 500-kV substations and 1, 7, 12, and 17 are 110-kV substations. The rest are 220-kV stations. Fourteen is the control center. The average node degree is 3.2, and the numbers on the links indicate the link length. The maximum impact radius of an earthquake is 500 km [21], the processing time is 0.01 ms [43]. The light speed in fiber is 2×10^5 km/s. The learning speed is 0.01, and the greedy factor is 0.9. In contrast with the general communication network, services in SGCN are not randomly distributed, and they happen between the substations, master stations, and the central stations. According to [34], the initial distribution of services in

SGCN is deployed as follows: 80% of the total services, the 500-kV substation is considered the source are chosen as the destination (source). The two terminals for the remaining services are randomly selected from \mathcal{V} .

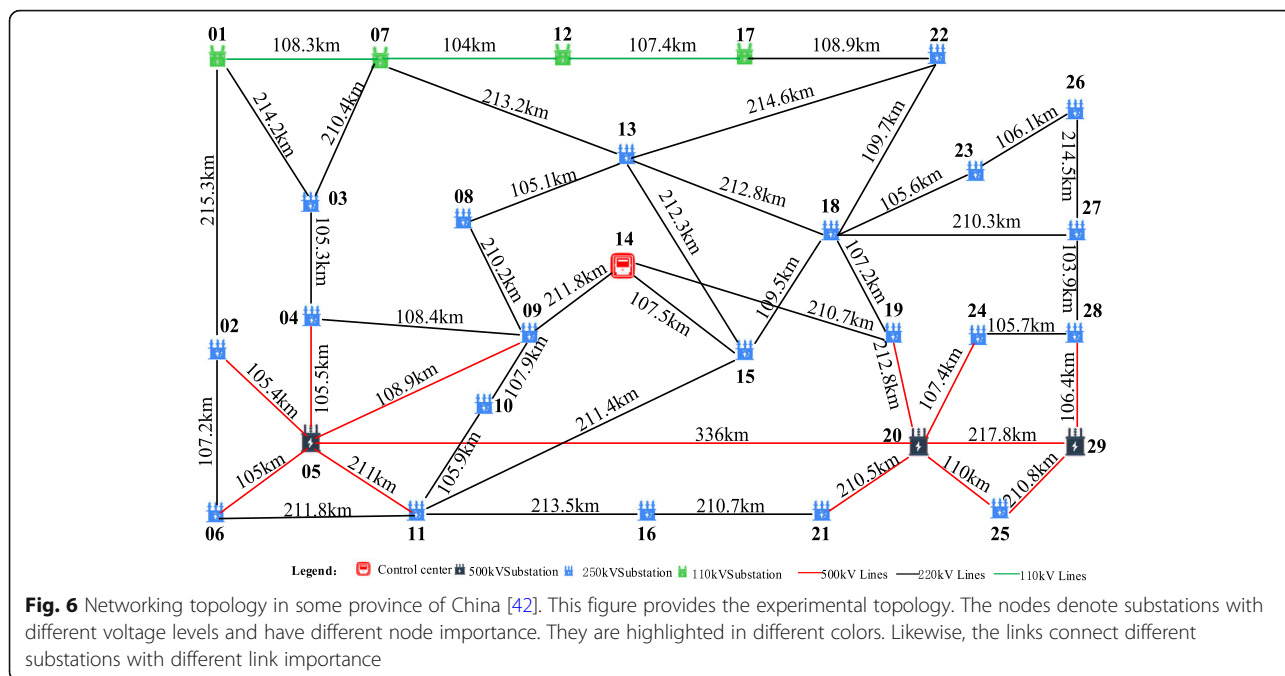
6.2 Overall performance of routing and algorithm

For simplicity and without loss of generality, there is an assumption that any node in the network can be considered the epicenter, and then we compute the average result for all the nodes. Since the process of node failures is the same as link failures, this research takes link failures as an example to validate the effectiveness of the proposed scheme. The changing trends of average reward with different approaches as well as convergence performance in different scenarios are demonstrated in Figs. 7 and 8. Here, the number of link failures (LF) is 3, and the number of affected services (AF) is set 15 and 25, respectively.

Figure 7 shows the average reward of the agent received from the environment varied with the number of episodes in the scenario of 3 link failures and 15 affected services. It can be seen that the average reward increases with the increasing number of episodes for all approaches, and Q-learning has a slighter advantage than the other two DQN-based approaches. This is because it is difficult to adjust the neural network parameters because of small state space and scarce training data. Furthermore, there is no obvious disparity between the prioritized replay resampling and uniform sampling because of the sparse samples. The overall performance of I-DQN approximates to the Natural-DQN. The performance of all algorithms is almost the same in the end.

