© 2016 IJSRSET | Volume 2 | Issue 5 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
Themed Section: Engineering and Technology

Deployment of P-Cycle in Optical Networks: A Data Mining Approach

Dr. S. Umarani[1], Dr. T. Rama Chaithanya[2], M. Divya

[1,2] SRM University, [2]PACE Institute of Technology & Sciences, Andhra Pradesh, India

ABSTRACT

The major challenge in survivable networks is the design of resource allocation algorithms that allocate network resources efficiently while at the same time being able to quickly recover from failure by rerouting the broken connection using the reserved spare capacity. This issue is particularly more challenging in optical networks operating under the wavelength continuity constraint, where the same wavelength must be assigned on all links in the selected path. This paper review the p-cycle deployment using widely used data mining algorithms and demands are calculated based on the Floyd's algorithm.

Keywords: P-Cycle, Floyd's Algorithm, Demand Calculation, Data Mining.

I. INTRODUCTION

The major challenge in survivable networks is the design of resource allocation algorithms that allocate network resources efficiently while at the same time being able to quickly recover from failure by rerouting the broken connection using the reserved spare capacity. This issue is particularly more challenging in optical networks operating under the wavelength continuity constraint, where the same wavelength must be assigned on all links in the selected path. The prime objective of most survivable routing algorithms is to minimize the consumption of network resources and reduce the restoration time, during a failure. However, in some survivable routing schemes, the objective is a trade-offs between reducing network resources required and decreasing the restoration time after failures (Hongsik Choi et al 2004). Moreover, P-cycles can reach good resource redundancy compatible to that of conventional survivable schemes used in mesh networks. In WDM wavelength-routed optical mesh networks, P-cycle techniques can be applied to ensure survivability against span failures (fiber cuts) under static and dynamic traffic environments.

1. Related Work

According to Zhenrong et al (2004) research, initially they found the all candidate cycles and efficiency ratio of each cycles. The maximum efficiency ratio of candidate cycle is selected for protection. Suppose, one or more candidate cycles have same efficiency ratio, select the candidate cycle randomly. But the design is based on joint formulation.

According to Grover (2002), this method simply rank the set of all distinct cycles by either topological Score and Apriori efficiency measures, and use only a limited number of the top ranked candidates for representation in the optimal solution model.

A novel heuristic algorithm (Taifei Zhao et al 2006), called Local-map Cycles mining algorithm for finding simple node-encircling and link candidate P-cycle. This algorithm produced good efficiency P-cycle in optical mesh networks without enumerating all cycles. However, this algorithm may sometimes use more backup capacity than it is actually needed and as the result the redundancy ratio is increased.

The new metric is called Route Sensitive Efficiency (RSE) is developed by Abdelhamid E.Eshoul & Hussein T.Mouftah (2009). This is used to reduce the number of candidate cycles and select high merit P-cycles, which is calculated, based on the locations of the primary routes. It is mainly reduced the number of candidate cycles without compromising the optimality of the solution. The following discussion is based on different mining patterns.

Kosala & Blockeel (2000), Joshi (2001), Chan et al (2002) and Ahmed et al (2011),are some of the related work taken as literature. It gives overall idea about for mining and mining pattern. Baglioni et al (2003) and Borges & Levene(2000) are mainly focused on data preposing and mining.

2. Problem Definition

In general, network traffic is unlikely to be symmetric in both directions between two nodes. This means that the number of working and protection wavelengths is not likely to be the same in both directions. According to Schupke et al (2002), the demands are routed using the shortest path algorithm and spare capacity of the network is obtained by using BFS (Breath First Search). The longer P-cycles have better efficiency. The connection of P-cycles is very dependent on the size of the network. P-cycles can be deployed in VWP and WP WDM networks. The capacity efficiency ratio of VWP is lower than 60% and WP is lower than 73% are achieved. A P-cycle can protect one working unit in the opposite direction for every on-cycle span, and two working units (one in each direction) for every straddling span. The number of spare units of a P-cycle is equal to the number of spans on the cycle. We define the ‘confidence of a P-cycle’ as the ratio of the number of working units that are actually protected by the P-cycle to the number of spare units of the P-cycle. A P-cycle with a greater confidence means that its spare units are utilized more efficiently than a P-cycle with a smaller confidence. The idea behind this, is the famous three mining algorithms are used to identify those P-cycles that can actually protect as many working units as possible, and hence to reduce the total spare units.

