Extraction of Parameters from Well Managed Networked System in Access Control

Akira Kanaoka*, Masahiko Katoh†, Nobukatsu Toudou*, and Eiji Okamoto*

*University of Tsukuba
1-1-1 Tennoudai, Tsukuba, Ibaraki, Japan,
{kanaoka, okamoto}@risk.tsukuba.ac.jp, toudou@cipher.risk.tsukuba.ac.jp

† IIJ Technology Inc.
1-105 Kanda Jinbo-cho, Chiyoda-ku, Tokyo, Japan
masa@iij-tech.co.jp

Abstract

Designing a networked system (NS), which consists of various network equipments and uses LAN technology, has become increasingly important. However, there have been few studies on NS design. Our previous study suggests that a well managed NS from an access control viewpoint has a fixed degree distribution, regardless of connection restriction. In this paper, we find an approximation function of its distribution using Genetic Algorithm. Found parameters show a well approximation result in all type of networked system. Furthermore, we propose a new measurement method for a well managed NS in access control using found parameters.

1. Introduction

When we want to provide a service through the Internet, the service is rarely provided by only one server. Usually, various network equipments are assigned to the system, and LAN technology is used to maintain redundancy, scalability, and security. We call such a system a Networked System (NS).

It is complicated to design an NS if it is a small system. Thus, recent business growth through the Internet requires higher these characteristics to NS. Although NS design is increasingly important, it depends on professional experience, and there has been little research on NS design.

Since an NS consists of several network devices such as servers, routers and switches, most previous studies considered networks of single role node [1]–[7] . Although we cannot apply these studies to the NS design, the model of NS proposed by A. Kanaoka et al. [8] realizes a expression of multiple roles in one model without loss of its characteristics. This model opens the way to build a methodology of NS design.

Constraints for network like cost, flow, are usually given for optimum network design. To discuss for optimazation of network design, a measurement method proper for constraints is significant for network design problem.

In this paper, the approximation function of degree distributions on well managed NS using the model by A. Kanaoka et al. [8], is obtained. Such approximation function $f(x)$ suggests a well managed NS has common characteristics in degree distribution. The meaning of "well managed NS" in this paper is an NS which communication link is controled , that is, access controled, by proposed efficient algorithm.

Moreover, the measurement method on NS design from access control point of view, is proposed using obtained approximation function $f(x)$. Using proposed measurement method as a design constraints, we can discuss NS design problem and its complexity.

Such measurement of NS which includes several network devices and communicates over several layers, is not realized before.

In Section 2, we outline previous studies on network design. Degree distributions of NS are discussed in Section 3. In Section 4, an approximation function of degree distributions is obtained using Genetic Algorithm (GA). A new measurement method showing a status of an NS from access control point of view, is proposed in Section 5. We then conclude the paper in Section 6.

978-0-7695-3612-5/09 $25.00 © 2009 IEEE
DOI 10.1109/ICIMP.2009.17

2. Related Work

Network design problems have widely studied. The main purpose of network design problem is to minimize cost or flow of networks or subgraph. .

Hayrapetyan, et al. suggest a new class of network design problems for information aggregation [1]. Sadagopan, et al. propose a systematic decentralized approach towards the design of networks [2]. Both are focused on sensor networks which is one of the hottest topic in network design. However, each node in sensor networks has basically the same role.

Chekuri studies three models to discuss network design problem and propose about robust network, vpn design, routing [3]. Lau et al. propose approximation alogorithms for survivable network design problem and subgraph problem [4]. Melkonian et al. consider directed graph in network design problem and focus on finding a minimum cost subgraph [5]. Rosenberg take hierarchical topological approach to design networks and proposed algorithm consider maximum network diameter [6]. Belotti et al. propose algorithms to realize networkd design with complicated node cost [7]. Although these studies acquire solutions using such constraints and requirements appropriately, most of these studies focused on networks of single role nodes, that is, homogeneous node networks.

On the other hand, networks of an NS consist of various roles, including Hub, Switch, Router, Firewall, Load Balancer, Server, Database, and Proxy, that is heterogeneous node networks. A.Kanaoka, et al. propose a new NS model that enables several pieces of network equipment without loss of its characteristics [8]. We call this model ”NSQ model” in this paper. NSQ stands for Networked system Security Quantification. Moreover, they suggests that a well managed NS from an access control viewpoint has a fixed degree distribution, regardless of connection restriction.

