

Congestion Aware Multipath Routing: Aggregation Network Applicability and IPv6 Implementation

Matej Kultan

Institute of Telecommunications, Faculty of Electrical Engineering and Information Technology, Slovak University of Technology in Bratislava, Slovakia E-mail: matej.kultan@ut.fei.stuba.sk

Martin Medveck ý

Institute of Telecommunications, Faculty of Electrical Engineering and Information Technology, Slovak University of Technology in Bratislava, Slovakia E-mail: medvecky@ktl.elf.stuba.sk

Abstract—Currently, the service provider network capacity utilization is low due to the shortest multipath based routing protocols Opens Shortest Path First (OSPF) and Intermediate System-to-Intermediate System (ISIS). Due to inefficient routing approach, certain paths can be overloaded and link capacity is required while alternative paths are unused. The overall network has to be dimensioned with higher link bandwidth requirements introducing additional line, linecard, routing engine and overall solution cos. This paper provides improved Congestion Aware Multipath Routing (CAMRv2) algorithm overview. The new network routing algorithm allows higher throughput, network load-balancing and stability to ensure lower congestion and data drop on critical links. The algorithm discovers unused network resources and dynamically adapts to the actual traffic load and displacement. The focus in this paper is on new parameters for path computation performance improvement. Additionally, detailed IPv6 source routing CAMRv2 implementation for parallel coexistence with present networks is presented. Finally, the new routing algorithm is simulated in several scenarios over aggregation network. The result of simulations have proved better performance and resource utilization of the proposed algorithm in sparse aggregation network in terms of load-balancing between uplinks to the core network

Index Terms—Multipath routing, stability factor, load-balancing, congestion avoidance, IPv6.

I. INTRODUCTION

The Congestion Aware Multipath Routing (CAMR) approach provides completely new options for network optimization in terms of higher throughput for any flow in the network, link stabilization, efficient usage of resources and thus overall cost efficiency.

Firstly, the packet networks over the world rely on the

set of standardized routing protocols e.g.: OSPF, ISIS providing only statistically optimal paths between source and destination. Even mechanisms for load-balancing rely on statistical methods in terms of packet and data distribution e.g.: ECMP, UCMP, LAG [1] etc. The CAMR is providing optimal path set selection, relying on link load and topology knowledge, thus able to react more efficient in dynamic environment. The selection based on knowledge instead of estimation, provides higher routing performance, far behind selecting normally, statistically suboptimal paths.

Frequently occurs, the traffic demands are not homogeneously spread in the network and in the time. The network demands on performance are changing specially when providing services on a common converged network for L2/L3 Business Service customers, Residential and high-bandwidth LTE market [2]. Therefore, there is a need of routing load-balancing mechanisms to distribute peak throughput demands over the network and avoid congestions and data discard.

Secondly, ISPs are designing the new network with traffic estimation growth, considering mainly the average and the peak bandwidth requirements observed from the past. The balanced and stable network provide better estimation for network designers for optimal hardware investments. The new CAMR algorithm also helps with loadbalancing, regardless of the topology, network density or link theoretical maximum capacity.

Proposed Congestion aware multipath routing CAMR offers multipath routing for unequal bandwidth links using live congestion feedback feature. The new model offers increased network resiliency, end-to-end bandwidth maximization, fair dynamic link balancing in time and therefore overall higher efficiency of the network. The CAMRv2 algorithm with the proposed CAMPv2 protocol address improved network throughput and stability.

In this paper we analyze in detail the new algorithm basic algorithm principles and phases, implementation principles and we describe the influence of introduced Stability factor to achieve better performance over the typical aggregation network infrastructure. As the protocol and algorithm are based on congestion feedback, the network routing decisions are changing more dynamically than decisions based on link nominal bandwidth. Therefore, we have decided to perform several scenarios related to node routing initialization, step flow level change in the aggregation network to verify the. Scenario results have proved expected stable and efficient algorithm behavior.

II. RELATED WORK

In the Adaptive Multi-Path (AMP) Routing approach [3], the algorithm is using per destination address hashing for flow classification. In comparison to CAMR, proposed algorithm during flow classification is hashing also source address and protocol number fields. This provides more granularity and better load-balancing properties. The Optimized Multi-Path (OMP) routing mentioned in the thesis, requires global signaling and it is distributed implementation of AMP.

