Computer Communications 116 (2018) 225–239

Contents lists available at ScienceDirect

Computer Communications

journal homepage: www.elsevier.com/locate/comcom

A hop-by-hop dynamic distributed multipath routing mechanism for link state network

Haijun Geng*[,a], Xingang Shi[c,d], Zhiliang Wang[c,d], Xia Yin[b,d]

[a] School of Software Engineering, ShanXi University, 030006, China
[b] Department of Computer Science and Technology, Tsinghua University, China
[c] Institute for Network Sciences and Cyberspace, Tsinghua University, China
[d] Tsinghua National Laboratory for Information Science and Technology (TNList), China

ARTICLE INFO

Keywords:
Multipath routing
Distributed algorithm
Loop-free
Reliability

ABSTRACT

High reliability is always pursued by network designers, operators and users. Multipath routing can provide multiple paths for transmission and failover, and is considered to be effective in the improvement of network reliability.

Existing multipath routing algorithms are tailored to find as many paths as possible and to guarantee loop-freeness, with their computation or communication overhead often overlooked. For example, a router either constructs multiple shortest path trees, or exchanges information with its neighbors for every destination, witnessing a particularly high cost when connecting with a large number of neighbors. Moreover, these algorithms typically perform a recomputation from scratch whenever the network state changes, leading to further resource scarcity.

We propose an algorithm, DMPA (dynamic multipath algorithm), for a router in a link-state network to compute multiple next-hops for each destination. The algorithm constructs one single shortest path tree based on ordinary network link states, and dynamically adjusts it in response to network state changes, so the sets of next-hops can be efficiently computed and incrementally updated. At the same time, DMPA guarantees loop-freeness of the induced routing path by implicitly maintaining a partial order of the routers underpinning it. And also, we present a further extension of DMPA (DMPA-e) by incorporating hop count, a topology dependent metric.

We evaluate the proposed algorithms and compare them with several latest multipath algorithms, using some real and inferred ISP topologies, and also a set of synthetic topologies. Our results show that our proposed algorithms can provide good reliability and fast recovery for the network with very low overhead.

1. Introduction

With the rapid development of the Internet, more and more services and applications are widely deployed, which pose stringent requirements on its effectiveness and reliability. Traditional routing algorithms are mostly concerned with finding a shortest path towards the destination, thereby cannot provide good connectivity under frequent network failures [1]. This highlights the need for mechanisms that possess fast and efficient recovery capabilities. Towards this goal, multipath routing [2–10] has been proposed to use multiple alternative paths for data transmission. It not only can improve a network's resilience to topology changes, routing failures and traffic bursts, but also can facilitate better load balance and higher throughput.

Existing multipath routing schemes can be divided into two subcategories, by whether special cooperation/signaling between routers are required for packet forwarding. Cooperation-free schemes compute multiple next-hops for each destination, and each router independently selects an appropriate next-hop for standard packet forwarding, where care must be taken such that the induced forwarding paths are loop-free. The benefit is that they can provide not only redundant backup links, but also other features such as load balancing and high throughput. The other sub-category of schemes compute, for a link to protect, a multi-hop repair path that is agreed by all routers on that path. Thus special cooperation mechanisms have to be employed to reroute packets along that path.

In this paper, we focus on the first type of schemes. And also, we confined our work in the link-state routing networks. Most Internet Service Providers (ISPs) prefer link-state routing instead of distance-vector routing in their intra-domain system [11], because of its merits like fast convergence and good support for metrics. Layer2 networks are

* Corresponding author.
E-mail address: genghaijunzt@163.com (H. Geng).

https://doi.org/10.1016/j.comcom.2017.12.008
Received 18 July 2017; Received in revised form 8 November 2017; Accepted 15 December 2017
Available online 15 December 2017

also incorporating link-state routing into their network architecture, such as the standardized TRansparent Interconnection of Lots of Links (TRILL) [12]. On the other hand, during topology changes caused by link or node failure, millisecond level fast convergency is preferred [13], which poses stringent performance requirement to route computation [14].

Since the performance of routing and forwarding is critical for the Internet, multipath algorithms must be highly efficient so as not to become a bottleneck. Existing approaches often focus on finding paths as many as possible, but do not take much effort in reducing their computation or communication overhead. For example, they either build multiple shortest path trees on a router [2,7,8], or exchange information for every destination between neighboring nodes[1][3], so the induced cost will be particularly high for high-degree nodes, i.e. routers with a large number of neighbors. Although ISPs tend to keep their network topologies secret, cases been reported that some run ISIS over a full-mesh MPLS topology [13], some have routers with more than 400 neighbors, and it may not take long to reach the order of thousands of neighbors [15], where even the calculation of a shortest path tree can cost substantial CPU time [13,14]. On the other hand, when a link changes its state or weight, existing multipath algorithms will recompute all the next-hops for each destination from scratch, causing further resource scarcity and convergence delay.

