

Network Working Group  
Internet-Draft  
Intended status: Informational  
Expires: May 28, 2016

S. Yang  
H. Yu  
UESTC  
X. Zhang  
N. Wu  
Huawei

November 25, 2015

## Ordered Metric Adjustment Implementation Report draft-yang-rtgwg-oma-impl-report-00

### Abstract

Ordered Metric Adjustment (OMA) as specified in [\[I-D.zxd-rtgwg-ordered-metric-adjustment\]](#) has provided a mechanism whereby global transient forwarding loops can be avoided through graceful link metric adjustment. This document provides an implementation report for OMA algorithm and mechanism on different network topologies. What's more, it gives an analysis on efficiency and performance of OMA, comparing with other well-known loop-preventing algorithm.

### Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC 2119](#) [[RFC2119](#)].

### Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of [BCP 78](#) and [BCP 79](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on May 28, 2016.

Internet-Draft

OMA Implementation Report

November 2015

## Copyright Notice

Copyright (c) 2015 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](#) and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

## Table of Contents

<a href="#">1.</a>	Introduction . . . . .	<a href="#">2</a>
<a href="#">2.</a>	Terminology . . . . .	<a href="#">3</a>
<a href="#">3.</a>	Overview of Algorithm . . . . .	<a href="#">3</a>
3.1.	Calculating the adjustment range of link metric for each node . . . . .	<a href="#">4</a>
3.2.	Merge different Seq(Yj) into a global sequence . . . . .	<a href="#">5</a>
<a href="#">4.</a>	Innovation point of algorithm . . . . .	<a href="#">5</a>
<a href="#">5.</a>	Simulation of Algorithm . . . . .	<a href="#">6</a>
5.1.	Environment Setup . . . . .	<a href="#">6</a>
5.2.	Measurement and Methods . . . . .	<a href="#">7</a>
5.3.	Simulation results . . . . .	<a href="#">8</a>
5.3.1.	Micro-loop risk . . . . .	<a href="#">8</a>
5.3.2.	Sequence length and computing time . . . . .	<a href="#">8</a>
<a href="#">6.</a>	Conclusion . . . . .	<a href="#">10</a>
<a href="#">7.</a>	IANA Considerations . . . . .	<a href="#">11</a>
<a href="#">8.</a>	Security Considerations . . . . .	<a href="#">11</a>
<a href="#">9.</a>	Acknowledgments . . . . .	<a href="#">11</a>
<a href="#">10.</a>	References . . . . .	<a href="#">11</a>
10.1.	Normative References . . . . .	<a href="#">11</a>
10.2.	Informative References . . . . .	<a href="#">11</a>
	Authors' Addresses . . . . .	<a href="#">12</a>

## [1.](#) Introduction

This document provides an implementation report for OMA algorithm and mechanism on different network topologies. What's more, it gives an

analysis on efficiency and performance of OMA, comparing with other well-known loop-preventing algorithm.

## [2.](#) Terminology

- o Reverse Shortest Path First Tree (RSPF Tree): A directed acyclic graph containing all the shortest path from the node of the network toward the root.
- o Link L: An edge in the network connecting node a to b will shut down and its metric value is K when link L is up.
- o X set: A node set contains the node impacted by the removal of a link L such that each node in X set must change its forwarding paths to avoid link L.
- o Y set: A node set contains all destination nodes in that may be concerned by the down event of the link L.
- o  $D(Y_j, X_i)$ : The shortest distance from node  $X_i$  to  $Y_j$  when link L is up.
- o  $D'(Y_j, X_i)$ : The shortest distance from node  $X_i$  to  $Y_j$  when link L is down.
- o  $COST(Y_j, X_i, max)$ : If node  $X_i$  switches its optimal path to  $Y_j$ , the reasonable upper metric of link L can be adjusted is  $COST(Y_j, X_i, max), COST(Y_j, X_i, max) = MIN(COST(Y_j, X_k, min))$ , where  $X_k$  is the father of node  $X_i$  in case of link L being up.
- o  $COST(Y_j, X_i, min)$ : If node  $X_i$  switches its optimal path to  $Y_j$ , the reasonable lower metric of link L can be adjusted is  $COST(Y_j, X_i, min), COST(Y_j, X_i, min) = D'(Y_j, X_i) - D(Y_j, X_i) + K$ .
- o  $R = (R_{min}, R_{max})$ : Here  $R_{min}$  equals  $COST(Y_j, X_i, min)$  and  $R_{max} = COST(Y_j, X_i, max)$ . When the link L's metric is set in the range of  $(R_{min}, R_{max})$ , node  $X_i$  can be switched to its optimal path to  $Y_j$  without forwarding loop.