II. METHODS AND MATERIAL

This section describes the working principles of the proposed algorithms with an illustration, in order to determine demand of the network and form the P-cycles. Firstly, the issues identified in the existing algorithms used for deploying P-cycles are discussed.

1. Issues

The basic formulation (Grover et al 2000) generates large problem files that can be difficult to solve optimality, primarily because of the size of the set of candidate cycles to consider. This is true when the jointly optimized problem is attempted. In general, the mining cost is an important issue to be considered while designing and developing a frequent pattern mining algorithm applied for large demand of the network. Actually, this cost involves I/O and CPU costs. However, the major cost associated with the mining process is the generation of potentially candidate paths (or sequences), called candidate cycles (Maged 2003). Many of the heuristic algorithms found that the solution is very close to that of optimal solutions and generating more protection cycle.

Several algorithms have been proposed to resolve the above said problems and most of the existing heuristic algorithms discussed in the related work are based on efficiency ratio.

To avoid candidate generation at the time of mining, this chapter discusses various algorithms, efficiency improvement of the algorithms and their performance. They are found to be better than efficiency ratio algorithms.

Various Mining Algorithms

The pre-configuration cycle is a promising approach for protecting capacities in optical mesh network. The most important step in P-cycle design is selecting optimal candidate cycles. The proposed approach is used to find the candidate cycle for given networks under the wavelength continuity constraint, where the same wavelength must be assigned on all links in the selected path(s). For a more optimal solution, the problem can be solved jointly, where the candidate working routes and the candidate backup P-cycles are jointly formulated as an Apfloyds Algorithm to minimize the total capacity required. However, the number of possible P-cycles grows exponentially with the average nodal degrees and number of nodes in the network(s). Further, the number of nodes (used to represent flow due to each light-path demand) increase with increase in number of links and number of wavelengths. By this, the problem’s complexity grows exponentially, which increases the problem’s solution time to unacceptable levels.

To reduce the complexity of the problem and make it more tractable, the routing problem is first solved using path flow technique. This generates the shortest routes for each s-d pair to be used as candidate routes in the optimization model. The candidate primary routes and candidate cycles are then used to formulate the overall problem as an Apfloyds Algorithm. To reduce the number of candidate P-cycles in the formulation, a novel pre-selection algorithm has been developed to select a reduced number of high merit cycles. Consequently, the final solution is a trade-off between the optimality of the solution and the complexity of the problem.

Apfloyds Algorithm is used to identify P-cycles that can protect the maximum of working units and hence reduce the number of backup units. Considering the mesh network topology and traffic demand, this confidence-based P-cycle design algorithm is summarized as follows:

Step1: Find all possible demands according to the All-Pair shortest path Algorithm.

Step2: For each candidate cycle, calculate the confidence of its P-cycle using Apriori algorithm (Han & Kamber 2007).

Step 3: Generate association rule for candidate cycles and select a P-cycle with maximum confidence. If multiple P-cycles have the same maximum confidence, then any one is randomly selected.

The above approach needs several iterations and uses uniform threshold value. Hence an alternative method is addressed in this work by using a non-uniform minimum support thresold.

This hybrid algorithm identifies those P-cycles that can actually protect the maximum number of working units, and hence to reduce the total backup units. Initially, all the demand from the given network and the support count of each node, are determined and then the infrequent node is discarded. After this, all the demands are arranged in decreasing order based on their threshold. The hybrid algorithm then reads all demands and constructs the tree until all the demands are mapped to the tree's path. The FP method (Aiman Moyaid Said et al 2009) is then applied to extract backup path from the tree. A sub tree is processed recursively to extract the backup path, after which solutions are merged as the steps are summarized below.

