**Optical Burst Switching (OBS) Network Performance Variables Description**

Adel Rajab

Department of Computer Science and Engineering,

University of South Carolina,

Columbia, SC, USA, 29208

rajaba@email.sc.edu

### OBS Performance Variables

One of the primary challenges in identifying the risks of Burst Header Packet (BHP) flooding attacks in Optical Burst Switching networks (OBS) is the scarcity of reliable historical dataset. Despite recent efforts by researchers in the research domains of OBS networks and machine learning there are no public datasets that record the key variables related to the OBS network performance so administrators and researchers are able to detect misbehaving edge nodes. This detection will enable blocking misbehaving edge nodes so other behaving edge nodes will be able to effectively reserve network resources, improving the resource management and QoS of the network. In this dataset, we record useful numbers of features that have proved to be effective in detecting the behaviour of edge nodes in OBS networks.

In this paper, we highlight important variables that possibly can be effective in detecting misbehaving nodes in OBS networks. More importantly, we describe the variables and the process of recording them.

**1.1 Binary Training Dataset Preparation**

To prepare the training dataset for the purpose of identifying misbehaving nodes that are causing BHP flooding attack, thousands of simulations runs were carried out in order to collect the different variables related to the OBS network performance. Significant variables recorded included the sending node number, allocated bandwidth, bandwidth used, bandwidth lost, packet transmitted, packet dropping rate, packet received, transmitted byte, received byte, and average delay time per second , and the percentage of BHP flooding attack. At the initial phase, the edge nodes were classified into only two classes, Behaving (B) and Not Behaving (NB), as the issue was to identify the Misbehaving nodes that reserve resources without the right usage and secure the network by the right action such as blocking it. We refer to this dataset as the ‘binary dataset’ since we have only the two classes B and NB.The following strategy has been developed to create the training dataset.

1. Set the duration of the simulation to 10 minutes and the number of edge nodes to M.
2. Record the numerous variables (see Table 1 for details). For illustration purposes, Table 1 highlights just four iterations for two of the edge nodes (3, 9).
3. The edge nodes’ bandwidth capacity varied during the simulations in order to assess different situations. This is done to ensure all possible cases – normal, contention and congestion are covered.
4. Repeat for *N* number of the simulations.

Table 1. Sample four iterations of two edge nodes

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Itr #** | **Node** | **FB** | **UB** | **LB** | **PSBy** | **PT** | **PR** | **PL** | **ByT** | **ByR** | **ADT** | **PDR** | **ByLR** |
| 1 | 3 | 100 | 97.279 | 2.721 | 1440 | 9048 | 8651 | 397 | 13029120 | 12457440 | 0.003 | 4.366 | 4.388 |
| 1 | 9 | 100 | 73.811 | 26.189 | 1440 | 9048 | 6564 | 2484 | 13029120 | 9452160 | 0.003 | 27.420 | 27.454 |
| 2 | 3 | 200 | 195.075 | 4.925 | 1440 | 18096 | 17351 | 745 | 26058240 | 24985440 | 0.002 | 4.106 | 4.117 |
| 2 | 9 | 200 | 149.693 | 50.308 | 1440 | 18096 | 13320 | 4776 | 26058240 | 19180800 | 0.002 | 26.382 | 26.393 |
| 3 | 3 | 300 | 292.466 | 7.534 | 1440 | 27092 | 26001 | 1091 | 39012480 | 37441440 | 0.001 | 4.020 | 4.027 |
| 3 | 9 | 300 | 224.629 | 75.371 | 1440 | 27092 | 19969 | 7123 | 39012480 | 28755360 | 0.001 | 26.281 | 26.292 |
| 4 | 3 | 400 | 390.285 | 9.715 | 1440 | 36140 | 34700 | 1440 | 52041600 | 49968000 | 0.005 | 3.976 | 3.985 |
| 4 | 9 | 400 | 296.258 | 103.743 | 1440 | 36140 | 26346 | 9794 | 52041600 | 37938240 | 0.001 | 27.095 | 27.100 |

**Variables’ Descriptions**

* **Itr**: The iteration number.
* **Node**: The edge node label.
* **FB**: Initial Reserved Bandwidth assigned (given) to each node, the user (usr) in the experiments assign these values.
* **UB**: This is what each node could reserve from the assigned Bandwidth from FB column. The drops here are due to congestions.
* **LB**: The amount of lost Bandwidth by each node from the assigned Bandwidth at column FB.
* **PSBy**: Packets size in Byte assigned specifically for each node to transmit. Note: 60 Byte will be added to the 1440 for the IP Header and the UDP Header ((Data size 1440 Byte) + (IP Header 40 Byte) + (UDP Header 20 Byte)) =1500 Byte.
* **PT**: Total transmitted packets (per second) for each node based on the assigned Bandwidth.
* **PR**: Total received packets (per second) for each node based on the reserved Bandwidth.
* **PL**: Total lost packets (per second) for each node, which based on the lost Bandwidth.
* **ByT**: Total transmitted Byte (per second) for each node.
* **ByR**: Total received Byte (per second) for each node based on the reserved Bandwidth.
* **ADT**: Average Delay Time (per second) for each node. This is (End-to End Delay).
* **PDR**: Percentage of Packets Drop Rate for each node.
* **ByLR**: Percentage of Lost Byte Rate for each node.

