Geodiverse Routing Protocol with Multipath Forwarding Compared to MPTCP

Yufei Cheng*, Truc Anh N. Nguyen*, Md. Moshfequr Rahman*, Siddharth Gangadhar*,
and James P.G. Sterbenz*†‡

*Information and Telecommunication Technology Center
Department of Electrical Engineering and Computer Science
The University of Kansas, Lawrence, KS, 66045, USA
† School of Computing and Communications (SCC) and InfoLab21
Lancaster LA1 4WA, UK
‡Department of Computing, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong
{yfcheng|annguyen|moshfequr|siddharth|jpgs}@ittc.ku.edu, jpgs@comp.{lancs.ac.uk,polyu.edu.hk}
www.ittc.ku.edu/resilinets

***Abstract*—This work presents a comprehensive performance comparison of our cross-layer resilient protocol stack, ResTP–GeoDivRP against Multipath TCP (MPTCP). A profile-based challenge model is used to better represent different failure scenarios. Furthermore, our resilient protocol stack is implemented in the network simulator *ns-3* and emulated in the KanREN testbed. The GeoDivRP routing protocol collects network statistics and calculates multiple geodiverse paths; these paths are provided upstack to our resilient transport protocol, ResTP, for resilient multipath communications. By providing multiple geodiverse paths, our ResTP–GeoDivRP protocol stack provides better path protection against regional failures than MPTCP.**

***Index Terms*—network resilience and survivability; multiple flow-diverse routing; geographically-correlated challenges; geodiverse protocols; network simulation and testbed experiments;**

I. Introduction and Motivation

The demands for Internet resilience have been increasing tremendously. Telecommunication networks are widely used for carrying Internet traffic and they rely heavily on physical infrastructure to maintain normal operation, and it is important to analyze their resilience to various faults and challenges [1]. Survivable optical networks under non-correlated failures have been a popular research topic [2], [3]; IP-level restoration mechanisms, have been studied [4] and shown to be effective. Recently, the research community has become more concerned about the potential damage caused by geographically-correlated challenges. This can result in significant damage to dependable network communications [5]. *Geographically correlated challenge* is defined as a failure or malicious attack that affects a series of nodes and links in a geographic vicinity. Since the failure effect is frequently long-term [6], a set of backup paths is required for survivable routing. Some may argue that in order to increase network resilience to regional failures, it should be considered during the network planning phase. However, the high cost and policy limitation of deploying new fibers have hindered the improvement of resilience through new physical-level diversity.

Network simulation has played a key role in failure analysis and protocol design. We have modeled both correlated failures and malicious attacks with simulation results [7]. We have proposed and analyzed two routing heuristics for the path geodiverse problem (PGD), in which any two nodes on disjoint paths are separated by a d-distance separation criteria, and demonstrated its effectiveness under geographically-correlated challenges [8]. We have studied traffic allocation optimization formulation [9] to statistically direct the rerouted traffic onto multiple d-distance separated paths that can better cope with network congestion when regional failure occurs.

In this work, we present comprehensive performance comparison of our resilient cross-layered flow-diverse protocol stack to MPTCP [10], [11]. It provides network protection and improves resilience by taking advantage of multiple geodiverse paths for each communicating node pair. The GeoDivRP routing protocol analyzes information collected from the link layer, in particular the failed nodes and links set along with the link delay and congestion level information. A *distance separation criteria* is calculated and used for the geodiverse path calculation. A parallel optimization process is performed using the collected information for optimal traffic allocation. The initial path set followed by the optimized traffic allocation set are passed up to ResTP for resilient flow establishment. Their relative performance is compared using ns-3 network simulator as well as our KanREN testbed.

The remainder of the paper is organized as follows. Section II describes background and related work. Section III presents our novel cross-layer design within the flow-diverse resilient protocol stack. Section IV details results comparing MPTCP to ResTP over GeoDivRP in the face of challenges and Section V concludes this paper.

II. Background and Related Work

Traditional intra-domain routing protocols are designed to form a single shortest path for each source–destination pair due to its simplicity and efficiency. However, this comes with the cost of not having the option to choose an alternate path in

978-1-4673-9023-1/16/$31.00

case the current one is unavailable due to failures or attacks. In order to quickly bypass the failed region, a resilient protocol is required to quickly find a single or multiple alternative paths for the communicating node pairs.

A. Current Intra-Domain Routing Protocol

Intra-domain routing protocols run within an Internet service provider (ISP) network, and most ISPs run a link-state routing protocol based on configurable link weight. The link weights are tuned by network operators, such as for load balancing, failure avoidance, or security. The two primary intra-domain routing protocols are Open Shortest Path First (OSPF) and Intermediate System to Intermediate System (IS-IS); we focus on OSPF in this work. Each router running OSPF has a static link weight configured for its outgoing link and the shortest path is calculated by Dijkstra's algorithm [12] based on the link-state advertisements (LSA) flooded throughout the network. Each router constructs its forwarding table based on the calculated shortest path. OSPF has several advantages: First, routing is very simple since it is based on a single metric, the link weight. Second, by flooding the information in the network using the LSA packet, each router has consistent view of the topology and routing loops are easily avoided. Finally, it is scalable using a hierarchy and reliable to single node failure due to its distributed nature.

However, there are some drawbacks due to its simple operation. First, the path is calculated only based on link weight: other information, such as end-to-end delay or congestion level cannot be considered during path calculation. Second, alternative paths are not provisioned and the path restoration process is slow. Although an end system or edge network could have access to multiple paths, routers are not able to use the path diversity or the geodiversity the network topology has to offer [13], [14]. To control route flapping, OSPF introduces several timeout values, which slows the protocol convergence. Since OSPF needs to reconverge whenever the topology changes due to network failures, the process becomes even slower with geographically-correlated challenges.