Step 1: Find all possible candidate paths according to the all pair shortest path Algorithm and determine the cost.

Step 2: Generate Tree in pass 1. It scans all the demands found in Step 1 and finds the support count, and then discards the infrequent path. Arrange all the demands in decreasing order based on their support.

Step 3: Generate Tree in pass 2. It reads the all demands one by one; construct the tree until all the demands are mapped to the path in the tree.

Step 4: Frequent path generation. Apply the FP method for extracts frequent path from the tree. A sub tree is processed recursively to extract the frequent path, and then solutions are merged.

The main disadvantage of using this hybrid algorithm is that it very computationally intensive if the number of network nodes is large. It is better than the Apfloyds algorithm**,** but generates a tree structure which is considered as unfit for the memory and it is also expensive to build**.**

Another efficient mining algorithm is the PrefixSpan algorithm (Poonam Sharma1 & Gudla Balakrishna 2011), which requires '*n*' number of demand scans for a path length '*n*', even though the path is searched in the minimized search space (projected demand) by applying divide-and-conquer technique. In the PrefixSpan algorithm, a pattern Y of length l+1 is searched in the projected demand prefixed with a pattern X of length $l$at level '*k*'. Moreover, at each level '*k*', a single scan is required to test whether the path Y is frequent or not, in the projected demand prefixed with X. In the PrefixSpan algorithm, the mining cost is mainly in the construction of projected demand and multiple scans. Here, the cost of constructing projected demand can be minimized by creating pseudo-projections in memory, so that they can be processed faster.

PrefixSpan Algorithm

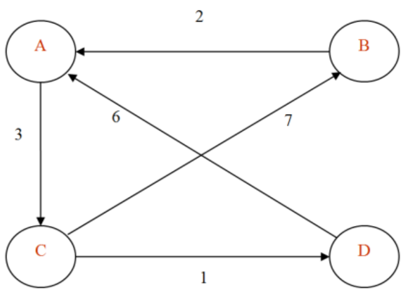
1. Scan all the demands once and find frequent 1-candidate set X

2. For each frequent 1- candidate set X
    2.1 Project all the demand prefixed with X
    2.2 Scan the demands once and find frequent 1-candidate set Y that could be merged with X and generate a candidate cycle X' as X+Y
    2.3 Output X' as a candidate cycle
    2.4 For each X' Construct X'-projected demand S|X' and go to step 2.2

III. RESULTS AND DISCUSSION

1. Method I

Consider the following Figure 1; find the source vertex (node) in the given network, the algorithm finds the path with lowest cost (i.e. the shortest path) between that node and every other node. It can also be used for finding costs of shortest paths from a single node to a single destination node by stopping the algorithm once the shortest path to the destination node has been determined. This algorithm is designed to find the least-expensive paths between all the nodes in a network. It does this by operating on a matrix representing the costs of links between nodes. Based on the matrix, List all the demands as shown in the following Table 1. The Table 1 contains 12 demands.



	A	B	C	D
A	0	0	3	0
B	2	0	0	0
C	0	7	0	1
D	6	0	0	0

	A	B	C	D
A	0	10	3	4
B	2	0	5	6
C	7	7	0	1
D	6	16	9	0

Figure 1. Mesh Network with 4 nodes (Traffic Matrix and Demand Matrix)

Table 1. Demands in the Mesh Network

S. No.	Demands	Cost
1	A->C->B	10
2	A->C	3
3	A->C->D	4
4	B->A	2
5	B->A->C	5
6	B->A->C->D	6
7	C->B->A	7
8	C->B	7
9	C->D	1
10	D->A	6
11	D->A->C->B	16
12	D->A->C	9

Consider the Table 1, consisting of 12 demands with minimum support count required is 20% and let minimum confidence required is 80%. We have to first find out the frequent path using step 2. Then, Association rules generated the paths using minimum support and minimum confidence. The set of frequent 1-Node, named as one candidate path, consists of the candidate 1- Node satisfying minimum support. In the first iteration of the algorithm, each node is a member of the set of candidate. The support count is calculated by using Table 1.

