year	month	day
2013	1	8

Department	Electronic and Information Engineering	ID	085401
Name	Chunghan Lee		

Supervisor	Kyoji Umemura

A b s t r a c t

Title	Measurements, Analysis, and Modeling for Network Throughput Prediction on the Internet

(800 words)

Data volume transferred on the Internet is growing explosively along with the advancement of computer technology and the popularization of smart mobile devices. To take account of the data volume, introducing alternative systems or redesigning the whole Internet might be adequate but cannot be achieved promptly. Introducing bandwidth reservation mechanism is one of possible directions of the adequate evolution. However, dedicated network infrastructures would be necessary for implementing such a kind of mechanisms. On the other hand, a method that efficiently utilizes the Internet is promising. Network throughput prediction on the Internet, that is another approach for the efficient utilization of the Internet, does not require such infrastructures, and it can help to enhance site selection on multiple sites and grid applications by reducing data transfer time.

Although many solutions have been proposed for predicting the network throughput, they suffer the following obstacles. First, it is hard to be modeled mathematically because distribution of traffic fluctuation is unclear. Second, changes in data are very large because the scale and bandwidth of network are rapidly increased each year. Third, there is noise occurred by abrupt changes in network state.

In this dissertation, throughput prediction methods and its application to efficient utilization of the Internet are explored.

As a first step, we introduce and discuss research issues of Internet traffic characteristics when virtualization technologies are used on the Internet. We gather throughput measurements on the Internet and appropriate prediction parameters are selected through statistical analysis before the actual prediction. In the throughput measurements, we clarify extra effects caused by the virtualization. Although network state is stable, the measurement results are fluctuated by the effects. A previous prediction parameter is inappropriate for the precise prediction, and noise and non-linear characteristics are found.

Next, we propose a throughput prediction method with improved prediction results. Machine learning techniques, that find patterns or characteristic features, are applied on the throughput measurements to automatically determine an appropriate regression curve and to deal with prediction models represented as noise and non-linear characteristics. The prediction results of our method compared to those of the previous prediction method are more accurate on the same condition.

Finally, we investigate the contributions of prediction results through grid simulation. Our prediction method and the previous prediction method were adopted by a grid scheduler and their performance was quantitatively compared to that of the scheduler without any throughput predictions. Through the simulation, the prediction results will not always contribute to reduce processing time on given tasks. Most results reduce overall processing time using our prediction method.