Traffic Dynamic in Modularity Structure of Complex Networks.

Suyong Eum*, Shin'ichi Arakawa*, and Masayuki Murata*
*Osaka University, Graduate School of Information Science and Technology
1-5 Yamadaoka, Suita, Osaka, 565-0871 Japan
Email: {suyong, arakawa, murata}@ist.osaka-u.ac.jp

***Abstract*— Modularity structure is often found in many different types of complex networks. Especially, this topological property is believed to provide a certain degree of robustness to networks. This belief leads us naturally to investigate the impact of modularity structure on the behavior of networks. In this paper, we carry out two simulation studies to monitor traffic dynamic in modularized and non-modularized scale free networks. In the first simulation, coarse-grained traffic is flooded through networks. This study demonstrates that a modularized structure localizes damages and stops the malicious effect to the whole system. In the second simulation, more fine-grained traffic is used to investigate how a modularity structure impacts on the saturation of networks. This result shows that a strong modularity structure is saturated much faster than non-modularized topology as traffic load increases, and additionally we show how this early saturation in modularized structure can be overcome.**

***Index Terms*— Modularity, Traffic Dynamic, Critical point.**

I. Introduction

Many different types of properties to characterize complex networks have been proposed, and some of these such as average path length, clustering coefficient, and degree distribution, have been adopted to explain robustness of complex networks. For instance, the degree distribution has attracted great attention from researchers after it was discovered that degree distributions of many different complex networks can be better described by a power law form $P(k) \sim k^{\gamma}$ rather than the conventional Poisson distribution. The power law degree distribution implies that a few nodes have extremely large number of links while large number of nodes have small number of links. Thus, if we assume that high degree nodes are more important than less degree nodes, the power law networks (also called scale free networks) provide clues that they are error tolerance - randomly chosen node is likely to be small degree nodes- and attack vulnerable - when high degree nodes are intentionally chosen [1]. In other words, complex networks with a scale free property are robust to a random failure and fragile to an intentional attack.

While adopting these basic properties have been well explored to explain robustness of networks, other types of properties such as modularity and hierarchy attracted less attention relatively because of its mathematical complexity in analysis. In spite of its difficulty, it is worth to investigate these structural properties for many different reasons such as evaluating the performance of networking protocols, assessing the effectiveness of proposed techniques to protect the network from nefarious intrusions and attacks, and developing improved designs for resource provisioning [2]. Especially, one of these topological structures called modularity is found in various complex networks including biological systems [3]. Since biological systems benefit from its evolutionary advantage of such an topological architecture, it is natural to ask why the topological architecture of biological systems ends up with a modularity structure through the evolution process. If we understand the advantage of the topological structure found in biological systems, we may be able to apply our understanding for building an artificial network so that the man-made systems can be as robust as biological systems.

As a result, in this paper we present some simulation results to observe the behavior of modularized and non-modularized topologies in terms of traffic dynamic. We make use of two simulation models which were proposed in [4] [5]. Firstly, traffic fluctuation is simulated with coarse-grained traffic which can be separated into internal and external fluctuations. This separation enables us to understand the origin of traffic fluctuation caused by two different components which are a topological structure and a traffic burstiness. Secondly, fine-grained traffic is flooded through topologies to observe the variation of critical points which represents the saturation of a network.

The rest of this paper is organized as follows. In Section II, we provide brief descriptions of two models which have been used to generate scale free topologies. This section also describes how the strength of modularity structure is defined. This is followed by a detail description of the first simulation scenario called coarse-grained traffic fluctuation in Section III. Section IV explains a different type of simulation to observe the saturation of networks with finer scale traffic. The both Sections include various numerical results based on two different simulation scenarios. Finally, we conclude the paper in Section V.

II. Two models for topology generation with modularity structure

Since network behaviors in modularized and non-modularized topologies are investigated, these two types of topologies are required for our investigation. Two well known models are used to generate these topologies, namely Barab*á*si and Albert (BA) model [6] and Fabrikant, Koutsoupias and Papadimitriou (FKP) model [7] models. They are based on two representative theories called the preferential attachment and the optimization process that explain the power-law phenomenon observed in many different fields of sciences.