3. Degree distribution of NS

3.1. NSQ model

In NSQ model, each network equipment like server, hub, switch, router, etc. is called a Module. A module has several nodes for each layer (Table 1), according to its role. For example, a router module (Layer 3 Relay : L3R) has several layer 3 (L3) nodes and no nodes above layer 4 (L4). Six modules are defined in NSQ model. The Service module provides a service such as WWW, e-mail, or DNS. The Internet module provides Internet service and is the source of communication with the system. The Layer 1 Relay (L1R), Layer 2 Relay (L2R), Layer 3 Relay (L3R), and Layer 4 Relay (L4R) modules are the relay modules at each layer.

Each node is connected by a link. There are two types of link: a link between nodes of different modules and the same layer, and a link between nodes of the same module and a different layer. The first type link expresses the possibility of direct communication of its layer. The second type link expresses the relationship between nodes in the module. Fig.1 and Fig. 2 show an example NS in the proposed model.

Table 1. Layer definitions

Layer 5	Abstracted service (WWW, DNS, etc.)
Layer 4	Services by port number (80, 53, etc)
Layer 3	IP
Layer 2	MAC address space
Layer 1	Physical object

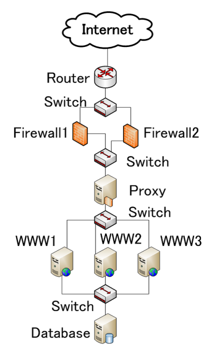

Figure 1. Current expression of NS

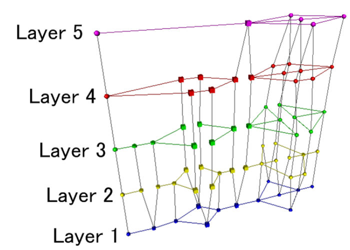

Figure 2. Expression of NS by the proposed model

3.2. Efficient connection algorithm

To realize optimum access control for NS, the number of links between nodes must be minimized without isolation of nodes in each layer, that is, minimizing access paths between nodes. However, the optimum access control method seems difficult to realize, and realizing optimum access control method is out of scope in this paper.

At first, relationships between abstracted services are drawn as L5 links and abstracted services are drawn as L5 nodes. Abstracter service (L5) nodes consist of several L1 nodes, that is, a partial L1 network is built in several modules. Then, L2 networks are decided automatically using L1 network.

Network connection on L3 and L4 depends on its network access control rules, capacity of processing, redundancy, and so on. From access control point of view, if an NS has a loose connection rule, most nodes are connected even though there are unnecessary communication paths. In this section, we describe a simple efficient algorithm which realizes a sufficient reduction in the number of links on L3 and L4. In this paper, we call an NS which is designed using the efficient algorithm as a well managed NS.

We first extract candidate nodes of L4 to be connected according to L5 connection and then connect candidate L4 nodes with a minimum number of links. Candidates are decided by existence of L4R modules inside L5 node. If an L5 node has an L4R module, L4 nodes of L4R are chosen as candidates. If not, the L4 nodes of all modules inside this L5 node are chosen as candidates. Next, L3 nodes are also connected according to L4 connection.

The algorithm of efficient connection on L3 and L4 is shown in Algorithm 1. The method $connect(S_A, S_B)$ in Algorithm 1 realizes the minimum number of links between S_A and S_B. $candidateL4nodes()$ is the method for selecting candidate of L4 nodes in one of the L5 nodes.

Algorithm 1 Efficient connection algorithm

Require: Link Set L
Ensure: Link Set L'
$L' \leftarrow L$
for all L5 nodes x, y such that $(x, y) \in L$ **do**
    $S_A \leftarrow candidateL4nodes(x, L)$
    $S_B \leftarrow candidateL4nodes(y, L)$
    $L' \leftarrow L' \cup connect(S_A, S_B)$
end for
for all L4 nodes x, y such that $(x, y) \in L'$ **do**
    $p \leftarrow$ L2 node such that (p, k) and $(k, x) \in L'$ for some L3 node k
    $q \leftarrow$ L2 node such that (q, l) and $(l, y) \in L'$ for some L3 node l
    $V \leftarrow searchRoute(p, q)$
    $L' \leftarrow connectRoute(x, y, V)$
end for

3.3. Degree distribution

When two nodes can communicate, there is a link between those nodes. If access control policies are different in two NS, these degree distributions are also different.