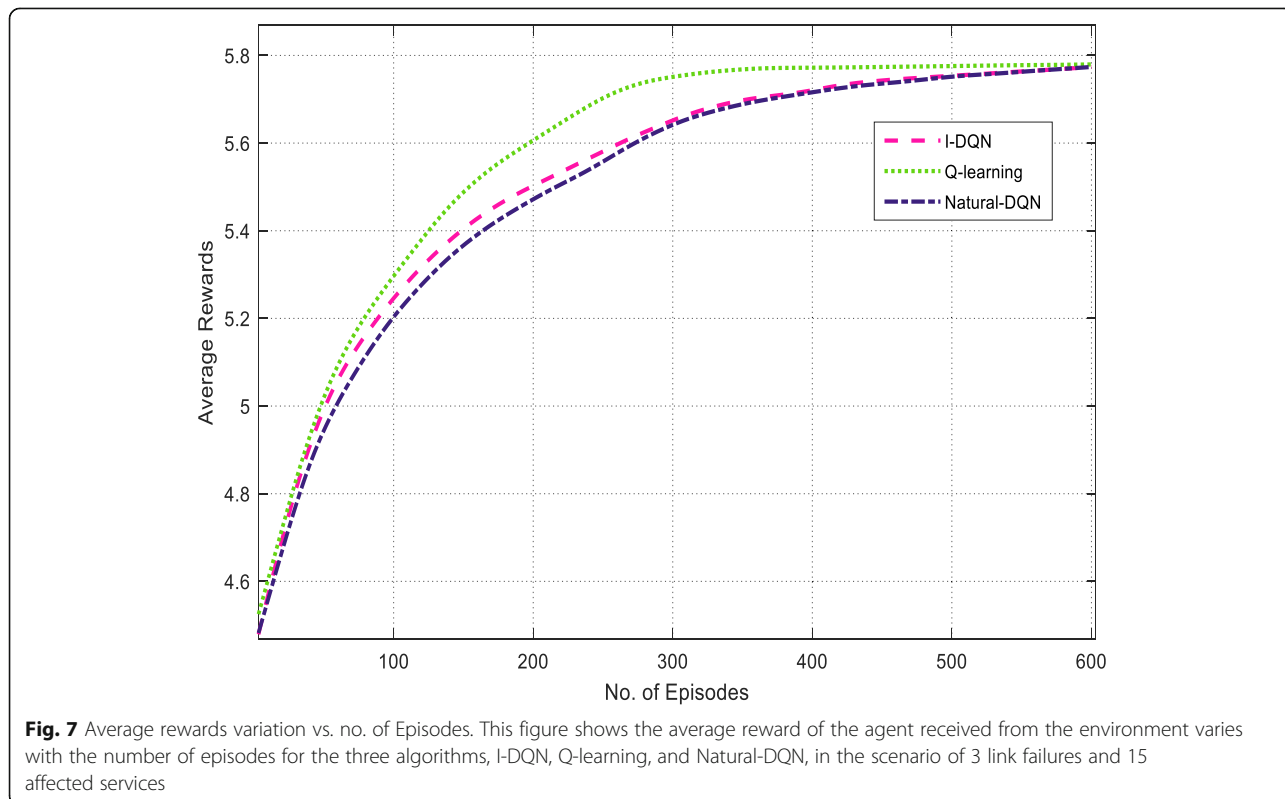
Figure 8 depicts the convergence performance for three approaches in different experimental scenarios. Figure 8a demonstrates that the average steps show a gradual decline trend with the increase of episodes for all algorithms with 3 link failures and 15 affected services. The metric of the average step refers to the experienced steps of the given episodes for the sampled transitions to achieve algorithm convergence. It is observed that the Q-learning decreases faster than other algorithms before 350 episodes and finally converges at 13 steps. After 350 episodes, the average steps of DQN-based algorithms have a faster speed than that of Q-learning, and they continuously decrease and finally are stable at 11 steps. The reason is that the neural network has not been trained well in the beginning so that the DQN-based approaches converge slowly. In the later episodes, more samples with a high quality are exploited to be trained, and the advantage of prioritized replay is becoming obvious.

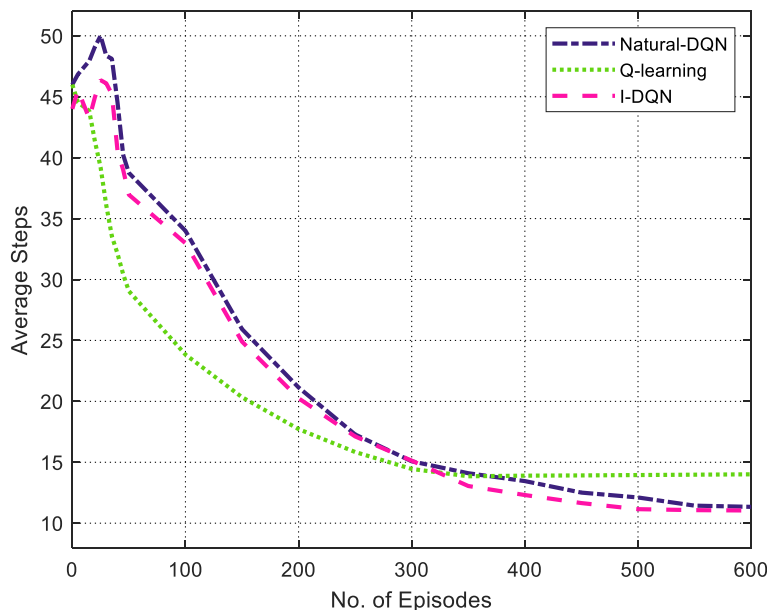
Figure 8b further compares the average steps with 3 link failures and 25 affected services for all algorithms. It is observed that the convergence speed of DQN-based algorithms drops faster than that of Q-learning, moreover, the approach of I-DQN expenses fewer average