A. Barabási and Albert (BA) model

Barabási and Albert (BA) [6] explained the emergence of power law of degree distribution in many complex networks using the preferential attachment mechanism. Based on the mechanism, they proposed a model to generate a power law topology. When a new node arrives, it is connected to an existing node which is chosen probabilistically from $\prod(k_i) = k_i/\sum_j k_k$, where k_i represents the number of degrees in a node i. It implies that a node with large degrees has a high probability to be chosen from a new arrived node. This phenomenon is also described as rich-get-richer. In the model is there a parameter m which decides the number of existing nodes to which a new arrived node attaches.

B. Fabrikant, Koutsoupias, and Papadimitriou (FKP) model

Fabrikant, Koutsoupias, and Papadimitriou (FKP) [7] proposed a model to explain of the power law distribution of degree in the Internet topology. The model imitates human behavior that minimizes the construction cost of a network. More specifically, they attach a new arrived node to an existing node by minimizing their proposed objective function $\alpha \cdot d_{ij} + h_j$, where d_{ij} is the Euclidean distance (i.e., geographical distance) between nodes i and j, and h_j is a hop counts distance between node j and a pre-specified $root$ node (node 0). The former distance represents construction cost and the latter does operating cost. Since two distances are complement each other, the parameter α gives weight to one of distances. When α has a low value such as 0, the geographical distance is ignored so that each new arrived node connects to a root node directly, so it results in a star topology. On the other hand, when the parameter is high, each node tries to connect to the geographically closest node so that it creates a topology which is similar to a random network. In this paper, we make use of values for α as used in [7]. Another interesting observation is that topologies from this model tend to have strong modularity structure. It is because geographically closed nodes are grouped together so that it forms a modularity structure.

C. Quantifying Modularity

Modularity has been studied in many different areas with different names such as community structure, graph partitioning in graph theory and computer science, and hierarchical clustering in sociology [8]. Basically, this property shows how easily a network can be divided into groups.

In [8], Newman et al developed an algorithm to quantify the strength of modularity structure. The method follows an iteration process. The first step involves finding of a link where most flows use (such a link is called a link with high betweeenness centrality). A link with the highest betweenness centrality is searched and removed continuously until the removal splits the network. When it is split, the strength of modularity is calculated using the following equation.

$$\mathrm{Q} = \sum_i (e_{ii} - a_i^2) \tag{1}$$

where Q is the quantified modularity value, e is a symmetric matrix which represents the connectivity among modules. The dimension of e is the same as the number of modules in the network. Also, a_i represents a row or column sum of matrix $a_i = \sum_j e_{ij}$. The value Q shows the ratio between the number of links inside modules and that of links between modules. Thus, relatively less number of links among modules produce a high value of Q.

Since the iteration process produces a series of modularity values, the maximum value is chosen to be the modularity value of the network.

III. Coarse grained traffic fluctuation.

In this simulation scenario, two traffic models are considered. In both models, a certain number of walkers W are located in randomly chosen nodes and move around a network M steps. The difference between two models is how the walkers are routed in a network. In the first model, a random routing is simulated which represents the behavior of an intruder in a system. In the second model, walkers follow along the shortest path between nodes. In both models, to obtain time to time fluctuation on each node i, different number of walkers are launched on different types of networks T times independently.

One way to characterize the dynamic of traffic in this scenario is to capture the relation between mean $E[V]_i$ and standard deviation $\sigma[V]_i$ of the number of visits of walkers in each node i. The relation is often represented using following equation.

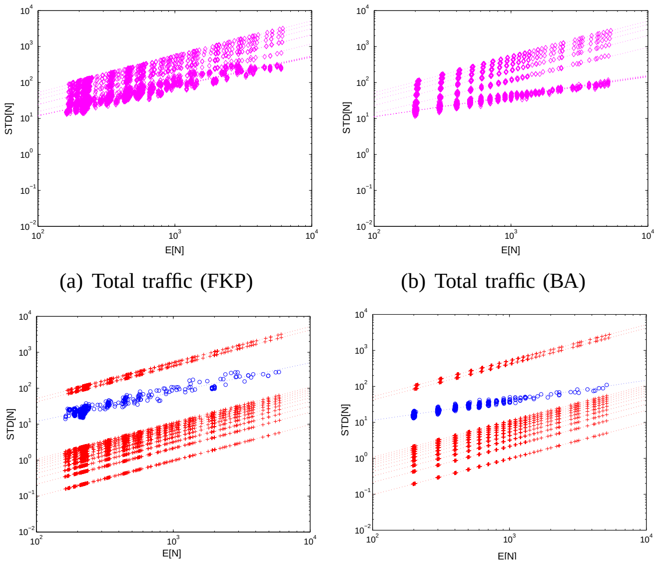
$$\sigma[V]_i = E[V]_i^{\alpha} \tag{2}$$