In the case of MultiRoute [4] approach only 1 bit is representing the congestion level. CAMR requires more granular representation of congestion level. This is required to be able to find shortest available multipath bandwidth instead of shortest available multipath paths. Low granularity implied by binary congestion level representation lead to route flapping and instability.

Max Flow Multipath Routing Algorithm (MFMP) [5] provides also multipath routing using similar to Edmonds-Karp, the Ford-Fulkerson algorithm. The algorithm tries to find full graph to provide maximum flow, but this approach requires high computation time. CAMR is stopping algorithm before finding full max flow graph matching criteria for efficient routing.

III. CAMRv2 DESIGN BASIC PRINCIPLES

The new CAMRv2 algorithm (Fig. 1.) is based on 5 principal phases, identical for all PE nodes:

- 1) Source/Destination selection
- 2) Path and flow search
- 3) Metric calculation
- 4) Metric to interval calculation
- 5) Encapsulation and proportional data distribution

All mentioned phases are independent in terms of computing and they run in parallel. In control plane, the Source/Destination selection phase is triggered by timer and topology update event. While the Path set and flow search phase is triggered by Link-State congestion vector arrivals. According the found Path Sets for individual destinations, metric calculations and interval assignments are done in highly parallelized environment due to extensive Routing Information Base and Forwarding Information Base sizes.

The last phase is completely performed in the data

plane. The encapsulation is triggered by node selected packets. As CAMR can be implemented in parallel with existing infrastructure and services, it is upon ISP its deployment model. The CAMR routing can be applied on network, access or hybrid interface mode.

In case of access mode, only limited set of traffic by implementing Service Access Point (SAP) approach. The SAP selects incoming traffic by specific rule, selecting by:

- Incoming physical interface, e.g. port 1/1/2
- Incoming VLAN Id e.g. 1/1/2:210
- Incoming Source/Destination prefix
- Filter L2/L3/L4 combined filtering

The access mode SAP can further provide accounting and statistics to the network.

The network mode is applied on the node for enabling the source routing. The network and hybrid modes allow to interface inspect Routing Header Extension and forward source routed packets to the next destination.

When hybrid mode is applied, the node is forwarding source routing based packets. If no Extension Header found on incoming packet, the appropriate source routing encapsulation will be applied.

A. Source/Destination selection

The source and destination selection is a phase when control plane takes care about prioritizing path calculation in the destination queues for highly loaded destinations. The computation frequency is linearly proportional to the volume of destination traffic in the past interval. Thus the probability of better high loads balancing in time increases and provides the algorithm higher efficiency and overall performance.

B. Path and flow search

CAMRv2 algorithm (Fig.2) searches for all paths p, from oriented graph G(V, E) - between nodes s and t and finds free capacity c for the flow f_i in order to maximize the flow f. Every algorithm iteration, the overall residual capacity G_{fi} is decreasing, until there is no path p_i between s and t. Then f(s, t) represents the maximum flow possible found by algorithm.

The path and flow search phase is derived from Breadth-First Search/Edmonds-Karp algorithm adapted for CAMRv2 purposes. For lowering the calculation complexity $O(VE^2)$ of the original CAMR algorithm, several conditions and mechanisms were applied.

The first improvement (1) is the algorithm interruption condition, when the found paths capacity for the destination is exceeding expected load with sufficient load margin B, distributed over sufficient number of found paths M. This improvement aims to lower processing overhead for less significant traffic loads. The distribution is fine for higher loaded destinations, the very low destinations will use best-effort balancing.

while
$$((M \ge |p|) \land (B \ge f))$$
 (1)



Fig.1. CAMRv2 protocol (CAMPv2) based IP/MPLS router architecture.



Fig.2. CAMRv2 improved path set search algorithm

The secondary interruption condition is based on rule of using only shortest path set and neglecting very long paths. In the 3rd and 4th algorithm phases, stability factor exponentially suppresses load for paths with much higher hop count than the shortest path. Thus it is desirable to interrupt max-flow algorithm to avoid costly long path set

calculations if long paths would not be used due stability factor in next steps.

while
$$(|p_i| < N + |p_1|)$$
 (2)

C. Metric calculation

In the third round the metric is generated for each path in the path set. The unique approach of CAMR algorithm is the 2^{nd} round of benefiting shortest path set selection and flow distribution by distance. For this purpose the compound CAMR metric s_f reflects the proportion of data sent over a specific path:

$$\rho(p_i) = \frac{c_{fi}^{e_f}}{d(t_i)^{s_f}}, \text{ where } s_f \in <0, \infty), e_f \in <0,1) (3)$$

The metric $\rho(p_i)$ is dependent from the path capacity c_{fi} , its length $d(t_i)$ in terms of the number of hops, the network stability s_f and exponential factor e_f .