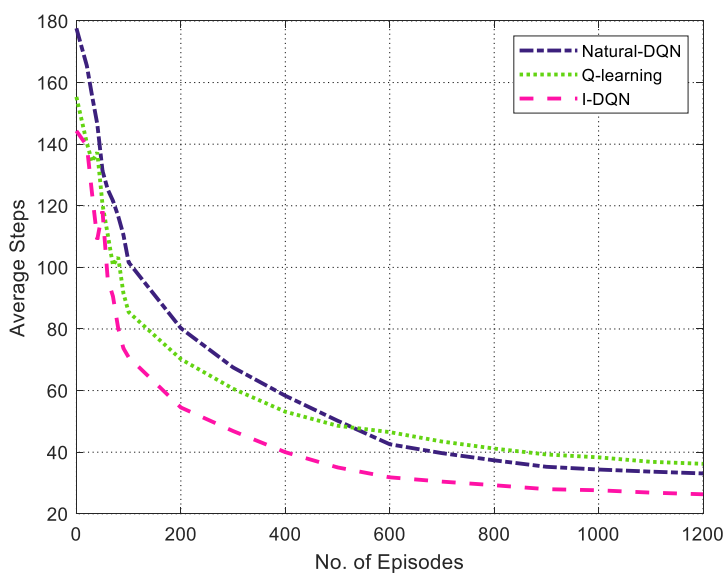
steps to find an optimal path combination than that of Natural-DQN and retains steady after about 800 episodes. The reason is that sufficient samples have to be stored with a prioritized resampling mechanism and make sure the neural network to be well trained.

Figure 9 depicts the average episodes as a function of the number of link failures. Note that the average episodes decrease for affected services with the increasing of link failures. Particularly, there is a sharp decline when the number of link failures is more than four. It is because more failures will take place around the





(a) LF=3, AS=15



(b) LF=3, AS=25

Fig. 8 Average steps variation vs. no. of episodes in different scenarios. This figure depicts the convergence performance comparison for three approaches, I-DQN, Q-learning, and Natural-DQN, in the situations of LF = 3, AS = 15, and LF = 3, AS = 25

epicenter in the scenario of large-scale failures. Therefore, the probability of information island apparently increases with the increase of link failures under the condition of the average node degree in the network is 3.2. As a result, there are comparatively fewer available paths for services which leads to smaller state space and action space, and thus the convergence speed is becoming faster. Also, we can find that the average episodes achieved the convergence state

increase with the increase of affected services leading to huge state space.

Figure 10 depicts the variation of average service recovery ratio versus the number of link failures. The recoverable service means that neither source nor destination of the service is isolated nodes and there are available paths satisfying service QoS requirements as well. The average recoverability can be represented as the ratio of the recoverable services and the total affected services. It is

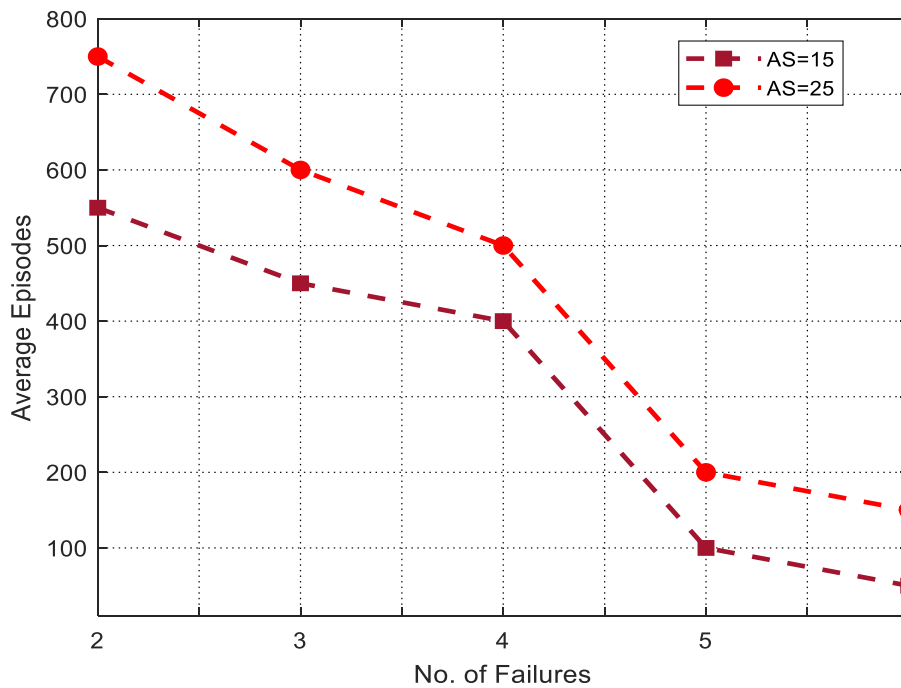


Fig. 9 Average episodes vs. no. of link failures. This figure shows the average episodes as a function of the number of link failures in the case of AS = 15 and AS = 25