Several mechanisms have been proposed to address the above mentioned drawbacks. Constrained shortest path first (CSPF) [15] protocol is an extension to calculate the paths fulfilling a set of constraints, with different classes of traffic forwarded to different paths. However, the constraints cannot be dynamic demands such as path delay or jitter. In this work, we introduce our optimization engine to consider dynamic traffic information for optimal traffic distribution. Second, equal-cost multipath (ECMP) is commonly deployed in which routers keep track of several shortest paths and then evenly split the flow among them [16]. Another mechanism for fast restoration is Fast-IP rerouting [17]; instead of only calculating one shortest path in the network restoration process, a second path is calculated to provide protection for the primary one.

B. Multipath Routing

Multiple shortest paths enable the network operators to balance load and provide better resilience by splitting traffic into k multiple paths. Since each router has a consistent view of the topology though LSAs, a k-shortest path algorithm can be used for each node pair. However, this is not realistic in practice due to its high computational cost. For example, for a single node pair, Yen's algorithm [18] has a complexity of $O(kn(m+n\log n))$ on a graph with n nodes and m links for a dense network. Furthermore, the forwarding table size will be k times larger. To reduce the complexity, path splicing [19] proposes a new routing mechanism that uses multiple instances of the link-state routing protocol; each link has a vector of link weights with each one tuned for different traffic class. For example, one can tune for high bandwidth use while another for low delay.

Several multipath architectures have been proposed for resilient traffic communication. A distributed traffic engineering heuristic, TeXCP, has been proposed that uses four paths for each demand [20], however, it may potentially be misguided by a near-optimal solution. The cross-layer routing paths problem has been proposed by maximizing the objective function for users implementing multipath routing [21]. When regional failures occur, the rerouting traffic has the tendency to share common links in the vicinity of the threat zone and increase the congestion possibility. Multipath mechanisms can minimize the after-challenge traffic impact on the *hot* links as well as the whole network. Furthermore, splitting traffic onto different paths strategically may potentially provide more throughput. Multipath routing has been widely studied as an effective mechanism to reduce congestion in *hot spots* by deviating traffic to unused network resources [16], [22], [23]. Several methods for load balancing using multipath routing without survivability measures have been researched. For example, optimization has been done to maximize the flow on each path in the ECMP routing algorithm [24]. Another optimization problem has been formulated by a weighted multipath routing based on ECMP; its objective function is to minimize the maximum link utilization [25]. Resilient Overlay Networks [26], [27] has been studied to improve the robustness and availability of current Internet paths. However, from a traffic engineering perspective, multipath routing is advantageous for small networks only for the all-commodity traffic scenario, yet the multipath gain diminishes as the network becomes large [28].

When multiple next-hop addresses are installed in router forwarding tables, forwarding and scheduling mechanisms have to be redesigned. Multi-Topology Routing (MTR) constructs multiple topologies with different link weight configurations and enables separate forwarding mechanisms on a per-topology basis. It is a simple mechanism in which each packet can switch among topologies [19], [29]. A source routing deflection mechanism uses tags to apply path diversity for multipath routing [30]. Another proposed approach is to forward traffic on all paths that make forwarding progress towards the destination [31] using only one set of link weights. Each router makes local forwarding decisions and loop-free paths are guaranteed since the forwarding decision at each hop is using a shorter-hop path towards the destination.

C. Resilient Multipath Architecture

Most of the traffic allocation and multipath routing studies assume normal network connectivity or single link failure [22], [32]. This is widely studied and considered as an effective fault-tolerant multipath routing mechanism for a single link failures [33], [34]. Geographically-correlated challenges have been studied under the current intra-domain routing framework. A circular disk failure model is used to model geographical failures [35]. The single-location challenge scenario has been analyzed and a polynomial algorithm has been formulated [36]. Correlated and simultaneous challenges have been discussed [37], and different circular-shaped vulnerability points have been identified [38]. During the geo-correlated challenge, the traffic allocation follows the widest paths disjoint with respect to the bottleneck links. The bottleneck links from multiple paths are mutually disjoint to increase resilience [39]. An optimization problem has been formulated to model the issues in a multi-source-destination routing environment, which leads to a pseudo-polynomial algorithm based on linear programming with a bounded buffer size and skew constraint [40]. For path skew analysis in the multipath routing context, past work calculates a number of shortest paths and selects the ones that meet the skew requirement. The returned paths are then used to solve the optimization problem [40]. A multipath flow optimization problem has been formulated with two objectives, total link utilization and bandwidth fairness, and has been solved with a nonlinear programming solver [41]. However, with the increasing importance of network resilience under large-scale failures or attacks, it is imperative to analyze multipath routing efficiency and understand the traffic allocation requirements under these challenges.

Several resilient transport protocols have been proposed. mTCP [42] can aggregate the bandwidth of several paths concurrently and improve resilience using the redundant path. Multipath transfer (CMT) can distribute data across multiple end-to-end paths in a multi-homed devices to achieve efficient parallel data transfer [43]. MultiPath TCP (MPTCP) [10], [11] enables simultaneous use of several network interfaces to establish multiple subflows for a host pair. The goal is to provide better throughput and survivability to failures while preserving the regular TCP interface to applications. However, there is no control over how the multiple paths for different subflows are calculated. Our ResTP [44]–[46] is a resilient general-purpose transport layer protocol. By employing a set of reliability mechanisms that are composable and tunable, it is flexible in efficiently supporting various application classes operating across different network environments with distinct characteristics. It establishes multiple transport flows for its data transmission by taking advantage of the geodiverse path set and the traffic allocation information provided by the GeoPath Diverse Routing Protocol GeoDivRP [9], [14]; ResTP can either actively spread the data over multiple available paths to survive a single path failure with no disruption or transmit the data on one path while leaving another as a hot-standby for rapid failover. This work considers only the multipath mode in which all paths transport traffic. In addition to multipath spreading capability, ResTP also provides other transport-layer services to the application layer, including multiplexing/demultiplexing, adaptive flow/subflow management, flexibly composable error control, and flow control and congestion control. As noted above, with the goal of supporting a variety of application types, each of these services is comprised of multiple composable mechanisms [47]. ResTP chooses among its various reliability mechanisms to satisfy the specific application it is servicing according to the particular mission requirements. In this work, we provide a comparison of our ResTP–GeoDivRP protocol stack with MPTCP to demonstrate the performance gain through geodiverse paths; we further present the effectiveness of our protocol stack in face of geocorrelated challenges.