- The $d(t_i)$ represents an integer value in terms of hops between the inspected tunnel t_i .
- The c_{fi} is representing path capacity derived from the minimal link bandwidth over selected path. The bandwidth is considered as a mean value from the last time the measurement was performed. The sampling period in simulations below are 14 seconds long.

- The e_f provides continuous flow redistribution for smoother convergence to the balanced state and flow flapping avoidance between paths. The e_f out of specified range (3) causes high-instability.
- The *s_f* suppresses distribution over longer path in favor the shortest path. Selecting longer paths increases overall throughput but on the other hand load generated in network by selecting too long paths may inefficiently congest the network.

D. Metric to interval calculation

The metric has only local significance and it is used as proportional value to calculate intervals of hash-function. The path p_i proportion of forwarded data is represented by the α_i - width of interval H_i . The interval H_i belongs to specific path p_i and it is dependent from $\rho(p_i)$ metric proportion to overall metric for destination *t*:

if
$$\delta_i = \frac{\rho(p_i)}{\sum_i \rho(p_i)}, 0$$
 then $\alpha_1 = \left\lfloor \frac{\delta_i}{2^m} \right\rfloor$, where $m = 16$
(5)

$$H_1 < 1, \delta_i$$
), for $i > 1$: $H_i(\delta_{i-1}, \delta_{i-1} + \delta_i)$ (6)

E. IPv6 Encapsulation

In order to achieve smooth integration of CAMR algorithm into existing packet based environment, the algorithm and protocol shall find its application in IETF and IEEE standard based networks. CAMR based routing is intended to provide IPv4, IPv6 and MPLS implementations from the first day, using source routing approach. The aim is to install the routing header information with waypoints into standardized IPv4 Strict or Loose Source Routing header option, IPv6 Routing Header, Routing Extension Header [6] and Segment Routing Label Stack encapsulation.

The most perspective implementation for to Deutsche Telekom Group TeraStream architecture [7] is applying IETF IPv6 approach. Firstly, TeraStream network is based purely on IPv6 standard. Secondly, there is no limit on Routing Header length. In case of fragmentation restriction the 1500B MTU allows up to 90 waypoints. The IPv6 packet header is 40 bytes long. The header contains Next Header field, identifying next header type. For CAMRv2 needs the value of Next Header is 43 identifying IPv6 Routing Header.

Routing Header format following fields: Next Header, Routing Header size in octets, Routing Header type, unvisited node index, Reserved and Waypoint addresses list. The Routing Header for Source routing is 0. The next waypoint in the waypoint address list is indexed by unvisited node index decremented when packet reaches the waypoint. The Waypoint address list contains list of 128-bit IPv6 addresses used for CAMR source based routing paths.

Additionally, there are other events and parameters extending the algorithm providing framework for practical CAMRv2 implementation: Congestion Aware Multipath Protocol (CAMP) and Congestion Feedback Protocol (CFP). The CAMP and CFP description is out of the scope of this paper.

IV. CAMRV2 ARCHITECTURE PRINCIPLES

As the control plane is designed separately from the control plane, CAMRv2 can be implemented in three different variants: centralized, distributed, local. centralized implementation, Considering the the congestion feedback and network status are collected in a single centralized node and subsequently computed routing table are distributed into nodes. Such option is suitable for small-sized networks with sparse connectivity. In distributed implementation, several geographically spread units compute routing tables for local router clusters. It is possible to reuse local datacenters for providing such computation service. Centralized and decentralized architecture options are based on Software Defined Network (SDN) principles [8] by managing local routers network decisions by a remote controller.

V. STABILITY FACTOR

Multipath routing over suboptimal paths results in burning unused network resources in the network to provide better load-balancing and higher end-to-end capacity for flows that need it at the moment.