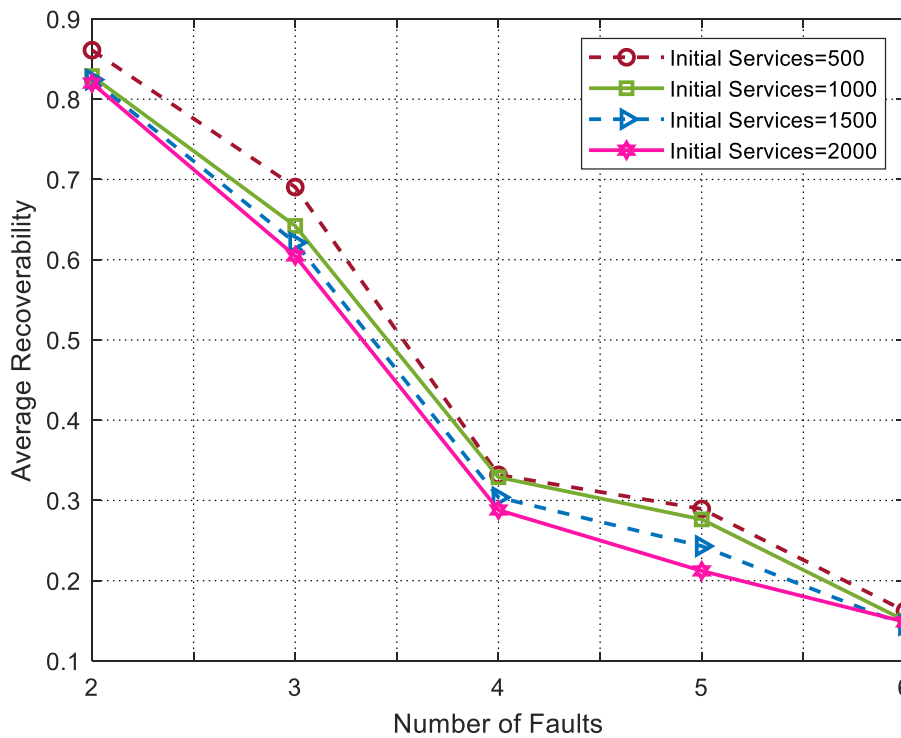


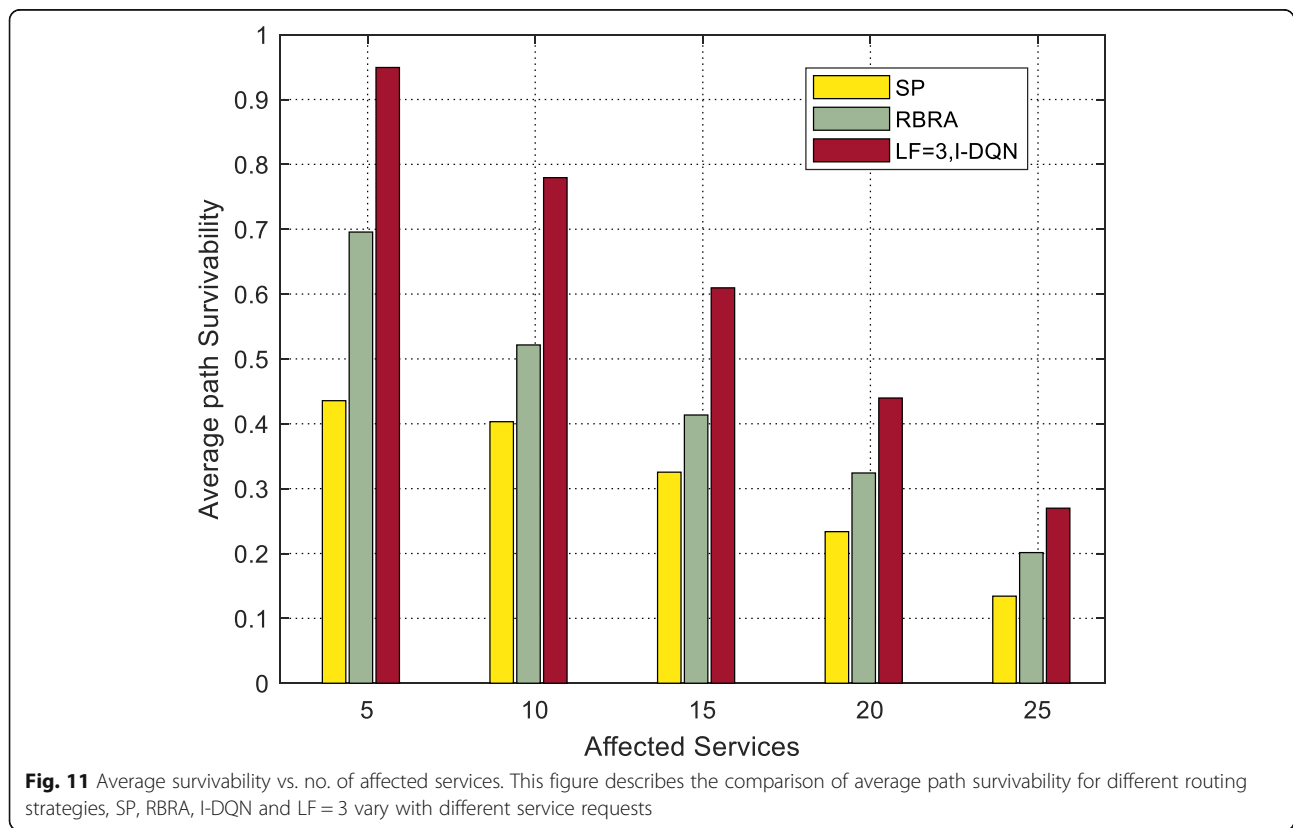
Fig. 10 Average recoverable ratio no. of link failures. This figure depicts the variation of average service recovery ratio versus the number of link failures when the initial number of services is 500, 1000, 1500, and 2000