When we plot standard deviations as a function of ascent sorted means in log scale, the exponent α characterizes the level of fluctuation of traffic. In [4], a model was proposed to identify the origin of traffic fluctuation. The model separates the original traffic fluctuation $f_i(t)$ at node i into internal $f_i^{int}(t)$ and external $f_i^{ext}(t)$ traffic fluctuations as follow

$$f_i(t) = f_i^{ext}(t) + f_i^{int}(t) \tag{3}$$

where $f_i^{ext}(t) = \frac{\sum_{t=1}^{T} f_i(t)}{\sum_{t=1}^{T}\sum_{i=1}^{N} f_i(t)} \sum_{i=1}^{N} f_i(t)$ and $f_i^{int}(t)$ can be obtained from subtracting $f_i^{ext}(t)$ from $f_i(t)$. It means that the measured $f_i(t)$ can be split into two components. The reason to choose this model for our investigation is that the internal fluctuation is related to the topological structure of a network.

Fig. 1 shows the relation among total f_i, external f_i^{ext}, and internal f_i^{int} fluctuations as the bustiness of traffic varies in two different types of topologies. The number of walkers W at time t varies from the uniform distribution $[N - \triangle W(t), N + \triangle W(t)]$, where N is the number of nodes, and $\triangle W$



(a) Total traffic (FKP) (b) Total traffic (BA)

(c) Internal & External traffic (FKP) (d) Internal & External traffic (BA)

Fig. 1. (a)(b) Before traffic fluctuation is separated into internal and external in FKP and BA topology respectively. From bottom to top, the $\triangle W$ is set to $[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 100, 200, 300, 400, 500]$. (c)(d) After traffic fluctuation is separated into internal and external in FKP and BA topology respectively. Blue circles show the fluctuation of internal traffic, and red crosses show external traffic fluctuations. From bottom to top, the $\triangle W$ is set to $[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 400, 500]$.

is from $[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 100, 200, 300, 400, 500]$. In Fig.1(a) and Fig.1(b), total traffic fluctuations f_i are plotted with fitting lines. The first five top lines in both figures show total traffic fluctuations when high fluctuated traffics ($\triangle W(t)$: [500,400,300,200,100]) are flooded , and the rest of lines (actually only one line can be seen since these lines are overlapped) are with low traffic fluctuation ($\triangle W(t)$: [10,9,8,7,6,5,4,3,2,1]). In both topologies, the first five top lines are nearly parallel with similar gradient close to 1 when high burstiness traffic are flooded. It is because the external fluctuation fully dominates the total fluctuation in both topologies. However, as the traffic burstiness (external fluctuation) decreases, the total fluctuation is under control of internal fluctuation so that different gradients start to appear when small bursty traffics are flooded. In Fig. 1(c) and Fig. 1(d), the internal fluctuations (fitting lines with blue circles) and the external fluctuations (fitting lines with red crosses) are separated from total traffic fluctuation. There are 12 different set of external fluctuations with $\triangle W$ from $[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 400, 500]$ (lines with red crosses), and one set of internal fluctuation (lines with blue circles) which is the counterpart of the 12 sets of external fluctuation. We observe only one set of internal fluctuation because internal traffic fluctuation is not affected by the variation of external traffic burstiness.

Since we make use of two different topologies, the different gradients of internal fluctuations (lines with blue circles) in both Fig. 1(c) and Fig. 1(d) are observed as 0.8154 and 0.5509 respectively. Thus, when the internal fluctuation dominates over external fluctuation, the total fluctuations in Fig. 1(a) and Fig. 1(b) are governed by internal fluctuation so that the gradient of total fluctuations are similar to that of internal fluctuation, however as the external fluctuations increase, the impact of internal fluctuation is fade out and the total fluctuation is totally dominated by the external fluctuation.