Thus, the total volume of traffic in the network in time



Fig.3. CAMRv2 IPv6 Routing Header Type=0 implementation

the shortest path routing. The stability and suppression factor s_f , mentioned in equation (3), is another key differentiator of CAMR from other multipath approaches. In order to avoid selfish routing and suppress long paths in the Path Set, dynamically changing s_f exponentially limits the long path traffic. The s_f value is dependent on the network overall load and stability. The appropriate value determination is done in control plane by analyzing the route flapping behavior. The s_f possible range is from interval $<0, \infty$), however by experiments it was proved that only values in the interval from 0 to 10 are relevant. Higher values than 10 provide same results, but need much higher computation and memory resources to perform.

The stability factor could be also perceived as parameter with limiting the network in between two extremes: maximum flow efficiency on one hand or equal cost multipath path load-balancing on the other.

VI. ALGORITHM STABILITY SCENARIOS

There are several common characteristics of converged network architecture regardless of ISP and the technology used. The ISP network is lavered into 3 or 4 main domains. Firstly, the Access network provides direct connectivity to Business customers and Residential and Mobile subscriber access infrastructure. Secondly, Aggregation domain (R1 nodes) collects all the traffic from Access nodes and each node is terminated redundantly by higher-capacity links to 2 independent Core routers creating Ring or DWDM Horseshoe Architecture[10]. The Core network (R2 nodes) due to resiliency create full-mesh topology to provide nonstop forwarding via geographically redundant paths in case of any single element failure at this level. In order to provide SecGW, SGW, IPTV, Voice, multimedia and other CDN based services, core nodes are directly connected to local datacenters. Tests performed in this paper were simulated on TeraStream 21 node network consisting of 6 R2 nodes and 15 R1 nodes.

In this paper, the CAMR algorithm stability and link load distribution are analyzed. The algorithm was tested by 3 basic scenarios in the aggregation network node to verify algorithm performance:

- Node boot-up and routing traffic from 0 into 66% link load
- Step increased traffic load on active node
- Continuously increased traffic load on active node

A. Scenario 1.: Node boot-up

This scenario is simulating standard case in an ISP network. There are several options representing the standard situation:

- New node installation
- Non-redundant control plane node restart

- Software upgrade without In Service Software Upgrade feature(ISSU)
- Node restart after power outage



Fig.4. Network under test - pure IPv6 TeraStream concept

It shall be verified that a new node introduced into the topology can handle the initial traffic load. Additionally, it shall be verified, that network discovering a new node can maintain stability.

The R1 node boots up and starts to route traffic over first time discovered paths (Fig. 5.). The sent traffic significantly influences link load on discovered paths in the boot-up phase. The consequence of the high step load are path selection and load distribution flapping. However flapping occurs during initial transmission period 0-60 steps, it keeps around optimal values.



Fig.5. Scenario 1.: Algorithm and Link stabilization period from 0 to 66% load, steps (0,60)

After the initial period, the algorithm converges and the step flapping from initial transmission period stabilizes into lower interval (Fig. 6.).

This scenario have proved the algorithm stability and convergence in case of new node boot-up or reset from idle into active state. Despite flapping the scenario also proved better performance in link load balancing being closer to optimal values than Shortest Path First (SPF) and SPF with Equal Cost Multipath (ECMP) routing.



Fig.6. Scenario 1.: Algorithm and Link stabilization period from 0 to 66% load, steps (150,220)

B. Scenario 2.: Step increased load

This scenario represents another typical cases in ISP environment, where the Aggregation node shall maintain forwarding:

- New access node connected into the aggregation node
- Access node restarted
- Access link restored after a failure in the node
- New incoming traffic from redundant dualhomed access link, due to the primary link failure

The scenario is considering stable environment with randomly generated traffic. In the step 302, the traffic from the aggregation node is increased by 20 Gb/s, meaning 58% higher load from the stable state. The network response after the step load increase was slightly unstable, but not leaving the stable margin variation – 2%, compared to 8,2% ECMP variation. This scenario again have proved better performance in step load balancing than SPF and SPF with ECMP routing (Fig. 7).



Fig.7. Scenario 2.: Algorithm and Link stabilization, 58% step load increase

C. Scenario 3.: Continuosly increased load

The last scenario (Fig. 8.) proves CAMRv2 high stability and load-balancing in stable environment and also in the environment where the traffic is continuously

increased by time.