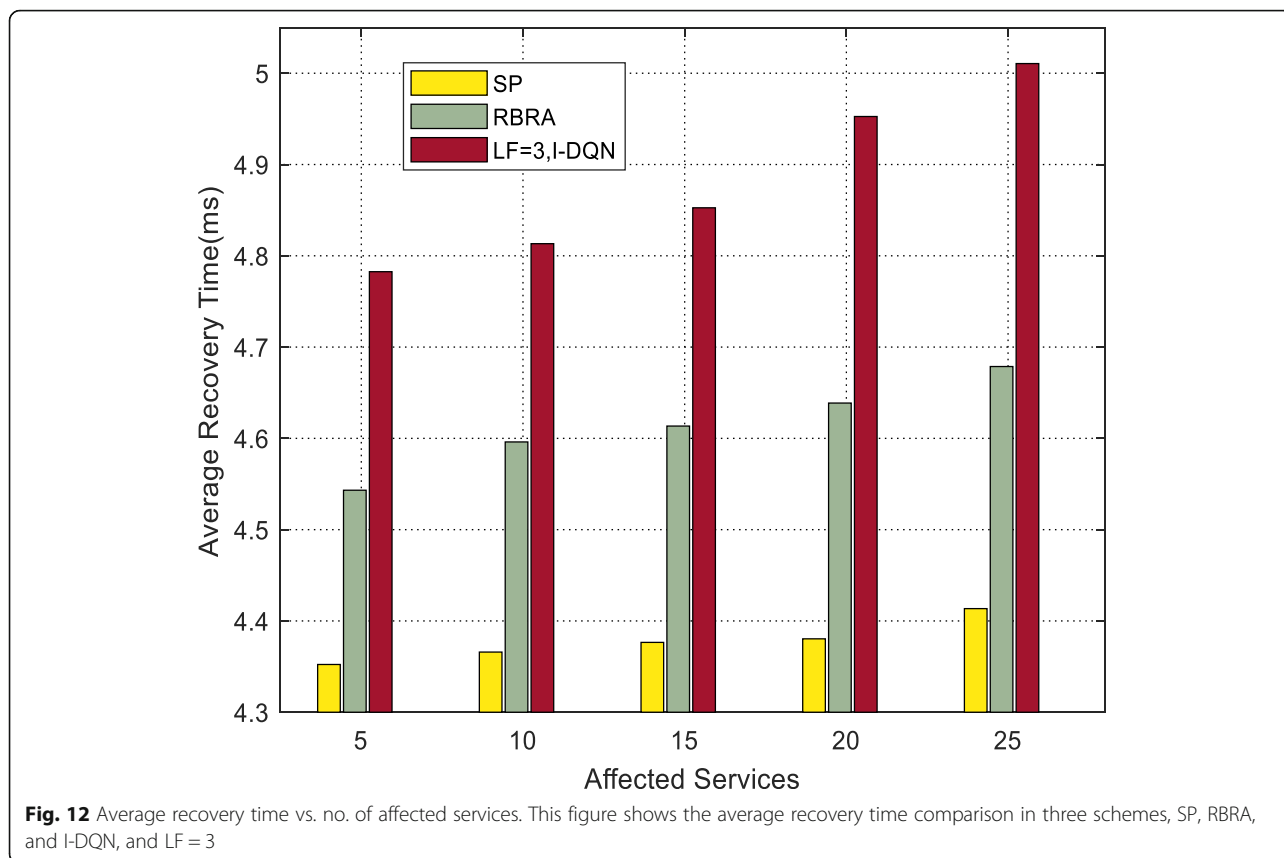
observed that the curve of the average recoverability suffers a sharp drop in the network. This is because the demands for specific services as well as the increase of link failures result in fewer available paths for services. Especially when the number of link failures is 6, the average recoverability is only 14.87%. Also, we notice that average recoverability shows a gradual decline trend with the increasing number of services. The reason is that more service requests cause more links to be unavailable due to insufficient bandwidth resources, and accordingly, fewer available routes are suitable for services transfer.