- o Seq(Yj): A metric range sequence for Yj contains all metric ranges of source nodes Xi affected by the removal of a link L. When link L's metric set to a value in each metric range of sequence Seq(Yj) in order, each node Xi will to its optimal path to Yj without forwarding loop.

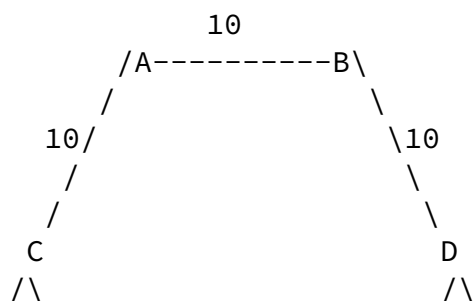
### 3. Overview of Algorithm

The Internet is the most popular network, which is a distributed system. Depend on its configuration, each network device communicates with its neighbor, calculate routes and generate the FIB individually, finally, the packet will be forwarded hop by hop. But

due to the difference of each device hardware capabilities and internal/external environments, the route calculation cannot be scheduled at same time, then micro-loop occur.

In draft [[I-D.zxd-rtgwg-ordered-metric-adjustment](#)] we introduces a method to prevent forwarding loop by adjusting link metric gradually for several times. In this section we will give an overview of algorithm and here we just talk about link down event and results computed in down event are also appropriate to link up event. We suppose link L connecting node A to B referenced in the following sections will shut down.

First of all, we have some definitions. Set X consist of Xi which on sub-tree below A in RSPF tree with root B such that each node in X must change its forwarding paths to Yj to avoid link L. And the notation Y that gives the set of "affected destinations". This set contains all destination nodes that may be concerned by the change of the link L and set Y consist of Yj which on sub- tree below B in RSPF tree with root A. In Figure 1, we have X = {A, C, F, G, H} and Y= {B, D, E}



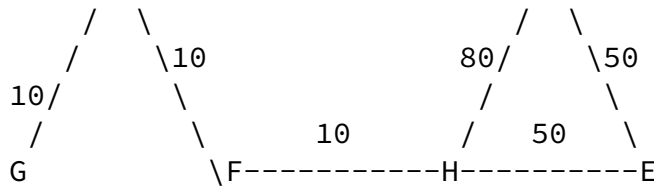


Figure 1 Topology

### 3.1. Calculating the adjustment range of link metric for each node

This part is used to calculate a reasonable adjustment metric range for each node  $X_i$  to switch its optimal path to a destination  $Y_j$  without forwarding loop. Each metric range of  $X_i$  related to  $Y_j$  make up of a metric range sequence for  $Y_j$ .

For destination  $Y_j$ , we calculate the RSPF tree with root  $Y_j$  when link  $L$  is up and down to determine the values of two variables  $R_{min} = \text{COST}(Y_j, X_i, \min)$  and  $R_{max} = \text{COST}(Y_j, X_i, \max)$ . When the link  $L$ 's

metric is set in the range of  $(R_{min}, R_{max})$ , node  $X_i$  can switch to its optimal path to the root node  $Y_j$  without forwarding loop.

Then metric ranges of  $X_i$  computed above make up of a metric range sequence for  $Y_j$  recorded as  $\text{Seq}(Y_j)$ . Then the metric of link  $L$  can be set to a value in each range of sequence in order. For example, in figure 1, we have  $\text{seq}(B) = \{[100, 120], [80, 100], [60, 80]\}$  and if we set the metric of link  $L$  to 60, 81 and 101, each node  $X_i$  will switch to its optimal path to the  $B$  without forwarding loop.

### 3.2. Merge different $\text{Seq}(Y_j)$ into a global sequence

According to [section 1.1](#), we have different metric range sequence corresponding to each destination  $Y_j$ . However, nodes need to react to the change of a link weight for all affected destinations. This part will show how to merge different sequences of each destination into a minimal sequence effectively, then multiple nodes in set  $X$  simultaneously switch optimal path to each root node without forwarding loop.

Compare the metric range of each  $\text{Seq}(Y_j)$  and choose range  $R = (R_{min}, R_{max})$  with largest lower bound in these metric ranges. Then

remove the metric range with upper bound larger than the lower bound  $R_{\min}$  of  $R$  in all sequence. Repeat operation above when the sequence of all destination is empty, then algorithm terminates.