Traditional approach of SPF and SPF with ECMP routing performance suffers from equal-cost paths absence for all destinations in the routing table. Meanwhile traditional SPF is always selecting only one link for destination, SPF-ECMP is selecting 1 up to 2 links, depending on destination location whether an another equal shortest path exist or not. Therefore, no real load-balancing is occurring in terms of efficient network resource usage.

The SPF routing algorithm have chosen for the most of destinations the link R1toR2b as heavily loaded destinations share the same R2 router in aggregation network. ECMP have distributed the load that maintain 2 shortest paths. CAMR does the loadbalancing regardless the small difference in hop-count metric.



Fig.8. Scenario 3.: Algorithm and Link stabilization, step and continuous traffic load increase

VII. CONCLUSIONS

Due to higher bandwidth requirements and optimal network resource utilization, new routing algorithms have to be implemented.

In this paper the improved CAMRv2 algorithm was presented for multipath congestion aware routing, improving the load-balancing performance compared to widely used OSPFv3 routing and ECMP approaches.

Additionally, the computation time of Edmonds-Karp was reduced by interrupting BFS algorithm computing unnecessary paths into the path set. This allows to integrate CAMRv2 into less powerful devices and decreases CAMRv2 convergence time.

Secondly, new architecture implementation options were described to support SDN based networks. This will benefit in flow manipulation and steering for future needs as planned outage detour or node graceful restart mechanism.

Thirdly, detailed implementation of CAMRv2 into IPv6 environment was presented applying source-based routing. The new method was analyzed by basic procedures, where the algorithm proved to be stable upon device boot, step and continuous traffic increase. In all scenarios, CAMRv2 algorithm proved better loadbalancing performance than traditional network routing approaches with OSPFv3 and ECMP implemented in many ISP networks. Simulations therefore provided clear advantage in application of the new method in aggregation networks.

VIII. FUTURE WORK

CAMRv2 in this paper was analyzed in sparse uplink network in the 21 node topology. CAMRv2 shall be analyzed in detail in Core network applications to prove benefits in higher throughput. The higher throughput shall be achieved by finding unused network resources, as the network is not loaded constantly equal in the time. Higher throughput over dense network could find applications especially in new datacenter migration scenario, where huge amount of data shall be transferred between two locations, exceeding the overall capacity of free shortest paths in the network.

An extensive set of Dynamic Source Routing (DSR) have been recently tested by experimental performance analysis [11]. To provide qualified benchmarking of proposed algorithm, multi-metric and constraint network used in the work shall be applied. The same applies for on comparison of proposed algorithm with on Demand Routing Protocol benchmarking [12].

Source based routing opens discussion for insufficient security in terms of DoS attacks. CAMRv2 is currently based on unsecure source routing principles. Attacker, accessing source routing enabled network, can with lowbandwidth IPv6 interface generate up to 255 times higher load. The attack is based on generating packets with maximum possible waypoints pointing to addresses ABABAB thus packets are incremented into the long cyclic path and congesting links bidirectionally. The load is cumulatively rising until TTL or Index of waypoints in the first packet are 0. Due to this fact, IETF does not recommend source based routing by default and provides explanation of IPv4 [13] and IPv6 [14] source based routing deprecation from original standard. As all of nodes can be authorized remotely from centralized server, all source based routing network can be secured via MACsec Security Encryption [15]. To ensure CAMRv2 security, L2 layer security from potential attackers shall be further investigated.

ACKNOWLEDGMENT

This article was created with the support of the Ministry of Education, Science, Research and Sport of the Slovak Republic within the Research and Development Operational Programme for the projects "Centre of Excellence for SMART Technologies, System and Services II", ITMS 26240120029 and "University Science Park of STU Bratislava", ITMS 26240220084, co-funded by the European Regional Development Fund. References.