Figure 11 describes the average path survivability for different routing strategies versus different service requests. We can find that the I-DQN acquires the maximal path survivability for the three approaches. This is due to that the approach of SP is in pursuit of the shortest end-to-end latency path for services while the routing strategy of RBRA is searching routings for services with the objectives of simultaneously minimizing the balanced network risk and the average communication latency. Neither of them considers the path survivability factor. Whereas with our proposed I-DQN scheme, the metrics of the path survivability and the site difference level are the received reward of an agent from the environment, which encourages the agent to choose the routes with higher survivability to meet the required reliability for

services. Given the obvious QoS discrepancies, the more service rerouting requests are, the fewer available paths satisfying service requirements become. And thus, some services tend to choose paths closer to the epicenter to guarantee high path survivability. Hence, the path survivability for all routing strategies gradually drops. However, the proposed I-DQN obtained higher path survivability compared to other routing strategies.

The average recovery time is a critical metric to measure the performance of different rerouting strategies in case of large-scale failures. As can be seen from Fig. 12, I-DQN achieves the maximal average recovery latency in comparison with other schemes. This is because that the I-DQN scheme stimulated by the rewards tends to choose the paths with higher survivability for services that are far away from the epicenter. However, the maximal average recovery time is only 5.02 ms with 25 affected services in the network, which does not exceed the service threshold in SGCN. Meanwhile, it is observed that the average recovery time increase by 10.87% and 5.18% with the increasing average path survivability by 87% and 47% in comparison with the approaches of SP and RBRA. In addition, the average recovery latency increases with the increasing service requests for all strategies, and similarly, the reason has been illustrated in Fig. 11.





6.3 Time complexity analysis

The neural network in the experiment has two hidden layers. The number of neurons of the input and the output layer corresponds to the dimensions of state space n_{feature} and the of the action space n_{action} . For any iteration, the feed-forward the calculation needs three matrix operations, that is, the number of calculations is $n_{\text{feature}} \times n_1$, $n_1 \times n_2$, and $n_2 \times n_{\text{action}}$, then the time complexity is $O(n_{\text{feature}} \times n_1 + n_1 \times n_2 + n_2 \times n_{\text{action}}) = O(T)$. It is assumed that the I-DQN converges at M episodes, and then the overall time complexity is $O(T \times M)$.

In summary, due to the fluctuation of the communication environment, the conventional heuristic algorithms need to be reinitialized when the network topology and carried services change in the event of natural disasters. Additionally, the time complexity and the space complexity are increasing rapidly as with the larger scale of the problem, and the real-time performance and the flexibility reduced. However, the reinforcement learning algorithm enables to optimize the service paths combination autonomously through a well-trained neural network once the training process is finished, which can improve the timeliness of the communication system and reduce computing complexity. The above experimental results show that the average service recovery time is 5.02 ms, which can satisfy the end-to-end latency

requirement of services in SGCN. It further validates the effectiveness of the proposed approach.

7 Conclusions and future work

In this paper, we design a fast and reliable concurrent rerouting mechanism under the framework of DRL for all affected services in SGCN in case of large-scale failures. The experimental results demonstrate that the proposed solution I-DQN has a better convergence in a large-scale failure scenario compared with Q-learning and Natural-DQN with respect to algorithm efficiency. Also, this solution achieved higher path survivability in comparison with SP and RBRA in terms of different routing strategies. This is significant to the routing planning and optimization of SGCN in the quake-prone regions. Since multi-agent reinforcement learning has more advantages over the multi-service concurrent rerouting. Our future work will concentrate on further improving rerouting efficiency with multi-agent RL in large-scale failures.

Abbreviations

ADSS: All-dielectric self-supporting optic fiber cable; DL: Deep learning; DMS: Distribution management system; DRL: Deep reinforcement learning; I-DQN: Improved deep Q networks; IEDs: Intelligent electric devices; IoT: Internet of Things; MDMS: Meter data management system; MPLS: Multiple protocol label switching; OPGW: Optical fiber composite overhead ground wire; PCE: Path Computation Elements; PMU: Phasor

measurement units; RL: Reinforcement learning; SG: Smart grid; SGCN: Smart grid communication network; TED: Traffic engineering databases; WAMS: Wide-area measurement system

Acknowledgements

The authors acknowledge the National Key R&D Program of China (grant 2018YFE0205502), National Natural Science Foundation of China (61672108), and Science and Technology Project from of State Grid Corporation of China: "Key technology Research and Application for Ubiquitous Inter-networking Security for Power wireless private network services (No. SGJSXT00GCJS1900120)".

