

WORLD WIDE WEB — A GRAPHICAL APPROACH

S. LAKSHMINARAYANA*

VISHWAS CHAVAN**

*/ National Institute of Oceanography, Regional Centre, Visakhapatnam - 530 017, India

** National Institute of Oceanography, Dona Paula, Goa - 403 004, India

ABSTRACT

World Wide Web (WWW) is a union of all networks that are present and accessible through a physical or logical network. In this paper, the fundamentals of WWW, the relative approach to the graph theory are discussed. Some flow parametric characteristics in terms of graph theory concepts are explained.

INTRODUCTION

A graph is defined on a given set V , which may be finite or infinite; whose elements be called points, vertices or nodes. By graph, we mean an ordered pair $G(V, E)$ where E is the set of two element subsets of V (in general, by an M -set we mean, a set of M elements) called edges of G . A digraph is a network with distinguished subsets of systems that act as source and sink as the case may be. These systems are connected through a media (physical or logical) for the transport of data packets between the systems. The efficiency of the network is dependent on the maximum rate at which the data packets can be transported through the media. The general belief is that the bandwidth is the measure of efficiency of the network but definitely not true because increasing bandwidth is increasing the velocity limit on a congested road. The reflexive, symmetric and transitive properties of the graph theory are explained in several books [1] and no explanation is needed. These properties can be imposed directly on the Web without any change except with the due changes in the nomenclature. Studies are to be arrived into the areas of prioritization of flow and other characteristic parameters.

DISCUSSION

The flow packets is only the interest of the user

which is an integer valued function f defined on the media such that $0 \leq f(a) \leq c(a)$ for all 'a' belongs to the media parameters that belongs to the set of payload factor, media access time, time required to transmit request packet and time required to transmit the response packet. The sum of the last three factors is denoted as Node Service Time (NST). Some factors such as total client service time, file server to workstation service time and client identification time do not contribute much to the flow and hence are not significant in the flow parameters. Here the values of $f(a)$ are a function of the rate at which the data packets are transported and $c(a)$ is the maximum capacity of the media [2].

$f(+v) \geq f(-v)$ for 'v' belongs to the total number of packets in the net at a particular time; where $f(+v)$ is the no. of data packets entered the network and $f(-v)$ is the no. of packets coming out of the network. In an ideal case $f(+v) = f(-v)$ and in all other cases $f(+v) \geq f(-v)$. $f(+v) - f(-v)$ is a variable function that depends on the network parameters and reflects the no. of loss of packets (LOP) and directly dependent on In band management, Out band management, configuration management, fault management, routing, security, performance, accounting asset, integration and many more tiny factors.

If a simple model of a network can be shown as follows :

State	Computer A	Media	Computer B
1	request data		
2		transmits the data packet	
3			takes time to respond
4			sends the data packet
5		transmits the data packet	
6	time takes to receipt		

In case of 1,3,4 and 6 states the media is idle. This idle time is used to propagate the data packets for a different set of nodes. The n-node digraph has nC_2 ways to propagate the datagram to next node. This walk of datagram will have n! no. of ways (Of course need not be true in real sense).

In a smaller network, it is not very difficult to maintain tables mapping names to addresses. If the table is too large, it will be difficult to handle. Unlike this the WWW network is a cluster of computers that spread across the globe and it is very difficult to manage all the information in a table or set of tables. The importance of holding a name to address table is obvious and there should be another method that resolves this job. The DNS database will help in resolving this issue.

The comparison will satisfy all relations of a tree/graph to smaller networks. In a wider networks and WWW; the situation is comparable to cyber /bond space theory where; each vertex act as a cluster of computers (network) and the only difference is incidence matrix. In case of bond space and digraph networks

$$\begin{aligned}
 m(v,a) &= 1 \text{ (if a is a link and v is the tail of a)} \\
 &= 0 \text{ (if a is a link and v is the head of a)} \\
 &= -1 \text{ (otherwise)}
 \end{aligned}$$

In case of undirected graphs

$$\begin{aligned}
 m(v,a) &= 1 \text{ (if a is a link and v is the node of a)} \\
 &= 0 \text{ (otherwise)}
 \end{aligned}$$

The host file will indicate the source of information to which a search has to be made. Further studies have to be carried out in the area of prioritizing the hosts by

assigning weight vectors, finding the nearest weight vector, updating of weight vector using different techniques. One such method is broadcast method [3]

$$W_j(t+1) = W_j(t) + C_k(t,r_j) [X(t) - W_j(t)]$$

Where $C_k(t,r_j)$ is the adaption function

(value varies [0,1]) r_j represents neighborhood function and the value is one if node is winner and the value is zero if the node is not a winner.