When the algorithm finishes, remaining metric ranges  $(R_{1\min}, R_{1\max}), (R_{2\min}, R_{2\max}), \dots, (R_{n\min}, R_{n\max})$  are what we want. In case of link  $L$  going to up, the adjustment process of metric of link  $L$  can be set to  $R_{1\min}+1, R_{2\min}+1, \dots, R_{n\min}$ , and configuration value  $K$ . In case of link  $L$  going to down, the adjustment process of metric of link  $L$  can be set to configuration value  $K, R_{n\min}+1, R_{((n-1)\min)}+1, \dots$ , and  $R_{1\min}$ .

In figure 1, In case of link  $L$  going to down, the adjustment process of metric of link  $L$  can be set to  $\{10, 41, 61, 81, 101\}$ , in proper order, then there is no forwarding loop during this adjustment.

#### 4. Innovation point of algorithm

There are several different algorithms solving forwarding loops by adjusting link metric gradually. Compared with other solution, we put forward two innovation points in each part of our algorithm.

First innovation point of our algorithm is to check whether destination  $Y_j$  is affected actually due to the down link. In part 1 of algorithm, if the size of set  $Y$  is  $|Y|$ , we need to compute  $2|Y|$  times RSPF tree and its complexity is in  $|N|\log|N|+|E|$  per times in

the best case. We just compute an RSPF tree when link  $L$  is up firstly. According to the Dynamic Dijkstra Algorithm, we know that nodes only on sub-tree below  $A$  in RSPF tree with root  $Y_j$  will change its optimal paths to destination if link  $L$  change. Obviously, if it is empty on sub-tree below  $A$  in RSPF tree with root  $Y_j$ , only node  $A$  will change its optimal path to root so that no forwarding loop for destination  $Y_j$  can occur when link  $L$  down. We call it "Unaffected destination" and in this case we can delete destination  $Y_j$  from the set  $Y$ . If there are  $|m|$  destinations are unaffected, we just need  $|2|Y|-|m||$  times computation.

The second innovation of our algorithm is that using metric range express the reasonable adjustment metric range for each node to change their optimal path to the root node and using a common range instead of several overlapping ranges to reduce the operation. In

part 2 of algorithm, we merge different range sequences into a global sequence, however, many ranges overlap with each other so if merge overlapping ranges into a common range, adjusting link metric several times can be instead of one time. Like this we can reduce a large number of metric ranges in a reasonably short time and the final global sequence will include a few ranges.

Through these two points, we have a minimal metric sequence length with the fastest computing time among the similar algorithms.

5. Simulation of Algorithm

This section aims at evaluating the performance of our algorithm from two aspects: sequence length and the time required to compute them on real IP networks. Firstly, we present the topologies we used in our algorithms. Then present the results of our simulation, analyzing the reasonable length, as well as the efficiency of our optimizations on computing times.

5.1. Environment Setup

Table I summarizes the main characteristics of our simulation topologies. |N| and |E| respectively represent the number of nodes and links in the graph. d<sub>m</sub> is the maximum degree, and ρ gives the weight distribution. Networks 1-6 of Table I are Rocketfuel inferred topologies obtained with traceroute campaigns [[inferring-link-weight-e2e](#)]. Networks 7-10 are generated obey power-law distribution. Networks 9 and 10 are topologies with special shape. Network 9 is a star topology and network 10 is a circle consists of several small circles. Networks 4 and 5 come with inferred IGP weights and other topologies come with random weights in [1,100].

+-----+-----+-----+-----+-----+-----+-----+
ID   Network    N     E    dm   w
+-----+-----+-----+-----+-----+-----+-----+
1   Abovenet  17   74   8   [1,100]
+-----+-----+-----+-----+-----+-----+-----+
2   Exodus   21   72   7   [1,100]
+-----+-----+-----+-----+-----+-----+-----+
3   Tiscali   27   128   18   [1,100]

4	Sprint	30	138	11	[1,100]
5	Geant	23	74	6	[156,9953]
6	AT&T	154	376	29	[1,222]
7	Pow-1	50	200	7	[1,100]
8	Pow-2	100	400	7	[1,100]
9	Pow-3	200	800	7	[1,100]
10	Pow-4	500	2000	7	[1,100]
11	star	13	44	6	[5,101]
12	circle	41	90	3	[8,99]

Table 1 Main graph properties of simulation network

## 5.2. Measurement and Methods

In the results of simulation we focus on the length of metric sequence and computing time. The shortest sequence length, the less operation will we do. Thus we expect to get a minimal sequence and computing time is as short as possible.