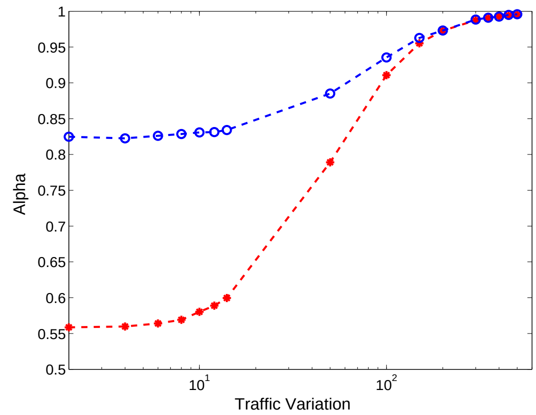


Fig. 2. The variation of exponent α as traffic fluctuation increases.

Fig. 2 shows the variation of gradients (exponent α) of total traffic fluctuation in FKP (blue circle) and BA (red asterisks) topologies respectively as traffic burstiness increases. With small external fluctuations, the total fluctuation is dominated by the internal fluctuation which is affected by a topological structure. Thus, the alpha values start from different points and gradually increases until $\triangle W(t)$ becomes around 10. Then, they rapidly increases, which show the domination of external fluctuation.

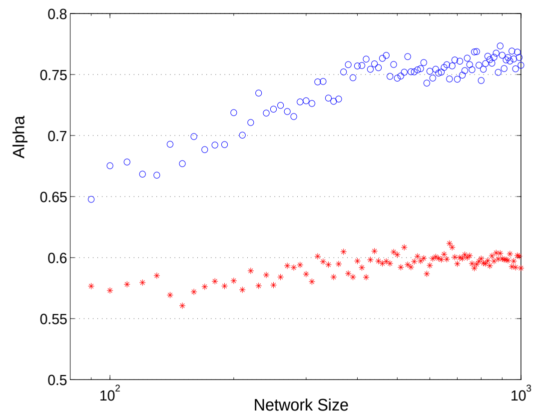


Fig. 3. The variation of exponent α as network size increases.

In Fig.3, with fixed external fluctuation, the size of network increases from 80 to 1000 nodes by 10. These topologies are generated from FKP model (blue circles) and BA model (red asterisks). The reason we increase the sizes of FKP and BA topologies is that the strength of modularity structure in FKP topology increases as the size of network increase, while, that of modularity structure in BA topology does not change much. From the figure, we can see that strong correlation between modularity structure and traffic fluctuation represented as an alpha which shows the relation between mean and standard deviation shown in Equation (2). The alpha value in FKP topologies proportionally increases from 0.65 to 0.75 as the size of topology increases, while for BA topology the alpha value does not change very much. It shows that topologies with strong modularity structure experience higher traffic fluctuation. In other words, when same amounts of traffics are observed in nodes in modularized and non-modularized topologies, the node in modularized topology experiences higher fluctuation than one in non-modularized topology. High traffic fluctuation in a modularized topology seems to be negative effect in terms of robustness point of view. Then, why many biological systems have strong modularized structure? This question leads us to investigate more about this phenomenon.

In the previous results, each walker moves around topologies randomly. However, data exchanged between nodes in real networks follow fixed routes in order to maximize performance of systems. Thus, we simulate another traffic model based on the shortest path routing. In Fig.4, each point represents the variance and mean ratio (VMR), and the mean of traffic in time series in each node. Especially, in Fig.4(a), Fig.4(c), Fig.4(e), and Fig.4(g), we classify the nodes according to their location in a network using our previous work [9]. There are four different groups in these figures. Data points (red upside-down triangles) in the most right side show data observed in nodes on the top layer of topology. Nodes on the top layer mean nodes which are used to connect among modules. Then, nodes in sub-layers are decided according to hop counts from the nodes on the top layer. Since topologies from BA model are fully connected each other, nodes are not classified into multiple layers as shown in Fig.4(b), Fig.4(d), Fig.4(f), and Fig.4(h). First thing we see is that the large amount of traffic is observed on nodes on the top layer in modularized topology in Fig.4(e) and Fig.4(g). It shows these nodes actually experience a bottle-neck phenomenon.