To discover the set of frequent 2-Nodes, named as two candidate path, the algorithm uses two candidate paths join one candidate path to generate a candidate path of 2-nodes, named as initial two candidate path. Next, the demands in D are scanned and the support count for each candidate node in initial candidate path is accumulated. The set of frequent 2-nodes, two candidate path is then determined, consisting of those two candidate path in initial two candidate path, having minimum support. Similarly, the three candidate path is generated. The steps are called 'join steps' and we found the confidence of each three candidate path. The highest confidence will be selected from three candidate path and these paths are called P-cycles. The entire candidate paths are shown in the following Table 2, 3, 4.

Table 2. One Candidate Paths

Candidate Path	**Support Count**
A	10
B	7
C	10
D	6

Table 3. Two Candidate Paths

Candidate Path	**Support Count**
A,B	2
A,C	7
A,D	2
B,C	2
C,D	3

Table 4. Three Candidate Paths

Candidate Path	**Support Count**
A,C,B	2
A,C,D	2

Table 5. A Confidence Ratio of P-cycles

Cycle Number	**Confidence Ratio**
A,C,B	100%
A,C,D	100%

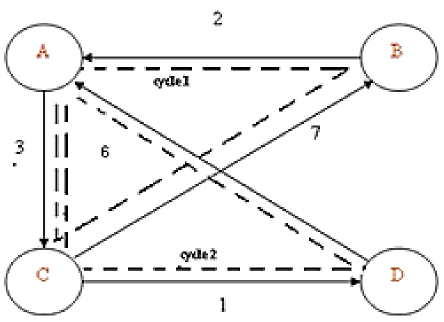


Figure 2. Mesh Network with P-cycles

There are two candidate cycles as shown in Figure 5, each of which can be in either the clockwise (or) counter clockwise direction. According to Agrawal & Srikant (1994), Győrödi & Győrödi (2002), generate the association rules for the Table 4 and formed two cycles. As cycle 1 has 3 on cycle spans, one straddling span and its current confidence is 100%.As cycle 2 has 3 spans and one straddling span and its confidence is 100%. The current confidences for all cycles in Figure 2 are given in Table5.As the P-cycle of cycle 1 and 2 have the maximum ratio.

2. Method II

This method is an efficient method of finding all frequent paths without candidate generation (Han et al 2000). For explanation of algorithm, consider the Table 1, consisting of 12 demands with minimum support count required is 20%. It is based on step1 already define in the previous section.

In step 2, consider the minimum support count be 20% then, the demands D= {(a, 10),(c,10),(d,7),(b,6)}, D is sorted in the order of descending support count. This order should maintain because each path of FP-tree will follow it.

In step 3, the root of the tree is labelled as NULL. The Fp tree reads one demands at a time and map it to a path and the nodes in each demands are processed in D order then the branch is created for each demands.

For example, the scan of first demands a, c, b which contains 3 nodes. The first branch of the tree with tree nodes {(a:1), (c:1), (d:1), (b:1)} where (a:1) is linked as a child of the root, (c:1)is linked to (a:1), (b:1) is linked to (c:1). The second demand there is no separate branch because already the demand was define in the first demand. The third demand, (d:1) branch out from (c:2) now, the support count of (a:3) and (c:3). The demands are defined in the following Figure 3.

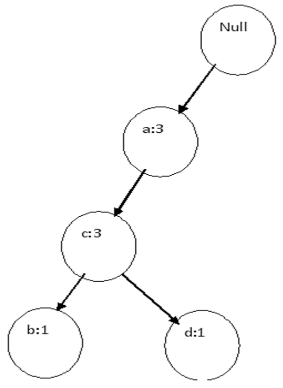


Figure 3. Fp-tree with Three Demands

For the remaining demands can be inserted in the same way in the Figure 3. To each tree traversal, header table is built so that it points to its occurrences in tree via chain of node-link.