III. Resilient cross-layer design

We have proposed ResTP–GeoDivRP protocol stack for path and flow diverse resilient traffic transmission [14], [48]. The application passes a service specification (ss) and threat model (tm) down to the transport layer protocol, ResTP (resilient transport protocol). ResTP then requests GeoDivRP to calculate geodiverse paths that meet the requirement tuple $(k, d, [h, t])$, where k is the total number of geodiverse paths requested, d is the distance separation criteria, $[h, t]$ is the optional path stretch (number of additional hops for diverse paths) and skew (delay difference across paths) t constraint. GeoDivRP interprets path stretch h as path delay (latency) l. Based on the configuration and network statistics ($[l, f]$) collected from the network monitor engine and the requirement tuple from ResTP, GeoDivRP calculates the geodiverse path set $P_k = P_0 ... P_{k-1}$. The statistic f represents the node and link failure information. It passes the configuration (P_k, l, t) to the optimization engine as shown in Figure 1. Based on the latency (l) and skew (t) requirement, the optimization engine returns the path set P_k along with its traffic allocation information X_k to GeoDivRP, which are then passed up to ResTP for establishing subflows. ResTP is still under active development and not complete; for this paper, we are only using the cross-layering feature of ResTP to take advantage of the geodiverse paths returned by GeoDivRP. We extend our protocol stack to reliable multiple path forwarding and provide detailed performance comparison against MPTCP.

We propose GeoDivRP [8], [14] to provide multiple geodiverse paths to ResTP for resilient traffic transmission. In order to decrease the complexity of geodiverse path calculation, our iterative WayPoint Shortest Path (iWPSP) heuristic [8], [29] is used. Using the distance separation criteria d provided by the failure monitor model, iWPSP returns a set of geodiverse paths for the input graph $G = (V, E, w)$, where V is the node set, E is the link set, and w is the link weight set. To demonstrate how the protocol stack calculates GeoPaths, we briefly explain how iWPSP works. As shown in Figure 2, for the case when $k = 3$, iWPSP first selects neighbor nodes v_{s_1} and v_{d_1} that are d-distance separated from source node v_{s} and destination node v_{d}, respectively. For simplicity, this

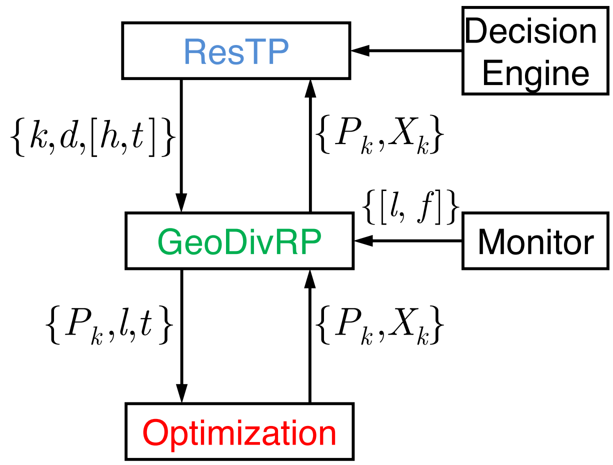

Fig. 1. Cross-layer protocol stack overview

work assumes that such nodes exist; otherwise, the nodes with the greatest distance will be chosen, iterating until nodes d apart are located. Assuming the shortest path connecting v_s and v_d is p_s, iWPSP selects waypoint nodes m' and m'' in the opposite direction that are distance $d+\delta$ apart from the middle node m in the shortest path, where the segment $m'mm''$ intersects the shortest path. Dijkstra's algorithm is performed for the two branches $v_{s_1}m'$ and $v_{d_1}m'$. By connecting the shortest path returned from the two branches, the heuristic obtains the first geodiverse path. The same mechanism repeats for waypoint node m'' for the second geodiverse path. The variable d is a user-chosen parameter based on the threat model, and δ is experimentally chosen for different network topologies to increase the probability of the heuristic to return a d-separated path. The δ parameter is also useful in preventing the links of the two geodiverse paths from interleaving and creating routing loops. By tweaking the value of δ, the heuristic can select a nearby waypoint node if the previous one fails running Dijkstra's algorithm. When the heuristic cannot select paths within the skew bound t, the model increases or decreases δ accordingly. The geodiverse path is the primary reason for performance gain when comparing to MPTCP during geo-correlated challenges. The geodiverse paths set is further passed to the optimization model for flow allocation information.

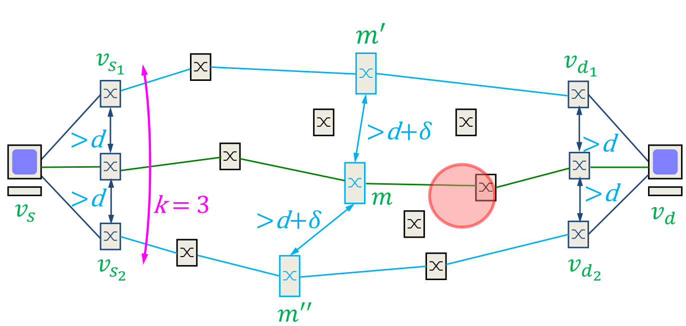

Fig. 2. Iterative waypoint shortest path heuristic

We design our optimization engine to be dynamically adapting to the path selections. This means that it can calculate flow distribution based on changing network metrics such as path delay and jitter. Since the optimization model requires a centralized view of the topology and fairly large processing power [9], the geodiverse paths without optimization are returned to ResTP for flow establishment immediately after calculation; the optimization engine is running in parallel and returns the flow distribution information when finished to ResTP for it to adapt to current network condition. This initial result is promising [9] and we plan to fully test the integration and present result in future work.