Most interesting observation in these results is that a modularized topology experiences high fluctuation when walkers are routed randomly as shown in Fig.4(a) compared to Fig.4(b). However, when walkers are routed along the shortest path, both modularized and non-modularized topologies experience the same fluctuation in Fig.4(c) and Fig.4(d). Under the assumption that the random routing represents the behavior of an intruder, the large fluctuation in Fig.4(a) can be translated that intruders are trapped inside modules so that they rarely damage nodes outside the module they belong to. It shows that a modularity structure localizes or isolates damages caused by an intruder.

IV. Fine grained traffic fluctuation

As the second part of experiment, a different type of simulation scenario is used based on a model in [5]. This scenario is involved with a fine traffic control. In each time step, packets are generated with a probability ρ from randomly chosen nodes and forwarded along the shortest path to randomly chosen the other nodes. In this scenario, every node has a queue. Thus, when traffic load increases by increasing the probability ρ, packets begin to be accumulated in queues and experience delay. These packets in queues are served based on FIFO(First-In First-Out) principle, and move to the next hop according to the probability calculated as follow:

$$\mu_{i \to j} = \frac{1}{(n_i n_j)^\gamma} \tag{4}$$

where n_i and n_j represent the number of queued packets in node i and j, and γ is a parameter which controls the speed of packet forwarding process in a node. Thus, the probability $\mu_{i \to j}$ means that the probability of a packet moving from node i to node j is inverse proportional to the number of packets in node i and node j and proportional to the control parameter γ. In order to observe the saturation of a given network, we use the order parameter [5] as follow:

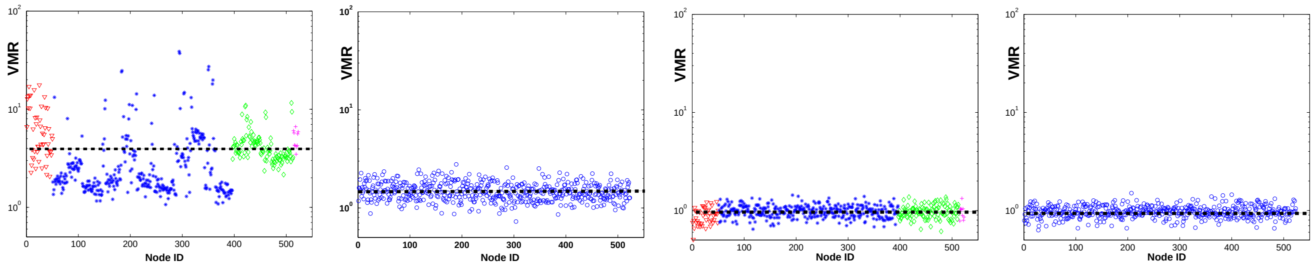
$$\eta(p) = \lim_{t \to \infty} \frac{1}{\rho S} \frac{< \triangle N >}{\triangle t} \tag{5}$$

where $\frac{1}{\rho S}$ shows the total number of packets offered to a network with the number of nodes S, and $\frac{<\triangle N>}{\triangle t}$ means the rate of packet increase in a network. Fig. 5 plots the order parameters as a function of ρ with fixed $\gamma = 0.01$ in modularized (blue circles) and non-modularized (red asterisks) topologies. The offered traffic ρ, where the order parameter η becomes non-zero, is called a critical point ρ_c that a network begins to saturate. In the figure, modularized topology reaches to the critical point much earlier than non-modularized topology. It is because, as we saw in Fig.4(e) and Fig.4(g), some nodes in modularized topology have high betweenness centrality [1] , so that these nodes cause modularized networks to be collapsed much earlier than non-modularized networks.

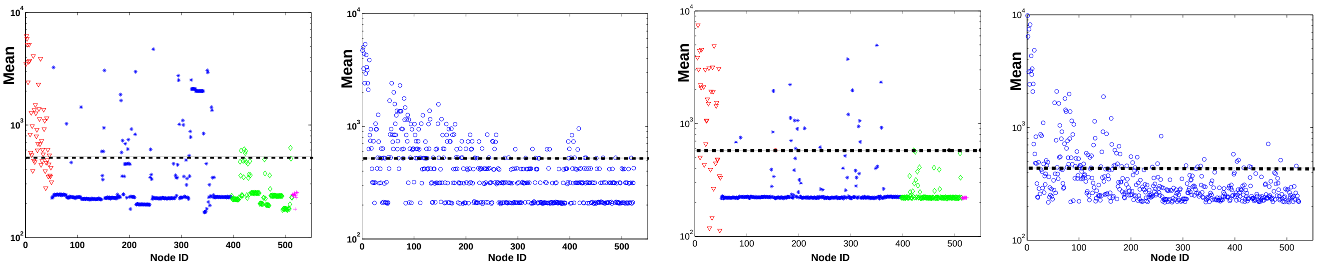
In Fig.6, the variation of critical points ρ_c are plotted as the size of both FKP (blue circles) and BA (red asterisks) topologies increase from 100 to 1000. Although, the critical points differ from each other, the critical values in both topologies are reduced at the similar rate. We believe that the scale free property in both topologies cause this result. It means that modularized topology keeps a certain characteristic of scale free property.