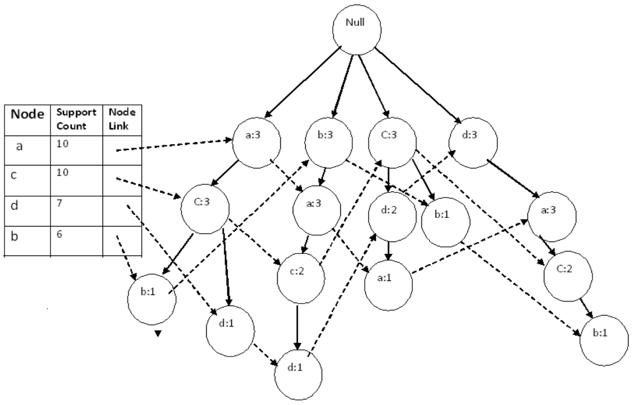


Figure 4. Fp-tree with 12 Demands

In step 4, apply the Fp method for extracts the frequent path from the tree. In our example, according to D, the complete set of frequent nodes can be divided into sub tree without overlapping.

The node occurs in two branches of the FP-tree of Figure 2. The occurrences of d can easily found by starting from the header table of d and following d's node links. the path formed by these branches are {(a:3) (c:3) (d:1) } and {(b:3) (a:3) (c:2) (d:1) } which formed its conditional pattern. Its conditional FP-tree contains only {(a: 3) (c: 3) (d: 1)} and {(b: 3) (a: 3) (c: 2) (d: 1)}. There are two P-cycle are formed in the F p tree for the same node d. Similarly, the same steps is followed by all the node and form the P-cycle when the link failure in occurred during the sending packet between source and destination. As cycle 1 has 3 on cycle spans, one straddling span. As cycle 2 have 3 spans. The cycles are already defined in the Figure 2.

The problem can be solved jointly. For a more optimum solution, the problem can be solve jointly, where the candidate working routes and the candidate backup P-cycles are jointly formulated as an Hybrid Algorithm to minimize the total capacity require. However, the numbers of possible P-cycle candidates grow exponentially with the average nodal degree and the number of nodes in the network. Furthermore, the number of variables used to represent flow, used to each light path demand increases with the number of links and the number of wavelengths. As a result, the complexity of the problem grows very rapidly, which makes the solution time of the problem unacceptable. To reduce the complexity of the problem and to make it more tractable, the routing problem is first solving using path flow technique. The idea is to generate the shortest routes for each s-d pair to be use as candidate routes in the optimization model. The candidate primary routes together with the candidate cycles are then used to formulate the overall problem as and Hybrid Algorithm. To reduce the number of candidate P-cycles in the formulation, a novel pre-selection algorithm has been developed to select a reduced number of high merit cycles.

3. Method III

Assume that the user given minimum support min-sup=3.Scan the demand once from the Table 1 and find frequent 1-node as {A, B, C, D}.Here, we need to generate sequential patterns prefixed with each frequent 1-node X.

Assume that we want to generate patterns prefixed with item say 'A'. To do this, let us first project the demands prefixed with 'A' as follows as shown in the Table 6. In this case, if there is no candidate after 'A' in a sequence, then it is discarded.

Now, we find the frequent 1-node satisfying the given min-sup in the projected demand. The item 'C' satisfies the constraint and it becomes frequent 1-node. Now, we generate a pattern as 'AC' and projected the demand again with prefix as 'AC'.