REFERENCES

- R. Krishnan, et al., "Mechanisms for Optimal LAG/ECMP Component Link Utilization in Networks, "*IETF draft-krishnan-opsawg-large-flow-load-balancing*", February 2014.
- [2] M. Mirahsan, R. Schoenen, H. Yanikomeroglu, "HetHetNets: Heterogeneous Traffic Distribution in Heterogeneous Wireless Cellular Networks," *IEEE Journal* on Selected Areas in Communications, vol. PP Issue 99, May 2015.
- [3] I. Gojmerac, "Adaptive multipath routing for dynamic traffic engineering" Dissertation thesis, Technischen Universitat Wien Fakultat f ür Elektrotechnik und Informationstechnik, April 2007.
- [4] A. Al-Shabibi, B. Martin "MultiRoute a congestionaware multipath routing protocol" *International Conference on High Performance Switching and Routing* (HPSR), June 2010, p. 88-93, ISBN: 978-1-4244-6969-7.
- [5] A.R. Mahlous, R.J. Fretwell, B. Chaourar, "MFMP: Max Flow Multipath Routing Algorithm" *In Second UKSIM European Symposium on Computer Modeling and Simulation, EMS* '08., September 2008. p. 482 – 487, ISBN: 978-0-7695-3325-4.
- [6] G. Huston, The ISP Column: An occasional column on things Internet 2007 "*Internet Society*", vol.2, June 2007.
- [7] P. Lothberg, TeraStream A Simplified IP Network Service Delivery Model, "*RIPE 67 Conference on IPv6*" October 2013.
- [8] M.R. Celenlioglu, H.A. Mantar, An SDN Based Intra-Domain Routing and Resource Management *IEEE International Conference on Cloud Engineering (IC2E)*, March 2015, pp 347 – 352.
- [9] E. Miranda, V. Cacchiani, T. Dorneth, et. al. Models and Algorithms for Robust Network Design with Several Traffic Scenarios Combinatorial Optimization *Lecture Notes in Computer Science*, Vol. 7422, April 2012, pp 261-272.
- [10] M. Gunkel, F. Wissel, W. Weiershausen, M. Franzke, V. Fürst, A. Mattheus, "Multi-layer interworking with rateadaptive transmission technology - benefit and challenges of a new use case", *In Proceedings of Photonic Networks*; *16. ITG Symposium*, May 2015, ISBN: 978-3-8007-3938-7.
- [11] Z. Mehmood, M. Iqgbal, W. Xingheng "Comprehensive Experimental Performance Analysis of DSR, AODV and DSDV Routing Protocol for Different Metrics Values with Predefined Constraints", *In International Journal of Information Technology and Computer Science(IJITCS)* Vol. 6, No. 7, June 2014.
- [12] B. Bushhan, S. Gupta, C.K. Nagpal "Comparision of on Demand Routing Protocols", *I.J. Information Technology* and Computer Science, March 2013, Vol.3., p. 61-68, DOI: 10.5815/ijitcs.2013.03.08.
- [13] A. Reitzel "Deprecation of Source Routing Options in IPv4, draft-reitzel-ipv4-source-routing-is-evil-00", [Online] https://tools.ietf.org/html/draft-reitzel-ipv4-source-routingis-evil-00.
- [14] J. Abley, P. Savola, G. Neville-Neil, "Deprecation of Type 0 Routing Headers in IPv6" *In IETF RFC 5095*, December 2007, [Online] https://tools.ietf.org/html/rfc5095
- [15] IEEE Standard, "802.1AE-2006 IEEE Standard for Local

15

and Metropolitan Area Networks: Media Access Control (MAC) Security'', August 2006, https://standards.ieee.org/findstds/standard/802.1AE-2006.html.

Authors' Profiles



Matej Kultan: was born in 1989. He received the B.E. and M.E. degrees in Faculty of Electrical Engineering and Information Technology of Slovak University of Technology (FEI STU) Bratislava in 2007 and 2009, respectively. At present he is a PhD student at the Institute of Telecommunications of FEI

STU Bratislava and IP Technical Pre-Sales Engineer at Alcatel-Lucent Slovakia.



Martin Medveck ý: was born in 1968. He received the Master degree and PhD degree in the Faculty of Electrical Engineering and Information Technology, Slovak University of Technology in Bratislava in 1991 and 1998, respectively. He is currently an Associate professor at the Institute of Telecommunications of FEI

STU Bratislava.

How to cite this paper: Matej Kultan, Martin Medveck ý,"Congestion Aware Multipath Routing: Aggregation Network Applicability and IPv6 Implementation", International Journal of Information Technology and Computer Science(IJITCS), vol.7, no.11, pp.9-16, 2015. DOI: 10.5815/ijitcs.2015.11.02