Firstly we have a loop-risk test on all simulation networks. We carry on 1000 times tests on each network making a link fail randomly, recording the total number that we need to use our algorithm to reconfigure fail link. This test will show that in most topologies the risk of loop will occur frequently when a link fail. So it is very important to search an efficient algorithm to prevent micro-loop.

Then we carry on  $|E|$  times tests in each network mentioned above ensuring that every link in topology will fail once, recording the sequence length and computing time each time and analysis it.

## 5.3. Simulation results

### 5.3.1. Micro-loop risk

Table 2 shows that possibility of micro-loop occurrence is more 50% in most topology. So it is very important to search an efficient algorithm to prevent forwarding loop.

Topo-ID	1	2	3	4	5	6
Loop risk	52.1%	66.8%	66.8%	64.1%	73.0%	29.1%

Topo-ID	7	8	9	10	11	12
Loop risk	77.6%	89.1%	91.6%	96.1%	56.3%	100.0%

Table 2 Micro-loop risk percentage

### 5.3.2. Sequence length and computing time

Tables 2 and 3 respectively give an overview of metric sequence length and computing time observed on our set of simulation networks.

**Minimal Metric Sequence Length:** The length of the sequence we obtained in most cases are relatively short. The actual network topology of 1-4,6 and 11, for more than half of the cases, no intermediate metric is needed, as the size of the sequence is equal to 0 and more than 90% of the cases, less than two intermediate metrics are needed to ensure a loop-free convergence in case of a link removal. For all networks, in most case approximately less than ten intermediate metrics can solve micro-loop during the convergence.

For all networks, there are several factors which influence the length of the metric sequence: 1) network size: Apparently when the size becomes larger, the number of source and destination nodes significantly increased, so more intermediate state are needed. In our simulation, network 10 is the biggest which has 500 nodes and 2000 links and we can notice that it need more intermediate state than others. 2) network shape: a ring network is easy to form a deep , so there can be many nodes on the ring forward its packets along the same direction to destination. When an upstream link is closed, if all these nodes go backward to avoid a failed link, it will produce a number of cycles, as the length of the sequence is longer than others certainly. Results of network 12 prove it. 3) weight distribution: When a network with a large weight range, there may be

less overlap range. So the common range remained will be more thus we get a longer sequence such as network 5.

Although in a few cases the length of metric sequence is a little long, but most of our sequence have a very reasonable size.

Computing Time: We will show that compared with others, our algorithm has the shortest computing time. Here we focus on a similar algorithm [[lightweight-algorithm-minimal-operation-impact](#)] (We call it to "Graceful" for short in the following content) and give a contrast between them. Table III gives computing time percentiles including worst-cases results.

According to the statistical, for all networks except network 10, the maximum computation time for calculating a sequence is below a few hundred milliseconds. Compared with the Graceful algorithm, there is no significant difference when network size is small. However, with network size increasing our algorithm performs better significantly which can be completed in a shorter time. That is because two innovative points in our algorithm. It can reduce the computation time effectively.

NID	Length distribution (%)								MAX
	MMS =0	<=1	<=2	<=5	<=8	<=10	<=15	MMS	
1	69.64%	89.29%	98.21%	100.00%	100.00%	100.00%	100.00%	3	
2	50.00%	80.65%	90.32%	100.00%	100.00%	100.00%	100.00%	5	
3	55.41%	83.78%	91.89%	100.00%	100.00%	100.00%	100.00%	4	
4	54.72%	83.02%	96.23%	100.00%	100.00%	100.00%	100.00%	4	
5	21.88%	29.69%	59.38%	82.81%	93.75%	87.50%	100.00%	15	
6	77.69%	96.15%	98.46%	100.00%	100.00%	100.00%	100.00%	4	
7	32.76%	60.92%	79.31%	94.25%	97.70%	95.98%	100.00%	14	
8	25.26%	49.21%	63.16%	84.74%	91.58%	87.63%	98.68%	21	
9	11.15%	23.23%	38.32%	68.50%	84.65%	75.72%	98.56%	34	
10	4.38%	10.56%	18.28%	44.85%	65.35%	73.79%	89.39%	87	

11	53.33%	93.33%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	2
----	--------	--------	---------	---------	---------	---------	---------	---------	---

12	0.00%	0.00%	1.11%	57.78%	84.44%	62.22%	94.44%	22
----	-------	-------	-------	--------	--------	--------	--------	----