IV. Performance Comparison

We present a comprehensive performance analysis of our ResTP–GeoDivRP protocol stack against MPTCP [11], [49] using multiple node-disjoint paths. The *ns-3* [50] simulator is used to demonstrate our protocol stack performance compared to an implementation of MPTCP [49]. All the nodes in the topology are ResTP–GeoDivRP enabled, and path protection using multiple geodiverse paths is provided by GeoDivRP. Three geodiverse paths are used in all the considered topologies for survivable routing. Various challenge profiles are used to represent different failure or attack scenarios. We further experiment our protocol stack on our KanREN [51] testbed and compare it to a kernel version of MPTCP [11]. As ResTP is not yet implemented in Linux kernel, we are using multiple TCP connections between a node pair to emulate ResTP capabilities.

A. Challenge profiles

A set of challenge profiles is used to systematically study their impact on the network connectivity and how our protocol stack performs in the face of these challenges.

- The Midwest challenge profile shown in green circles represents a super-storm sweeping trajectory
- The coastline challenge profile along the East Coast shown in blue circles represents a hurricane trajectory
- The cascading challenge profile such as power blackout affects a region growing in size as shown in red circles

As shown in Figure 3, the Sprint physical network [52] is presented with several challenge profiles. The movement for the Midwest profile is from the southwest to the northeast direction representing a super-storm, while the coastline profile moves in a similar direction but on the East Coast representing a hurricane. It also has a larger challenge radius compared to the Midwest profile. The profiles provide a better understanding of how different challenge locations and trajectories affect the protocol stack performance. For example, the cascading challenge profile shown in Figure 3 as red circles concentrates on the infrastructure with most shortest path occurrences. It renders the network more difficult to maintain normal network connectivity in the face of challenges since the affected nodes are forming a large percentage of shortest paths connecting the west and the east coast. In this work, we apply a cascading challenge profile in the Sprint [52] and Level 3

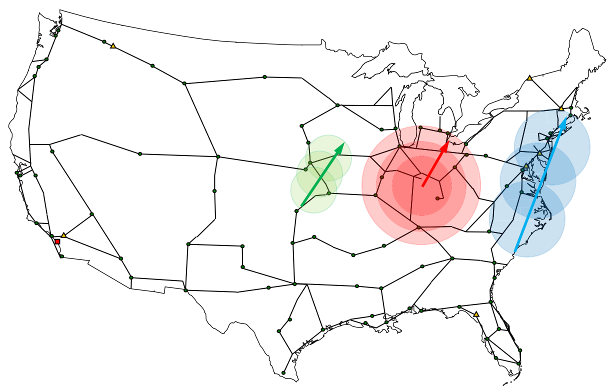

Fig. 3. Sprint network topology challenge profile

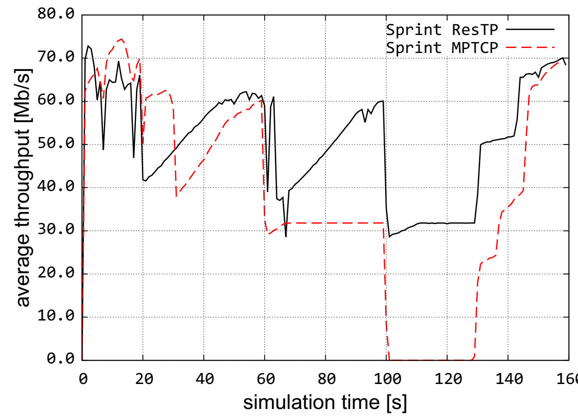

Fig. 5. Sprint network ResTP throughput compared to MPTCP

networks [53], and also study how the Midwest profile affects our KanREN [51] testbed.

B. Simulation results

The cascading challenge profile is applied in the Sprint network shown in Figure 4. In this profile, three challenge scenarios are included. The challenge begins at 20 s and grows larger in radius with each challenge lasting for 20 s. It originates around Nashville and grows larger in range; the green small circle challenge occurs at 20 – 40 s, the yellow circle challenge at 60 – 80 s, and the red large challenge at 100 – 120 s. The traffic originates from Oklahoma City to Washington D.C. and the bandwidth on each link is 100 Mb/s. The dashed line represents the paths calculated for MPTCP. These are node-disjoint paths calculated using Suurballe's algorithm [54], [55], and it cannot guarantee all the paths are geographically disjoint. In this experiment, the requirement tuple from ResTP has $k = 3$ and a varying failure distance d. MPTCP uses St. Louis, Kansas City, and Atlanta as its next-hops for the three paths. The second challenge shown in yellow circle fails two of MTCP's paths and the third challenge fails all three of them. For all the simulation cases in this work, we only carry out one simulation for each protocol.

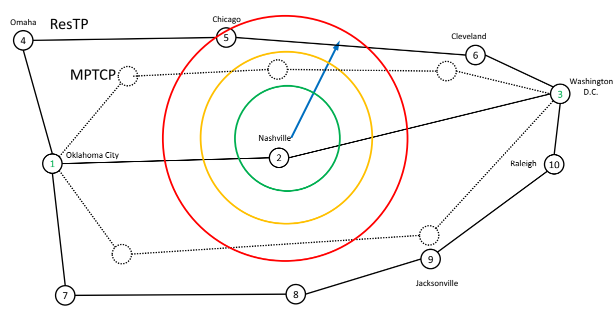

Fig. 4. Sprint network topology with multiple paths

On the other hand, GeoDivRP guarantees the paths are geographically disjoint, and therefore with all the subflows created by ResTP geodiversity, the protocol stack can provide higher throughput and better resilience to cascading challenges. As shown in Figure 4, GeoDivRP uses Omaha, Nashville, and Houston as its next-hops. This guarantees that for any regional challenges with a radius no larger than the distance separation criteria d, 2/3 of the paths will survive the regional challenge when the topology allows; for the last challenge scenario, since there is no alternative path to the north of Chicago for the Sprint network, GeoDivRP can only provide one backup path. The traffic still originates from Oklahoma City to D.C. with the solid line representing the paths calculated for ResTP. The paths are provided by GeoDivRP using iWPSP heuristic [14] with the d-distance separation guaranteed.