We propose a tree based dynamic multipath algorithm, DMPA, for a link-state network. A node in the network only needs to compute a single shortest path tree (SPT), locally and independently, without exchanging with other nodes any information other than the ordinary link states. Rules to select the next-hops are designed such that when the tree is fully built, for each destination in the network, a specific set of next-hops are derived. Any routing path induced from the distributed computation results, i.e., hop-by-hop routing based on the next-hop sets computed by the nodes, is loop free. In addition, those next-hops can be incrementally updated in response to any link state change, instead of being computed from scratch. DMPA greatly improves algorithmic efficiency without sacrificing result quality, as will be demonstrated by our evaluations. Although a bunch of dynamic algorithms, i.e., Multi-Mesh of Trees (MMT) [16,17], have been proposed for ad-hoc networks, DMPA is the first dynamic multipath algorithm for loop-free path computation in networks running link state protocols like OSPF or IS-IS.

The rest of the paper is organized as follows. Section 2 introduces some background and related works. Section 3 presents DMPA and its properties in detail. In Section 4, the next-hop selection criterion in DMPA is further relaxed based on a specifically defined partial order. Then Section 5 compares DMPA with some latest multipath algorithms, and finally, Section 6 concludes the paper.

2. Related works

Network failures occure frequently in today's Internet. Therefore, lots of methods have been provided to deal with this problem in different ways. IETF has drafted a framework named IP Fast ReRoute (IPFRR) [18], which aims to provide fast recovery from network failures. The basic idea is that, when a node detects the failure of a link directly connected to itself, it can immediate switch to backup paths that are specifically computed for this failure.

Equal-cost multi-path (ECMP) [19] allows packets to be routed along multiple paths of equal cost, which can be specifically tuned by network operators. However, in certain cases it is just impossible to achieve equal costs no matter what link weights are used [20], where ECMP cannot offer good reliability.

Discount Shortest Path Algorithm (DSPA) [21] computes K-shortest paths and takes into account both path quantity and path independence. However, computing the K-shortest paths is still much more computationally intensive than finding a single shortest path. O2 [30] compute at least two disjoint next hops towards any given destination for each node. However, the time complexity of the approach is in $O(|V|^3)$.

Multi-topology routing [7,8] pre-computes multiple routes based on backup topologies tailored for specific failures, either by removing the corresponding edges or by increasing their associated costs. However backup paths computed in this way may not work well in the case of unplanned failures. In [22], multiple instances of a link-state routing protocol are employed to offer multiple choices, where each link is associated with a vector of weights which can be adjusted by end systems. Path splicing [3] is an enhancement to multi-topology routing. It creates a set of slices for the network based on random link-weight perturbations, and end system can control which slices the routers should use by embedding control bits in packet headers. The paper [23] presents a new routing method that combines the concept of permutation routings with joker links, called joker-capable permutation routings. IDAG [24] can be resilient under all single link faults by building two node/link independent configurations on a given topology, called the red tree and the blue tree.

Several multipath routing algorithms [2,25–27] compute loop-free alternatives based on the rule named Downstream Criterion (DC) [28,29], where packets can be forwarded to a next-hop which is closer to the destination than the local node, so that loop-freeness can be guaranteed. However, the cost of the criterion validity assessment increases proportionally to the degree of a node. The computational complexity increases as the number of network node degree, which is particular high when a node has a high degree. Source selectable deflection [2] requires shortest path tree computation for each neighbor or information exchange with each neighbor, so the overhead increases linearly with the degree of a node. SMRS [25–27] aims at designing a method to find more diverse loop-free paths, which is the extension of the deflection approach to 2-hop history information. Instantaneous loop-freeness has also been studied, where link-state [30–33] as well as distance-vector [5,34,35] multipath algorithms achieve loop-free at every instant.

Several algorithms [36,37] compute a Directed Acyclic Graph (DAG) for each destination to avoid loops, so in a network with $|V|$ nodes, their time complexity is in the order of $|V|$ times the cost of computing a single DAG, which is already more complicated than computing a SPT. More next-hops can be found at the cost of increasingly sophisticated DAGs, but there is no guarantee to find an alternate next-hop for each destination (or each link). In MARAs [36], authors propose a family of novel multipath route calculation algorithms, which construct a DAG that includes all edges in the network graph structure, in order to provide a significant number of alternative paths among all nodes to a destination. However, the time complexity of their approach is extremely high. In [37], authors proposed a centralized computation method to offer an effective alternative routing for guaranteed network connectivity. Several methods [38–41] employs the local rerouting and pre-computes interface-specific forwarding tables. However, all the algorithms need a very complicated pre-computation algorithm to handle network failures. In [42], authors investigate the optimal node-disjoint and link-disjoint solutions for maximizing network availability. For the first time, authors in [43] put forward the scheme to deal with single network component failure in segment routed networks. In [44], authors propose an algorithm tunneling on demand (TOD) approach to enhance the network reliability while bring little overhead. However, TOD need to maintain tunnels in most of the cases, which brings the burden on the router and network.

3. DMPA: a dynamic multipath algorithm

3.1. Notations and basic idea

Before formally describing our tree based dynamic multipath

[1] In this paper, we use router and node interchangeably.