Therefore, in this section, we compare degree distributions between loose connection and well managed access control connection. The target of NS is providing Web service, which is the most typical internet service. The requirement of services in an NS is fixed at four: Internet (I), Web Server (W), Application Server (AP), and Database (DB). Each service is assigned as a L5 node. Its possible number of 4-node L5 network is 27 except isomorphic network. Based on these L5 networks,

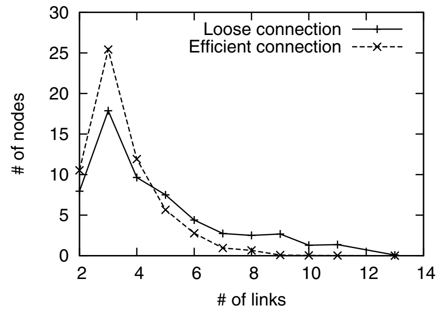

Figure 3. Degree distribution: # of nodes = 58

we made dataset of all considerable NS. Actually we made two datasets of NS: dataset of NS with loose connection and dataset of NS with well managed connection. In these datasets, the maximum number of module on an NS is 15 and the maximum number of nodes on an NS is 72. The number of NS in one dataset is 10764.

Fig. 3 shows the degree distribution of NS in which has 58 nodes.

There are peaks on the number of links with 3 (L=3) in both loose and efficient cases. The efficient access control case has a larger number of nodes in L=3 than that in the loose case. In L=4, the numbers of nodes are similar. Then, L>5, and the loose access control case has a larger number of nodes. The efficient case has a higher peak and lower skirts, and the loose case has a lower peak and relatively higher skirts, that is, it is flatter than the efficient case.

In this paper, due to space limitations, we present the results for only one case. However, the results are similar for all cases.

Next, influences of module with respect to degree distribution is investigated. Fig. 4 shows the degree distribution of the 63-node case of the NS with and without L1R, and the degree distribution of the 63-node case of the NS with and without L3R.

There is little difference between degree distributions for efficient cases. These result suggest that a well managed NS from an access control viewpoint has a fixed degree distribution, regardless of connection restriction.

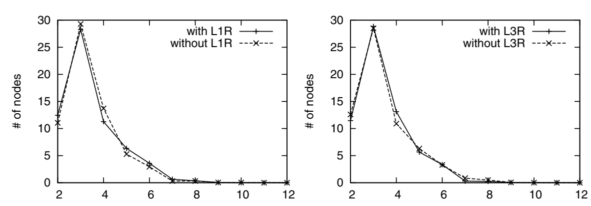

Figure 4. Degree distribution on efficient connection: # of nodes = 63

4. Approximation of degree distribution

In the field of graph theory or complex networks, the approximation of degree distribution is realized in several cases. For example. Faloutsos et al. showed that the degree distribution for the network of Autonomous Systems (AS) is approximated like $f_d \propto d^O$, where f_d is the frequency of an outdegree d and O is the constant value [9]. Such constant value O is used as a parameter of characteristic for network.

Although such approximation is represented using exponential or power-law, distributions which started from L=2, has a peak at L=3 and skirts after L=4 like Fig. 4 is hard to approximate using exponential or power-law.

We list several distribution types used in statistical field, and try to fit for degree distributions.

4.1. Target types of distribution

Types of distribution are six. Although these types are based on popular distributions in statistics, like χ square distribution, Poisson distribution, etc. , these types are selected to allow wider potential of identification for target degree distribution. For example, the probability density function of χ^2 distribution is $f(x;k) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{(k/2)-1}\mathrm{exp}^{-x/2}$ and has only one paramter k. However, in this paper, to allow similar distribution for χ^2 distribution, we take $f(x) \propto x^\alpha \mathrm{exp}(\beta x)$ and has independent two parameters (α, β). In each type, we use α, β, and/or γ as parameters.

4.1.1. Type A : F distribution type. Type A distribution is represented as follows:

$$f(x) \propto \frac{x^\alpha}{(1+\beta x)^\gamma} \quad (1)$$

The number of parameters is 3 (α, β, γ).

4.1.2. Type B : Exponential and Power-law.

Type B-1.

Type B-1 distribution is represented as follows:

$$f(x) \propto x^\alpha \mathrm{exp}(\beta x) \quad (2)$$

The number of parameters is 2 (α, β). It includes χ square distribution and Γ distribution.

Type B-2.