Figure 5 plots the average throughput across the three paths against the simulation time. The throughput starts from zero and approaches 70 Mb/s at the beginning of the simulation until the first challenge occurring at 20 s. MPTCP does not guarantee the distance separation among the multiple paths used in the simulation; therefore, with circular radius d challenge, each circular failure can take down two or all paths at the same time if occurs at the right location. From 20 – 40 s, with the shortest path failed, both ResTP and MPTCP reduce in throughput. From 60 – 80 s, ResTP obtains 2/3 of the full average throughput while 1/3 for MPTCP since two of its paths are failed. The worst performance for MTCP occurs from 100 – 120 s as all three of its paths are failed. Overall, ResTP presents around 30% to 40% performance increase compared to MPTCP in face of regional challenges.

A similar challenge profile is applied in the Level 3 network [53] as shown in Figure 6. The failure region is shown in the KU TopView [52], [56] to better present the overall topology and how the challenge affects nodes and links. Similar to the previous experiment, the failures in color-coded circles represent the cascading challenge growing in size. Two red dots represent the source node at Denver, CO and destination node at Indianapolis, IN. The outbound red arrow from Denver shows the shortest path, while the two green arrows represent the two alternative paths calculated by GeoDivRP.

As shown in Figure 7, when the first challenge is introduced at 20 s, both ResTP and MPTCP see a drop in throughput. At

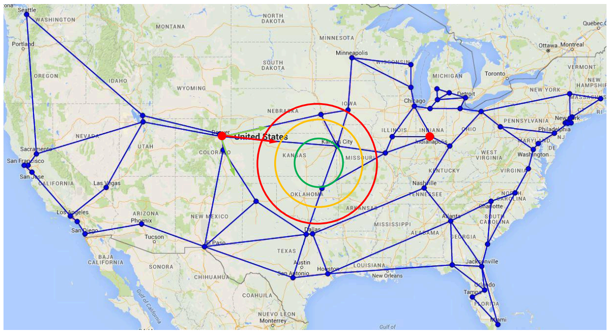

Fig. 6. Level 3 cascading challenge scenario

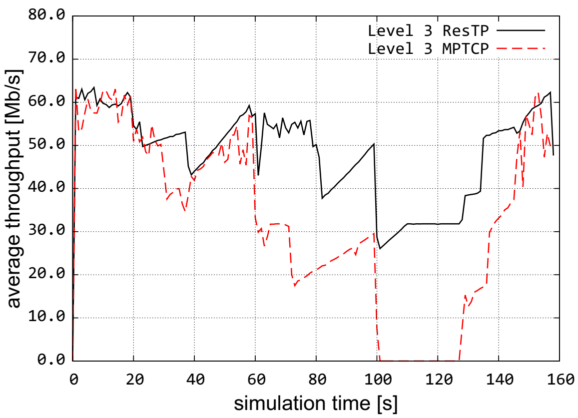

Fig. 7. Level 3 network ResTP throughput compared to MPTCP

60 s, ResTP is able to restore 2/3 the throughput by using the geodiverse path provided by GeoDivRP. On the other hand, the throughput for MPTCP further reduces as it lost another path. The worst performance for MPTCP occurs at 100 – 120 s since all its three paths are failed during this period, while ResTP with cross-layer path information can still achieve 30 Mb/s throughput. After 120 s with all the challenges elapsed, both of the protocols restore back to normal operation.

C. KanREN Testbed Result

In our KanREN-GENI testbed, we have deployed OpenFlow-enabled switches in the Kansas Research and Education Network (KanREN) [51], which is a logical ring throughout the state of Kansas connecting institutions of higher education. Eight Brocade NetIron CES 2024C [57] OpenFlow switches have been deployed at these institutions, as shown in Figure 8. A full-mesh topology is deployed as an OpenFlow overlay from which any arbitrary virtual topologies can be initialized through Multiprotocol Label Switching (MPLS) [58] tunnels. A ring topology is used in our experiment. Floodlight [59] is an OpenFlow controller based on Java that works with both physical and virtual switches. Our resilience routing framework controls the switches through Floodlight.

We study our KanREN physical switch topologies using two failure scenarios as presented in Figures 9 and 10, in

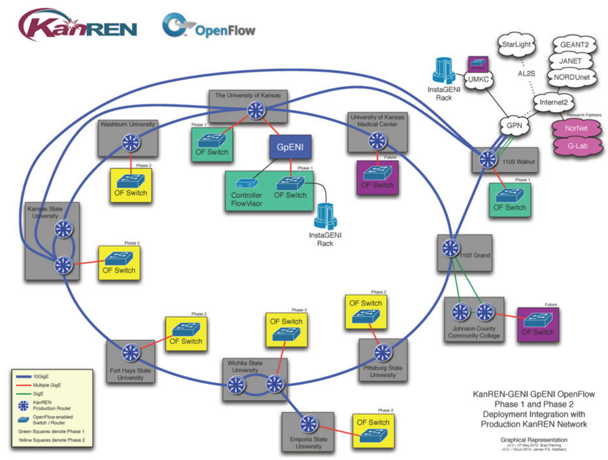

Fig. 8. KanREN OpenFlow switches deployment

which multiple paths exist between Lawrence and Wichita. The traffic originates from Lawrence to Wichita with all the link bandwidths at 100 Mb/s. A Midwest challenge profile is applied over the topology starting from Pittsburg and moving towards Topeka. The initial failure takes effect at 20 s bringing down the Pittsburg switch and lasts for 20 s. The next failure occurs at Emporia starting at 60 s lasts for 20 s as well. The last failure circle encompasses Topeka at 100 s for 20 s. Finally, the failure moves away from the topology with all switches up after 120 s. When the smaller failure radius is introduced as shown in Figure 9, only one path is failed at any given time. Since the paths calculated by GeoDivRP are d-distance separated, the challenge cannot affect two paths at the same time and our protocol maintains normal communication throughout the simulation time. We are using multiple TCP connections to emulate ResTP abilities in the testbed.

For the larger failure radius shown in Figure 10, two paths are failed during each challenge. The challenge begins at Pittsburg and follows the same trajectory as the previous case. Each challenge lasts for 20 s as well. But with the larger radius, each failure can take down two paths at the same time. For example, failure f_2 fails both Emporia and Pittsburg and failure f_3 fails both Topeka and Emporia. GeoDivRP maintains at least one working path during each failure and the worst case performance for our protocol is at 30 Mb/s.