The early saturation experienced by modularized topology can be overcome in many different ways. Since the main reason of this early saturation is that the top layer nodes in a modularized topology cause a bottleneck phenomenon,

[1] shows how heavily a node is used by flows that exist between every two nodes in a network



(a) Random Walk (FKP) (b) Random Walk (BA) (c) Shortest Path (FKP) (d) Shortest Path (BA)



(e) Random Walk (FKP) (f) Random Walk (BA) (g) Shortest Path (FKP) (h) Shortest Path (BA)

Fig. 4. Variance and mean ratio(VMR), and mean of traffic variation in each node of highly modularized ((a)(c)(e)(g)) and non-modularized topology ((b)(d)(f)(h)). (a)Nearly 70% (365/523) of total nodes are under the average VMR(3.9488 black dot line). (b)55% (288/523) are under the average VMR (1.4947- black dot line). (e)80% (427/523) nodes are under the total mean average(523), also the average mean traffic of nodes located in the first layer (1480) (red down side triangles) is well over 2 times larger than total average (black dot line). (f)70% of nodes are under the average(523 black dot line)(c)A half of total nodes (260/523) are under the average VMR(0.9711 - black dot line). (d)Around 45% (235/523) are under the average VMR (0.9454 - black dot line).(g)90% (471/523) nodes are under the total mean average(580.38), also the average mean traffic of nodes located in the first layer (3359) (red down side triangles) is well over 2 times larger than total average (black dot line).(h)79% of nodes are under the average(437.28 black dot line).

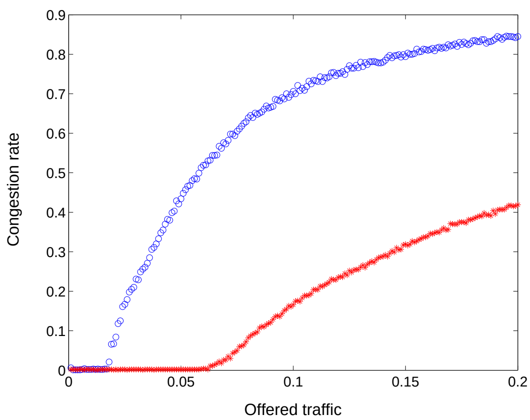


Fig. 5. The variation of critical points as the offered traffic increases in modularized network (blue circles), and in non-modularized network (red asterisks)

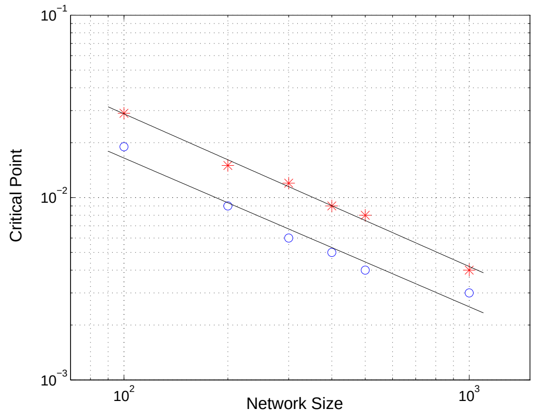


Fig. 6. The observation of critical points in modularized network (blue circles), and in non-modularized network (red asterisks)

increasing the speed of packet processing only in these nodes releases the networks from the early saturation. Fig.7 shows the variation of critical points as the speed of packet processing increases in modularized topology. We decrease the values of γ in Equation (4) to increase the speed of packet processing in nodes. The blue circles show the variation of critical points when the packet processing power of top nodes increases, and the red triangles show a result when nodes other than top layer nodes are chosen to increase the packet forwarding power. It demonstrates that actually top layer nodes cause the bottleneck and this bottleneck phenomenon can be resolved by increasing the capacity of these nodes on top layer.