In the first iteration t varies from 0 to t-1 (t being the total no of data vectors, in case of WWW number of hops that data packet passes) [4]

$$\begin{aligned}
 C_k(t_0) &= 1 - (t+1)/t \text{ for the first pass} \\
 &= 0.25 (1 - (t+1)/t) \text{ for the second pass}
 \end{aligned}$$

$$\begin{aligned}
 C_k(t_1) &= (1 - (t+1)/t)^2 \text{ for the first pass} \\
 &= 0.0 \text{ for the second pass}
 \end{aligned}$$

Once the address resolution part is successful, the only matter is to establish a link between them and get connected. The congestion, bandwidth and all communication concepts will play a great role in doing this. To avoid congestion different methodologies are adopted. A method widely used is time-out based congestion control, where a packet is killed if it retains for more than a specific time in the net. Some other methods are delay based congestion avoidance where constant measuring the delay and taking corrective action on the flow and negative feed back congestion control [4]. The net gets advantage in implementation of ATM and other type of high-speed hardware.

The WWW net will have no control on these factors except each system has its own characteristics of flow of packets within it. The flow is successful, if the packet falls within the limit of characteristic parameters of the system or through any system that is within the route of a packet. The route of packets is not same for all of them and the IP address, host and destination address on the data packet will decide and resolve the complexity of the transport of the packet.

While comparing the smaller network to a WWW, the following issues are important and are classified into two categories. (I) Issues within our net and (II) Issues out side the net.

Issues within the net are IP address, address resolution, IP routing, Control Message Protocol Domain Name Server (DNS) and issues out side the net are SLIP, PPP, Network Information System, Network File System,

UUCP accounts, Mail and NNTP.

The IP address of our network usually shares addresses from a logical valid IP address. If there are several such IP addresses, the same may be on sub net. Similarities can be explained in terms of a graph, where in a graph two nodes cannot be called with the same node name. A host server does address resolution in case of local, where in a WWW this is done by the environment (!). To activate such resolution the host server should be configured for binding with outer net. The concept could be compared with a unification of several trees into a graph and a packet is moving from a leaf node of one tree to other. Control Message Protocol will deal with error conditions with the net and in WWW the same could be achieved through another host.

CONCLUSIONS

In the case of larger networks the source (X0) and the destination (XN) will have a path through several systems (networks). The data packet will undergo through different protocols before reaching the destination. The 'traceroute' or similar function attempts to show the walk of the datagram and for example if X0U1 X1U2 X2U3 X3..... XN is the walk of a datagram, where Xi (i=1,N) are the nodes in case of graph theory and hops in case of WWW; the only comparison interested here is the Ui (i=1, N-1), where Ui can be explained in terms of weight vectors in case of graph theory and protocol characteristics in case of WWW.

The set of Xi (systems) has different technologies to identify the shortest path by classification of servers / identification of IP address and server resolution methods. The priorities (Ui) have to be studied carefully, because they will play very important role in forwarding a datagram through it (!). The user thus in effect gets

the sum of $X_i * U_j$ (i=1,N) and '*' indicates the binary operator [5]. The permissible meaning of the binary operator could be evaluative linkage between the system and other characteristic parameters of an adjacency matrix (Say Uij) defined by $U_{ij} = W [E_{ij}]$; where W is the weight vector or priority that amount on the walk of the data packet across the source and destination, and E ij is the effective data transport or flow parameters between the node I to the node J within the walk.

ACKNOWLEDGMENTS

The authors express their thanks to Dr. E. Desa, Director, National Institute of Oceanography, Dona Paula, Goa for his permission to publish this paper and Dr. K.S.R. Murthy Scientist-in-Charge of National Institute of Oceanography, Regional Centre, Visakhapatnam, and Albert Gouveia of NIO, Goa for their support in doing this work.

REFERENCES

- [1] Harary, F., Graph theory, Addison Wesley Pub. Com., 1972
- [2] Andrew S. Tanenbanum, Computer Networks Prentice-Hall (India) Ltd., 2nd Edition, 1986.
- [3] Brahmabatt and Shah, R, Image data compression for broadcast methods, CSI Communications, 1995
- [4] Ganesh R.A. Congestion control in computer networks, CSI Communications, Feb 1996.
- [5] Peay E.R., Matrix operations, properties of networks and directed graphs, Jour. of Mathematical Phy 15(1), 1977, pp 89-101.