In order to increase the statistical significance of the variables by smoothing their values and reducing the data variations or sudden drops in an iteration for each variable, the initial dataset (binary) and the augmented dataset (multi-class) in section 1.2 were pre-processed. This was done by computing the average for 10 consecutive iterations per variable, and for each node as one new data instance. More specifically, for each node variable, the value of the data instance at its first iteration was the first 10 consecutive values’ averages in iteration 1 to 10, while at iteration 2, the second new instance value is iteration 2-11’s average values, etc. The statistical mean from the primary data was used to compute the new data instances in order to boost the variables’ statistical meaning through levelling their values and limiting the variations amongst the data as well as rapid drops for each variable. Finally, the binary dataset’s class values were assigned by a domain expert using a rule of thumb based on two of the variables: the premeditated false resource utilization rate (percentage of BHP flooding attack)and the actual packet drop rate.

**1.2 Multi-Class Training Dataset Preparation**

To improve the presentation of data further, and to better demonstrate the BHP flood attack scenario during the simulations, the binary dataset was augmented and a new dataset (multi-class dataset) was created. The target classes of the multi-class dataset was linked with four possible values, i.e. Behaving-No Block (No Block), Not Behaving-No Block (NB-No Block), Not Behaving-Wait (NB-Wait), and Not Behaving-Block (Block). These were assigned to a new column called “New Class: Action” based on the level of BHP flood attacks for 200 runs and using two edge nodes. Information from the five sample runs from the two nodes over a simulation of 10 minutes is shown in Table 2. The simulator was used to produce the BHP flooding attacks, but without pre-setting values. Situations for the edge nodes that (in simulation runs) would end up in random levels of BHP flood attacks were created. A point has been made to show scenarios in which there are occupied resources in the OBS network without utilization with different occupancies.

For illustration purposes, Table 2 shows how the edge nodes for every iteration was used to send data, and included the iterations which predominately featured misbehaving actions from both edge node. Iteration #1, for example, shows that node 9 has been assigned class = Block, as the BHP sent by this node did not contain data (high BHP flooding) meaning that a significant proportion of the BHPs had been reserved without use. However, iteration #3 shows misbehaving actions in node 9, the node was still permitted to transfer data. Iteration #5 appears more complicated, as each of the nodes shows misbehaving actions, but not at the level of a significant BHP flood attack. As its dropping rate was smaller, node 3 was given higher priority over node 9.

Table 2 variables have been derived from Table 1.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Iteration** | **Node#** | **10-Run-AVG-PDR** | **10-Run-AVG-Bandwidth-Use** | **10-Run-ADT** | **Node Status** | **Flood** | **Class** |
| 1 | 3 | 0.135 | 0.813 | 0.001 | B | 0 | No Block |
| 1 | 9 | 0.539 | 0.239 | 0.002 | NB | 0.512 | Block |
| 2 | 3 | 0.062 | 0.886 | 0.001 | B | 0 | No Block |
| 2 | 9 | 0.503 | 0.328 | 0.001 | NB | 0.387 | Block |
| 3 | 3 | 0.074 | 0.869 | 0.001 | B | 0 | No Block |
| 3 | 9 | 0.336 | 0.437 | 0.002 | P NB | 0.134 | NB-No Block |
| 4 | 3 | 0.080 | 0.866 | 0.001 | B | 0 | No Block |
| 4 | 9 | 0.342 | 0.503 | 0.002 | P NB | 0.087 | Block |
| 5 | 3 | 0.172 | 0.737 | 0.00092 | B | 0.017 | NB-No Block |
| 5 | 9 | 0.339 | 0.490 | 0.00094 | P NB | 0.091 | NB-Wait |

**Variables’ Descriptions**

* **Itr**: The iteration number.
* **Node**: The edge node label.
* **10-Run-AVG-PDR**: Average packet drop rate for ten consecutive iterations
* **10-Run-AVG-Bandwidth-Use**: Average Bandwidth utilized for ten consecutive iterations
* **10-Run-ADT**: Average delay time for ten consecutive iterations
* **Node Status**: initial classification of nodes based on PDR, Bandwidth Use and ADT. B= Behaving, NB= Not Behaving and P NB= Potentially Not Behaving.
* **Flood**: % of flood per node based on PDR
* **Class**: The final classification of nodes based on PDR, FB, Iteration #, Bandwidth Use, ADT, etc.