V. Conclusions

In this paper, traffic fluctuation in modularized and non-modularized topologies has been investigated with various simulation scenarios.

In the first scenario, we simulated coarse-grained traffic fluctuation. Initially, we separated the observed traffic fluctuation into internal and external fluctuations as proposed in [4] to identify the impact of topological impact on traffic fluctuation.

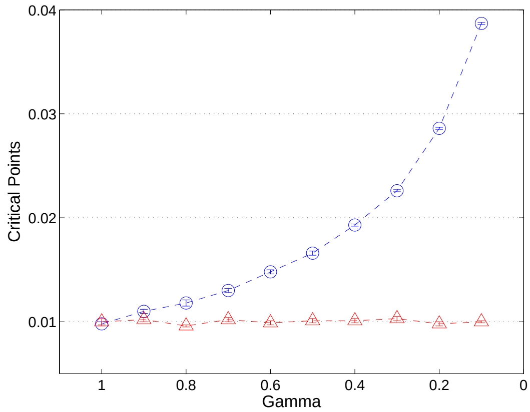


Fig. 7. The variation of critical points as γ in Equation (4) varies in nodes on the top layer of modularized network (blue circles), and in nodes on other than top layer (red asterisks)

We found that a topology with strong modularity structure tends to experience high traffic fluctuation. However, this fluctuation is observed only when traffic is routed randomly which represents the behavior of intruders in a network. It implies that modularity structure provides a natural protection for the system against attacks from an intruder.

In the second scenario, fine-grained traffic is flooded through networks to observe how modularity structure impacts on the saturation of different types of networks. A topology with high modularity structure becomes saturated earlier as traffic load increases than one with non-modularized structure because of a bottleneck phenomenon of modularized topology. In addition, the saturation points decreased at the similar rate in both topologies, we believe, it is because both are scale free topologies although their structures are different. This bottleneck phenomenon can be overcome by increasing the speed of packets processing in nodes which are used to connect modules.

Finally, it is worthwhile to emphasize that the FKP model, we used to generate topologies with strong modularity structures, needs to be improved since it was initially designed for ISP-level topology with a scale free property. Thus, we leave the improvement as our future work.

Acknowledgment

This work was supported in part by a grant of Special Coordination Funds for Promoting Science and Technology: Yuragi Project from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

[1] R. Albert, H. Jeong, and A. Barabsi, "The Internet's Achilles' Heel: Error and attack tolerance of complex networks," 2000.

[2] L. Li, D. A, W. W, and J. D, "A first-principles approach to understanding the internet's router-level topology," in *SIGCOMM '04: Proceedings of the 2004 conference on Applications, technologies, architectures, and protocols for computer communications*. New York, NY, USA: ACM Press, 2004, pp. 3–14.

[3] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A. L. Barabasi, "Hierarchical Organization of Modularity in Metabolic Networks," *Science*, vol. 297, pp. 1551–1555, 2002.

[4] M. A. de Menezes and A.-L. Barabasi, "Fluctuations in network dynamics," *PHYS.REV.LETT*, vol. 92, p. 028701, 2004. [Online]. Available: http://www.citebase.org/abstract?id=oai:arXiv.org:cond-mat/0306304

[5] A. Arenas, A. Díaz-Guilera, and R. Guimerà, "Communication in Networks with Hierarchical Branching," *Phys. Rev. Lett.*, vol. 86, no. 14, pp. 3196–3199, April 2001.

[6] A. L. Barabasi and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, October 1999.

[7] A. Fabrikant, E. Koutsoupias, and C. Papadimitriou, "Heuristically optimized tradeoffs: A new paradigm for power laws in the internet," in *Heuristically optimized tradeoffs: A new paradigm for power laws in the internet*. 34th Symposium on Theory of Computing, 2002.

[8] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," August 2003.

[9] S. Eum, S. Arakawa, and M. Murata, "A new approach for discovering and quantifying hierarchical structure of complex networks." in *ICAS '08: Proceedings of The Fourth International Conference on Autonomic and Autonomous Systems*, Gosier, Guadeloupe, 2008.