Results from the above challenge profile is shown in Figure 11. The traffic is generated using iPerf [60], a network framework for evaluating the network's maximum bandwidth. For the first 20 s, the average throughput for the two scenarios is around 90 Mb/s. Starting at 20 s, the average throughput drops to just above 60 Mb/s due to the challenge f_1. At 50 s, the throughput for the smaller radius scenario increases to 85 Mb/s, yet the larger failure scenario stays at 30 Mb/s. At 90 s, the throughput for the smaller failure scenario increases to 90 Mb/s, while that of the smaller scenario stays at 60 Mb/s due to its additional failed path. At 100 – 120 s, the failure f_4 occurs and both scenarios perform the same due to the similarly failed path. Overall, the smaller failure scenario has

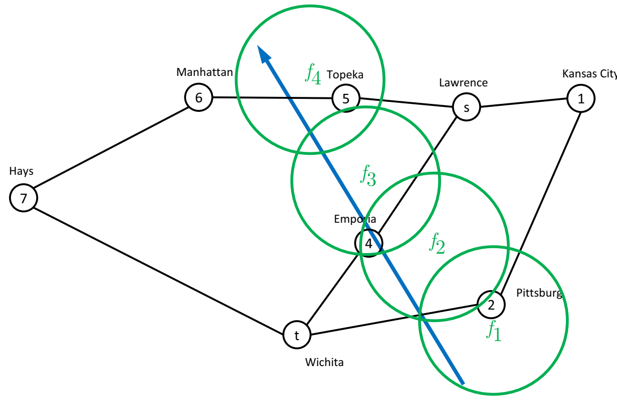

Fig. 9. KanREN small failure radius

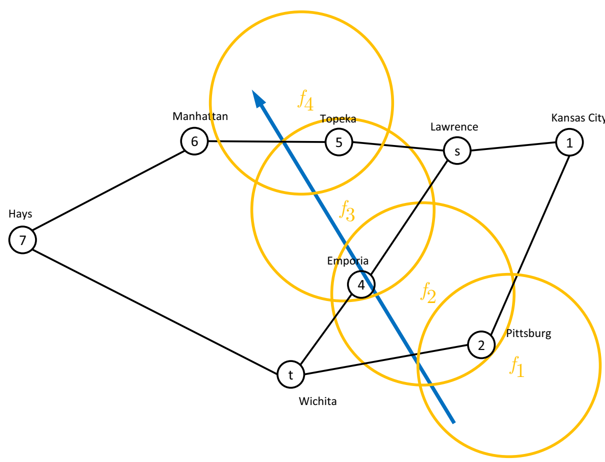

Fig. 10. KanREN large failure radius

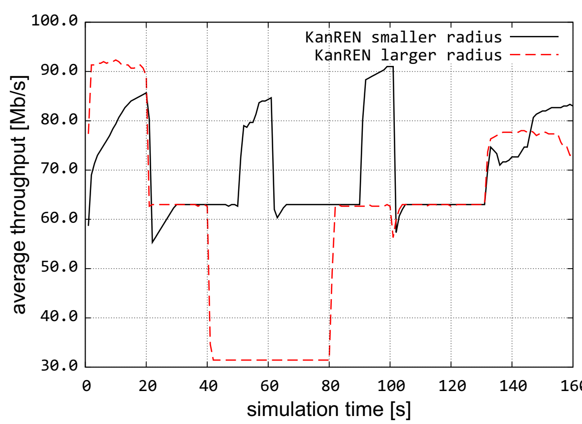

Fig. 11. KanREN testbed experiment results

around 30% of performance gain.

V. Conclusion and Future Work

This paper has presented the ResTP–GeoDivRP protocol stack and demonstrated its efficiency in bypassing the challenged regions and its improvement in terms of throughput compared to MPTCP. By calculating and selecting multiple geographic diverse paths, the protocol stack meets the resilience requirement for different applications. ResTP–GeoDivRP guarantees two working paths for resilient data transmissions as long as the topology allows, which guarantees reliable traffic delivery if erasure coding is considered. Overall, our proposed protocol stack provides higher throughput and resilience than MPTCP in the face of geographically-correlated challenges.

For future work, we plan to explore how erasure coding improves our protocol performance. Furthermore, we plan to test the real-time failure detection module of our testbed. Finally, we plan the complete integration of our ResTP–GeoDivRP protocol stack with different resilient modes enabled and compare it to MPTCP in different challenge profiles.

Acknowledgments

The authors would like to thank the members of the ResiliNets group for discussions to led to this work. This research was supported in part by NSF grant CNS-1128122 (KanREN-GENI) and CNS-1219028 (Resilient Network Design for Massive Failures and Attacks).

References

[1] J. P. G. Sterbenz, D. Hutchison, E. K. Çetinkaya, A. Jabbar, J. P. Rohrer, M. Schöller, and P. Smith, "Resilience and survivability in communication networks: Strategies, principles, and survey of disciplines," *Computer Networks*, vol. 54, no. 8, pp. 1245–1265, 2010.

[2] R. Cohen, K. Erez, D. ben Avraham, and S. Havlin, "Resilience of the Internet to Random Breakdowns," *Phys. Rev. Lett.*, vol. 85, pp. 4626–4628, November 2000.

[3] D. Magoni, "Tearing down the Internet," *IEEE J.Sel. A. Commun.*, vol. 21, pp. 949–960, September 2006.

[4] G. Iannaccone, C.-N. Chuah, S. Bhattacharyya, and C. Diot, "Feasibility of ip restoration in a tier 1 backbone," *IEEE Network*, vol. 18, pp. 13–19, Mar 2004.

[5] Y. Cheng, J. Li, and J. P. G. Sterbenz, "Path Geo-diversification: Design and Analysis," in *Proceedings of the 5th IEEE/IFIP International Workshop on Reliable Networks Design and Modeling (RNDM)*, (Almaty), September 2013.

[6] M. J. F. Denise M. B. Masi, Eric E. Smith, "Understanding and Mitigating Catastrophic Disruption and Attack," *Sigma Journal*, pp. 16–22, September 2010.