Table 6. Demands Prefixed with 'A'

S.No.	**Demands**
1	A->C->B
2	A->C
3	A->C->D
4	A->C
5	A->C->D
6	A->C->B
7	A->C

However, the newly projected demand is same as the above one. Since the node 'B' and 'C' have the support count as 2, they cannot become frequent nodes. Hence, the algorithm that generates patterns prefixed with ‘A’ terminates here. Otherwise, if min-sup is 2, then patterns A->C, A->C->B, A->C->D will be generated as frequent patterns. As per the above discussion, this method is yield two P-cycles.

IV. CONCLUSION

The standard three mining algorithms are used to deploy the P-cycles. The aim of this work is how the P-cycles are formed. The P-cycles are completely connected graph. If a P-cycle is completely connected, traffic can be routed from one node to any other node in one logical hop. Such P-cycles will enhance node processing capability and reduce hardware cost. Also, it will result in better band-width utilization, as sharing of a light path between traffic flows due to different node pairs, is reduced. The first two algorithms need more number of scan and the execution time also high. But the prefix algorithm is minimizing the number of scans while mining frequent path in a demand.

V. REFERENCES

[1] Abdelhamid, E, Eshoul & Hussein T Mouftah 2009, ‘Survivability Approaches Using P-cycles in WDM Mesh Networks Under Static Traffic’ , IEEE/ACM transaction on Networking, vol. 17, no.2.

[2] Ahmed, CF, Tanbeer, SK & Jeong, BS 2011, ‘A Framework for Mining High Utility Web Access Sequences’, IETE Technical Review, vol.28, no.1, pp.3-16.

[3] Aiman Moyaid Said P D D, Dominic & Azween B Abdullah 2009, ‘A Comparative Study Of FP-Growth Variations’, IJCSNS International Journal of Computer Science and Network Security, vol.9, no.5.

[4] Baglioni, M, Ferrara, U, Romei, A, Ruggieri, S & Turini, F 2003, ‘Preprocessing and Mining Web Log Data for Web Personalization’, Proceedings of the 8th Italian Conference on Artificial Intelligence, NCS vol.2829, pp.237-249.

[5] Borges, J & Levene, M 2000, ‘Data Mining of User Navigation Patterns’, Web Usage Analysis and User Profiling, vol.1836, pp.92-111.

[6] Chan, K et al 2002, ‘Comparison of Machine Learning and Traditional Classifiers in Glaucoma Diagnosis’, IEEE Trans. Biomed. Eng., vol.49, no.9, pp. 963-974.

[7] Grover WD & Doucette J 2002, ‘Advances in optical networks design with P-cycles: Joint optimization and pre-selection of candidate P-cycles’, in proc. IEEE/LEOS, All-Optical Networking Conf., Mont Trem-blant, Qc, Canada, pp.WA2.49-WA2 -50.

[8] Hongsik Choi, Suresh Subramaniam & Hyeong-Ah Choi 2004, ‘Loopback Recovery from Double-Link Failures in Optical Mesh Networks’, IEEE Transaction on Networking,vol.12, no.6.

[9] Joshi, A 2001, ‘Web/data mining and personalization’, University of Maryland Baltimore County (UMBC) eBiquity Research Area, http://ebiquity.umbc.edu/project/html/id/17/Web-Data-Mining-and-Personalization.

[10] Kosala, R & Blockeel, H 2000, ‘Web Mining Research: A Survey’, In ACM SIGKDD Explorations, vol.2, pp.1-15.

[11] Poonam Sharma & Gudla Balakrishna 2011, ‘PrefixSpan: Mining Sequential Patterns by Prefix-Projected Pattern’, International Journal of Computer Science & Engineering Survey (IJCSES), vol.2, no.4.

[12] Taifei Zhan, Hongfang Yu & Lemin Li 2006, ‘A Novel Algorithm For Node-Encircling And Link Candidate P-cycles Design’, in WDM Mesh Network vol.29, no.7, pp.1227-1233.

[13] Zhenrong Zhang, Wen-De Zhong, & Biswanath Mukherjee 2004, ‘A Heuristic Method For Design Of Survivable WDM Networks With P-cycles’, IEEE Communications letters, vol.8, no.7.