Authors' contributions

BJL proposed the main idea, performed the simulation, and analyzed the result. She is the main writer of this paper. PY gave some important suggestions for the paper. XSQ and LS have revised the paper. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Author details

¹State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, China. ²Pingdingshan University, Pingdingshan, China. ³Department of Computing, Carlow Institute of Technology, Carlow, Ireland.

Received: 4 September 2019 Accepted: 16 January 2020

Published online: 19 May 2020

References

1. Y Saleem, N Crespi, M H Rehmani, Internet of things-aided smart grid: Technologies, architectures, applications, prototypes, and future research directions. <https://arxiv.org/abs/1704.08977> [cs.NI] (2019)
2. S.S. Reka, T. Dragicevic, Future effectual role of energy delivery: A comprehensive review of Internet of Things and smart grid. *Renew. Sust. Energ.Rev.* **91**, 90–108 (2018). <https://doi.org/10.1016/j.rser.2018.03.089>
3. J. Kim, F. Filali, Y. Ko, Trends and potentials of the smart grid infrastructure: From ICT sub-system to SDN-enabled smart grid architecture. *Appl. Sci.* **5**(4), 706–727 (2015)
4. M. Karakus, A. Durrresi, Quality of service (QoS) in software defined networking (SDN): A survey. *J. Netw. Comput. Appl.* **80**, 200–218 (2017)
5. H C. Leligou, T. Zahariadis, L. Sarakis, E. Tsampasis, Smart grid: a demanding use case for 5G technologies. *IEEE PerCom Workshops*, Athens, Greece, pp. 215–220 (2018)
6. S. De Dutta, R. Prasad, Security for smart grid in 5G and beyond networks. *Wirel. Pers. Commun.* **106**(1), 261–273 (2019)
7. R. Chaudhary, G.S. Aujla, S. Garg, N. Kumar, SDN-enabled multi-attribute-based secure communication for smart grid in IIoT environment. *IEEE Trans. Ind. Inform.* **14**(6), 2629–2640 (2018)
8. X. Yin, A. Guo, T. Zhao, S. Li, Analysis of regional tectonic stress field characteristics of Changning Ms6.0 earthquake in Sichuan Province, J. *Earthquake Eng of China.* **41**(5), 1215–1220 (2019)
9. G. Zhang, Theory study on the relationship between the repetition time of large earthquakes and the duration of aftershocks. *Seismology Geology* **01**, 17–24 (1996)
10. A. Izaddoost, S.S. Heydari, Enhancing network service survivability in large-scale failure scenarios. *J. Commun. Netw.* **16**(5), 534–547 (2014)
11. Mubashir Husain Rehmani, Software defined networks based smart grid communication: A comprehensive survey. <http://arXiv:1801.04613v4> [cs.NI] 1–34, 27 Mar 2019
12. S. Rinaldi, F. Bonafini, P. Ferrari, A. Flammini, Characterization of IP-based communication for smart grid using software-defined networking. *IEEE Trans. Instrum. Meas.* **6**, 1–10 (2018)
13. B. Liu, P. Yu, F. Chen, F. Chen, Q. Xue-song, L. Shi, *Risk-Aware Service Routes Planning for System Protection Communication Network in Energy Internet* (IFIP/IEEE Symposium on Integrated Network and Service Management (IM), Washington D.C, 2019), pp. 295–303
14. C. Ji, P. Zhao, W. Li, *Comprehensive Vulnerability Assessment and Optimization Method for Smart Grid Communication Transmission Systems* (IFIP/IEEE Symposium on Integrated Network and Service Management (IM), Lisbon, 2017), pp. 975–978
15. X. Chen, P. Zhao, P. Yu, B. Liu, Risk analysis and optimization for communication end-to-end link interruption in SG cyber-physical system. *Int. J. Distrib. Sens. Netw.* **14**(1–12) (2017)
16. O. Dong, P. Yu, *A Service Routing Reconstruction Approach in Cyber-Physical Power System Based on Risk Balance* (IEEE/IFIP Network Operations and Management Symposium (NOMS), Taipei, 2018), pp. 1–6
17. J.O. Abe, H.A. Mantar, A.G. Yayimli, k-Maximally Disjoint Path Routing Algorithms for SDN, *International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery*, Xi'an, pp.499–508 (2015)
18. Peng, X. Gong, L. Guo, A survivability routing mechanism in SDN enabled wireless mesh networks: Design and evaluation. *China Commun.* **13**(7), 32–38 (2016)
19. S. Lin, I.F. Akyildiz, P. Wang, *QoS-Aware Adaptive Routing in Multi-layer Hierarchical Software Defined Networks: A Reinforcement Learning Approach* (International Conference on Services Computing, San Francisco, 2016), pp. 25–33. <https://doi.org/10.1109/SCC.2016.12>