[7] E. K. Çetinkaya, D. Broyles, A. Dandekar, S. Srinivasan, and J. P. G. Sterbenz, "Modelling Communication Network Challenges for Future Internet Resilience, Survivability, and Disruption Tolerance: A Simulation-Based Approach," *Telecommunication Systems*, vol. 52, no. 2, pp. 751–766, 2013.

[8] Y. Cheng, M. T. Gardner, J. Li, R. May, D. Medhi, and J. P. Sterbenz, "Optimised Heuristics for a Geodiverse Routing Protocol," in *Proceedings of the IEEE 10th International Workshop on the Design of Reliable Communication Networks (DRCN)*, (Ghent, Belgium), pp. 1–9, April 2014.

[9] Y. Cheng, D. Medhi, and J. P. Sterbenz, "Geodiverse Routing with Path Delay and Skew Requirement under Area-based Challenges," *Networks (Wiley)*, vol. 66, pp. 335–346, Dec. 2015.

[10] A. Ford, C. Raiciu, M. Handley, and O. Bonaventure, "TCP Extensions for Multipath Operation with Multiple Addresses." RFC 6824 (Experimental), Jan. 2013.

[11] C. Paasch and S. Barré, "Multipath TCP in the Linux Kernel." http://www.multipath-tcp.org, January 2013.

[12] E. W. Dijkstra, "A Note on Two Problems in connection with Graphs," *Numerische Mathematik*, vol. 1, pp. 269–271, 1959.

[13] R. Teixeira, K. Marzullo, S. Savage, and G. M. Voelker, "Characterizing and Measuring Path Diversity of Internet Topologies," *SIGMETRICS Perform. Eval. Rev.*, vol. 31, pp. 304–305, June 2003.

[14] Y. Cheng, M. T. Gardner, J. Li, R. May, D. Medhi, and J. P. Sterbenz, "Analysing GeoPath Diversity and Improving Routing Performance in Optical Networks," *Computer Networks*, vol. 82, pp. 50–67, May 2015.

[15] M. Ziegelmann, *Constrained Shortest Paths and Related Problems - Constrained Network Optimization*. Saarbrücken, Germany, Germany: VDM Verlag, 2007.

[16] C. Hopps, "Analysis of an Equal-Cost Multi-Path Algorithm." RFC 2992 (Informational), Nov. 2000.

[17] M. Shand and S. Bryant, "IP Fast Reroute Framework." RFC 5714, Oct. 2015.

[18] J. Y. Yen, "Finding the k Shortest Loopless Paths in a Network," *management Science*, vol. 17, no. 11, pp. 712–716, 1971.

[19] M. Motiwala, M. Elmore, N. Feamster, and S. Vempala, "Path Splicing," in *Proceedings of the ACM SIGCOMM*, (Seattle, WA), pp. 27–38, August 2008.

[20] S. Kandula, D. Katabi, B. Davie, and A. Charny, "Walking the Tightrope: Responsive Yet Stable Traffic Engineering," in *Proceedings of the 2005 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, SIGCOMM '05, (New York, NY, USA), pp. 253–264, ACM, 2005.

[21] M. Wang, C. W. Tan, W. Xu, and A. Tang, "Cost of Not Splitting in Routing: Characterization and Estimation," *IEEE/ACM Trans. Netw.*, vol. 19, pp. 1849–1859, Dec. 2011.

[22] C. C. Aggarwal and J. B. Orlin, "On Multiroute Maximum Flows in Networks," *Networks*, vol. 39, no. 1, pp. 43–52, 2002.

[23] B. H. Shen, B. Hao, and A. Sen, "On Multipath Routing using Widest Pair of Disjoint Paths," in *High Performance Switching and Routing, 2004. HPSR. 2004 Workshop on*, pp. 134–140, 2004.

[24] M. Dzida, M. Zagozdzon, M. Zotkiewicz, and M. Pioro, "Flow Optimization in IP Networks with Fast Proactive Recovery," in *Telecommunications Network Strategy and Planning Symposium, 2008. Networks 2008. The 13th International*, vol. Supplement, pp. 1–9, Sept 2008.

[25] J. Zhang, K. Xi, L. Zhang, and H. Chao, "Optimizing Network Performance Using Weighted Multipath Routing," in *Computer Communications and Networks (ICCCN), 2012 21st International Conference on*, pp. 1–7, July 2012.

[26] D. Andersen, H. Balakrishnan, F. Kaashoek, and R. Morris, "Resilient Overlay Networks," in *Proceedings of the Eighteenth ACM Symposium on Operating Systems Principles*, SOSP '01, (New York, NY, USA), pp. 131–145, ACM, 2001.

[27] S. Savage, T. A. A. Aggarawl, T. Anderson, A. Aggarwal, D. Becker, N. Cardwell, A. Collins, E. Hoffman, J. Snell, A. Vahdat, G. Voelker, and J. Zahorjan, "Detour: a Case for Informed Internet Routing and Transport," *IEEE Micro*, vol. 19, pp. 50–59, 1999.

[28] X. Liu, S. Mohanraj, M. Pioro, and D. Medhi, "Multipath Routing from a Traffic Engineering Perspective: How Beneficial Is It?," in *Network Protocols (ICNP), 2014 IEEE 22nd International Conference on*, pp. 143–154, Oct 2014.

[29] M. Gardner, R. May, C. Beard, and D. Medhi, "Using Multi-Topology Routing to Improve Routing during Geographically Correlated Failures," in *Design of Reliable Communication Networks (DRCN), 2014 10th International Conference on the*, pp. 1–8, April 2014.

[30] X. Yang and D. Wetherall, "Source Selectable Path Diversity via Routing Deflections," in *Proceedings of the ACM SIGCOMM*, (Pisa), pp. 159–170, September 2006.

[31] D. Xu, M. Chiang, and J. Rexford, "Deft: Distributed exponentially-weighted flow splitting," in *IEEE INFOCOM 2007 - 26th IEEE International Conference on Computer Communications*, pp. 71–79, May 2007.