Table 3 Length Distribution

NID	Computing Time Distribution (ms)							
	50th		80th		100th		Mean	
	OMA	Graceful	OMA	Graceful	OMA	Graceful	OMA	Graceful
1	1	1	1	1	2	5	0.64	0.96
2	1	1	4	3	4	8	1.56	1.82
3	1	1	1	4	7	8	1.22	2.23
4	1	1	3	4	7	10	1.79	2.33
5	2	3	3	6	20	15	2.44	3.41
6	6	13	49	175	189	294	35.68	69.11
7	3	4	8	18	24	52	4.76	9.32
8	12	21	37	86	113	351	20.95	49.50
9	71	172	192	650	589	2541	111.55	346.83
10	907	2467	2656	6886	7365	24206	1445.28	3689.79
11	1	0	1	1	2	2	0.80	0.47
12	6	18	9	31	14	131	6.17	26.27

Table 4 Computing Time Distribution

## 6. Conclusion

In document [[I-D.zxd-rtgwg-ordered-metric-adjustment](#)] we introduces a method to prevent forwarding loop by adjusting link metric gradually for several times. In this document we mainly give the simulation data. We have a simple review firstly then focus on the innovation in our algorithm. Then give the simulation results such that our algorithm have the best performance among the current mainstream algorithms to prevent forwarding loop.

Yang, et al.

Expires May 28, 2016

[Page 10]

---

Internet-Draft

OMA Implementation Report

November 2015

## [7.](#) IANA Considerations

This document includes no request to IANA.

## [8.](#) Security Considerations

This document is not currently believed to introduce new security concerns.

## [9.](#) Acknowledgments

TBD.

## [10.](#) References

### [10.1.](#) Normative References

[I-D.zxd-rtgwg-ordered-metric-adjustment]

Zhang, X. and G. Yan, "Algorithm for Ordered Metric Adjustment", [draft-zxd-rtgwg-ordered-metric-adjustment-00](#) (work in progress), October 2013.

[inferring-link-weight-e2e]

Mahajan, R., Spring, N., Wetherall, D., and T. Anderson, "Inferring link weights using end-to-end measurements", November 2002.

[lightweight-algorithm-minimal-operation-impact]

Clad, F., Merindol, P., Pansiot, J., Francois, P., and O. Bonaventure, "Graceful Convergence in Link-State IP Networks: A Lightweight Algorithm Ensuring Minimal Operational Impact, IEEE 1063-6692", February 2014.

## 10.2. Informative References

- [RFC1195] Callon, R., "Use of OSI IS-IS for routing in TCP/IP and dual environments", [RFC 1195](#), DOI 10.17487/RFC1195, December 1990, <<http://www.rfc-editor.org/info/rfc1195>>.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), DOI 10.17487/RFC2119, March 1997, <<http://www.rfc-editor.org/info/rfc2119>>.
- [RFC2328] Moy, J., "OSPF Version 2", STD 54, [RFC 2328](#), DOI 10.17487/RFC2328, April 1998, <<http://www.rfc-editor.org/info/rfc2328>>.

Yang, et al.

Expires May 28, 2016

[Page 11]

---

Internet-Draft

OMA Implementation Report

November 2015

- [RFC5715] Shand, M. and S. Bryant, "A Framework for Loop-Free Convergence", [RFC 5715](#), DOI 10.17487/RFC5715, January 2010, <<http://www.rfc-editor.org/info/rfc5715>>.
- [RFC6976] Shand, M., Bryant, S., Previdi, S., Filsfils, C., Francois, P., and O. Bonaventure, "Framework for Loop-Free Convergence Using the Ordered Forwarding Information Base (oFIB) Approach", [RFC 6976](#), DOI 10.17487/RFC6976, July 2013, <<http://www.rfc-editor.org/info/rfc6976>>.

### Authors' Addresses

Shiqi Yang  
UESTC  
No.2006, Xi Yuan Ave., Westn High-Tech Zone  
Chengdu 611731  
China

Email: yangsq90309@163.com

Hongfang Yu  
UESTC  
No.2006, Xi Yuan Ave., Westn High-Tech Zone  
Chengdu 611731

China

Email: yuhf.uestc@139.com

Xudong Zhang

Huawei

Huawei Bld., No.156 Beiqing Rd.

Beijing 100095

China

Email: zhangxudong@huawei.com

Nan Wu

Huawei

Huawei Bld., No.156 Beiqing Rd.

Beijing 100095

China

Email: eric.wu@huawei.com