20. X. Huang, T. Yuan, G. Qiao, Deep reinforcement learning for multimedia traffic control in software defined networking. *IEEE Network* **32**(6), 35–41 (2018)
21. Y. Du, F. Zhang, L. Xue, A kind of joint routing and resource allocation scheme based on prioritized memories-deep Q network for cognitive radio ad hoc networks. *Sensors* **18**(7), 2119–2140 (2018)
22. W. Zhang, X. Du, K. Nygard, T. Wang, *Self-Protecting Networking Using Dynamic p-Cycle Construction Within Link Capacity Constraint* (IEEE International Conference on Communications (ICC), Dresden, 2009), pp. 1–6
23. T. Elhourani, A. Gopalan, S. Ramasubramanian, I.P. Fast, Rerouting for multi-link failures. *IEEE-ACM Trans. Netw.* **24**(5), 3014–3025 (2016)
24. M. Bhor, D. Karia, *Network Recovery Using IP Fast Rerouting for Multi-link Failures* (International Conference on Intelligent Computing and Control (I2C2), Coimbatore, 2017), pp. 1–5
25. Y. Yang, M. Xu, Q. Li, Fast rerouting against multi-link failures without topology constraint. *ACM Trans. Netw.* **26**(1), 384–397 (2018)
26. B. Bassiri, S. Heydari, *Network Survivability in Large-Scale Regional Failure Scenarios* (ACM, Montreal, 2009), pp. 83–87
27. M Kiese, V. Marcheva, J. Eberspacher, D. Schupke, Diverse Routing Based on Shared Risk Link Groups, 7th International Workshop on Design of Reliable Communication Networks, Washington, DC, pp. 153–159(2009)
28. R. Li, X. Wang, X. Jiang, *Network Survivability Against Region Failure* (IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), Xi'an, 2011), pp. 1–6
29. S. Neumayer, G. Zussman, R. Cohen, E. Modiano, Assessing the Vulnerability of the Fiber Infrastructure to Disasters. *IEEE-ACM Trans. Netw.* **19**(6), 1610–1623 (2011)
30. Z. Cheng, X. Zhang, Y. Li, Congestion-aware local reroute for fast failure recovery in software-defined networks. *J. Opt. Commun. Netw.* **9**(11), 934–945 (2017)
31. D. Silver, A. Huang, C. Maddison, Mastering the game of Go with deep neural networks and tree search. *Nature.* **529**, 484–489 (2016)
32. Y. Saleem, K.A. Yau, H. Mohamad, N. Ramli, M.H. Rehmani, Q. Ni, Clustering and reinforcement-learning-based routing for cognitive radio networks. *IEEE Wirel. Commun.* **24**(4), 146–151 (2017)
33. R. Ding, Y. Xu, F. Gao, Deep reinforcement learning for router selection in network with heavy traffic. *IEEE Access* **7**, 37109–37120 (2019)
34. B. Li, C. Lu, D. Jing, C. Zhu, Y. Sun, B. Qi, An optimized routing algorithm with load and risk joint balance in electric communication network. *Proc. CSEE* **39**(09), 2713–2723 (2018)
35. B. Fan, Y. Zeng, L. Tang, Vulnerability assessment of power communication network based on information entropy. *J. Electron. Inf. Technol.* **36**(09), 2138–2144 (2014)
36. GB/T2900.50–2008, Electrotechnical terminology-generation, transmission and distribution of electricity General (2009)
37. J. Liu, F. Gao, X. Luo, Survey of deep reinforcement learning based on value function and policy gradient. *Chin. J. Comput.* **42**(06), 1406–1438 (2019)
38. Q. Liu, J. Zhai, Z. Zhang, S. Zhong, X.U.A. Jin, Survey on deep reinforcement learning. *Chin. J. Comput.* **41**(1), 1–27 (2018)
39. X. Chen, L. Cao, C. Li, Z. Xu, M. He, Deep reinforcement learning via good choice resampling experience replay memory. *Control Decis.* **33**(04), 600–606 (2018)
40. Mnih V, Kavukcuoglu K, Silver D, Playing Atari with deep reinforcement learning [EB/OL]. <https://arxiv.org/abs/1312.5602>, [2017-3-16]

41. C. Bai, P. Liu, W. Zhao, J. Tang, Active sampling for deep Q-learning based on TD-error adaptive correction. *J. Comput. Res. Dev.* **56**(02), 262–280 (2019)
42. B. Qi, S. Liu, B. Li, S. Chen, Optimized scheduling algorithm for power demand response service based on multi-priority services and SRLG. *Power Syst. Technol.* **43**(07), 2393–2402 (2019)
43. K. Qin, C. Huang, C. Wang, J. Shi, D. Wu, X. Chen, Balanced multiple controllers placement with latency and capacity bound in software-defined network. *J. Commun.* **37**(11), 90–103 (2016)

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