[32] R. Banner and A. Orda, "Multipath Routing Algorithms for Congestion Minimization," *Networking, IEEE/ACM Transactions on*, vol. 15, pp. 413–424, April 2007.

[33] M. Menth, R. Martin, A. M. Koster, and S. Orlowski, "Overview of Resilience Mechanisms Based on Multipath Structures," in *Design and Reliable Communication Networks, 2007. DRCN 2007. 6th International Workshop on*, pp. 1–9, IEEE, 2007.

[34] A. Iselt, A. Kirstadter, A. Pardigon, and T. Schwabe, "Resilient Routing using MPLS and ECMP," in *High Performance Switching and Routing, 2004. HPSR. 2004 Workshop on*, pp. 345–349, IEEE, 2004.

[35] Y. Kobayashi and K. Otsuki, "Max-flow min-cut theorem and faster algorithms in a circular disk failure model," in *IEEE INFOCOM 2014 - IEEE Conference on Computer Communications*, pp. 1635–1643, April 2014.

[36] S. Neumayer, G. Zussman, R. Cohen, and E. Modiano, "Assessing the Vulnerability of the Fiber Infrastructure to Disasters," *IEEE/ACM Transactions on Networking*, vol. 19, no. 6, pp. 1610–1623, 2011.

[37] P. K. Agarwal, A. Efrat, S. Ganjugunte, D. Hay, S. Sankararaman, and G. Zussman, "Network Vulnerability to Single, Multiple, and Probabilistic Physical Attacks," in *Military Communications Conference, 2010 - MILCOM 2010*, pp. 1824–1829, 2010.

[38] Y. Cheng and J. P. Sterbenz, "Critical region identification and geodiverse routing protocol under massive challenges," in *Reliable Networks Design and Modeling (RNDM), 2015 7th International Workshop on*, pp. 14–20, Oct. 2015.

[39] S. Nelakuditi and Z.-L. Zhang, "On Selection of Paths for Multipath Routing," in *Proceedings of the 9th International Workshop on Quality of Service*, IWQoS '01, (London, UK, UK), pp. 170–186, Springer-Verlag, 2001.

[40] T. Anjali, A. Fortin, G. Calinescu, S. Kapoor, N. Kirubanandan, and S. Tongngam, "Multipath Network Flows: Bounded Buffers and Jitter," in *INFOCOM, 2010 Proceedings IEEE*, pp. 1–7, March 2010.

[41] T. Ishida, K. Ueda, and T. Yakoh, "Fairness and Utilization in Multipath Network Flow Optimization," in *Industrial Informatics, 2006 IEEE International Conference on*, pp. 1096–1101, Aug 2006.

[42] M. Zhang, J. Lai, A. Krishnamurthy, L. Peterson, and R. Wang, "A transport layer approach for improving end-to-end performance and robustness using redundant paths," in *Proceedings of the USENIX Annual Tchnical Conference*, (Boston, MA), June 2004.

[43] J. Iyengar, P. Amer, and R. Stewart, "Concurrent Multipath Transfer Using SCTP Multihoming Over Independent End-to-End Paths," *Networking, IEEE/ACM Transactions on*, vol. 14, pp. 951–964, Oct 2006.

[44] T. A. N. Nguyen, J. P. Rohrer, and J. P. G. Sterbenz, "ResTP–A Transport Protocol for FI Resilience," in *Proceedings of the 10th International Conference on Future Internet Technologies (CFI)*, June 2015.

[45] J. P. Rohrer, A. Jabbar, and J. P. Sterbenz, "Path Diversification for Future Internet End-to-End Resilience and Survivability," *Springer Telecommunication Systems*, vol. 56, pp. 49–67, May 2014.

[46] J. P. Rohrer, R. Naidu, and J. P. G. Sterbenz, "Multipath at the Transport Layer: An End-to-End Resilience Mechanism," in *Proceedings of the IEEE/IFIP International Workshop on Reliable Networks Design and Modeling (RNDM)*, (St. Petersburg, Russia), pp. 1–7, October 2009.

[47] D. Feldmeier, "An overview of the TP++ transport protocol project," in *High Performance Networks: Frontiers and Experience* (A. N. Tantawy, ed.), vol. 238 of *Kluwer International Series in Engineering and Computer Science*, ch. 8, Boston, MA, USA: Kluwer Academic Publishers, 1993.

[48] Y. Cheng, M. M. Rahman, S. Gangadhar, M. J. Alenazi, and J. P. Sterbenz, "Cross-Layer Framework with Geodiverse Routing in Software-Defined Networking," in *ManSDN*, pp. 348–353, Nov. 2015.

[49] D. G. Fernández, "Multipath tcp in ns-3." https://github.com/dgomezunican/multipath-ns3.13, 2013.

[50] "The ns-3 Network Simulator." http://www.nsnam.org, July 2009.

[51] "The Kansas Research and Education Network (KanREN)." http://www.kanren.net/, December 2009.

[52] J. P. Rohrer, M. J. F. Alenazi, and J. P. G. Sterbenz, "ResiliNets Topology Map Viewer." http://www.ittc.ku.edu/resilinets/maps/, January 2011.

[53] "Level 3 Network Map." http://maps.level3.com.

[54] J. W. Suurballe, "Disjoint Paths in a Network," *Networks*, vol. 4, no. 2, pp. 125–145, 1974.

[55] J. W. Suurballe and R. E. Tarjan, "A Quick Method for Finding Shortest Pairs of Disjoint Paths," *Networks*, vol. 14, no. 2, pp. 325–336, 1984.

[56] A. Cosner, J. P. Rohrer, M. J. F. Alenazi, and J. P. G. Sterbenz, "ResiliNets Topology Map Viewer Version 2." http://www.ittc.ku.edu/resilinets/maps/v2, January 2015.

[57] http://www.brocade.com/, 1995.

[58] E. Rosen, A. Viswanathan, and R. Callon, "Multiprotocol Label Switching Architecture." RFC 3031 (Proposed Standard), Jan. 2001.

[59] "The Floodlight Controller." http://www.projectfloodlight.org/, 2015.

[60] "iPerf - The network bandwidth measurement tool." https://iperf.fr/, March 2003.