Type B-2 distribution is represented as follows:

$$f(x) \propto x^\alpha \mathrm{exp}(\beta x^2) \quad (3)$$

The number of parameters is 2 (α, β). This type includes inverse Gaussian distribution.

Type B-3.

Type B-3 distribution is represented as follows:

$$f(x) \propto x^\alpha \mathrm{exp}\left(\beta x + \frac{\gamma}{x}\right) \quad (4)$$

The number of parameters is 3 (α, β, γ). This type includes log-normal distribution.

4.1.3. Type C : Discrete distribution.

Type C-1.

Type C-1 distribution is represented as follows:

$$f(x) \propto \sum_{k=0}^{x} \frac{\alpha^k \mathrm{exp}(-\alpha)}{k!} \quad (5)$$

The number of parameters is one (α). This type includes Poisson distribution.

Type C-2.

Type C-2 distribution is represented as follows:

$$f(x) \propto \sum_{k=0}^{[x/2]} \frac{\alpha^{x-2k}\beta^k}{(x-2k)!k!} \quad (6)$$

The number of parameters is 2 (α, β). This type includes Hermite distribution.

4.2. Approximation method

Approximation is executed using Genetic Algorithm (GA). GA evaluates each gene (solution domain) to choice for selection, crossover and mutation using fitness function. Since a simple summation of square-error is dependent heavily on number of nodes in NS, the following S is used as fitness function.

$$S = \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i \hat{y}_i^2} \quad (7)$$

when, $\hat{y}_i = f(x_i)$.

The parameters of a gene are parameters of each type and a scale coefficient Θ. (Ex. a gene of Type A, $g = (\Theta, \alpha, \beta, \gamma)$ on $f(x) = \Theta \frac{x^\alpha}{(1+\beta x)^\gamma}$)

4.3. Approximation result

To achieve results of approximation, we divided 10764 NS patterns into groups which have same number of nodes in NS. Then GA is executed on each group. Such approximation is carried out by each Type. That is, $378 = 54$ groups $\times$ 7 types approximation is calculated.

Several results are eliminated to evaluate fitness by GA. First, we eliminate results of group which has less than 200 cases since insufficient number of cases. Second, we eliminate results of group which has large S value, as a failure of GA.

	Avg.(> 35)
Type A	0.0133
Type B-1	0.0166
Type B-2	0.0232
Type B-3	0.0161
Type B-4	0.3102
Type C-1	0.3963
Type C-2	0.0476

Table 2. Average of S in each GA approximation

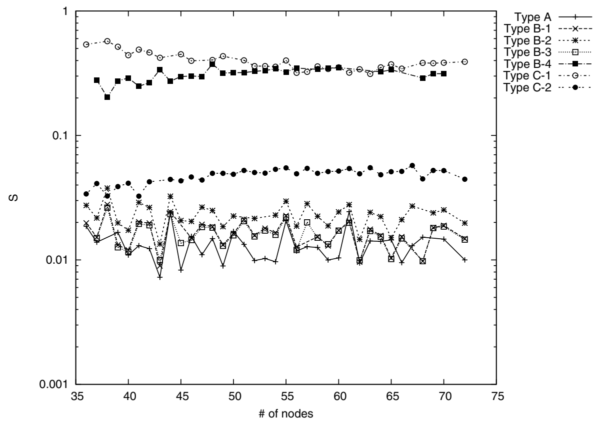

Figure 5. S in each GA approximation

Values of S are shown in Fig.5, and the average S value among eliminated group results are shown in Table 2 .

If $S = 0$, it means that the degree distribution completely fit to approximation function. S value close to 0 is better as evaluation. From Table 2 , Type A achieves good approximation to each NS degree distribution.

Next, we fix parameters except a scale coefficient Θ by averaging each parameters. (Ex. $\alpha = \sum \alpha_i / M$, M is the number of cases in summation) whether each distributions of a certain distribution function type have common parameters.

A scale coefficient Θ is decided by number of nodes. The peak value of each distribution is in L=3, and have similar ratio to total number of nodes N in NS. Fig. 6 shows the average number of nodes which has $L = 3$. The average ratio among the case which $N > 35$, is 0.43859. Θ is obtained by solving following equation.

$$0.43859N = f(3) \tag{8}$$

Table 3 shows an average value of S in fixed parameters and Fig. 7 shows S value in fixed parameters. In this result, Type A is the best function as approximation for degree distribution of efficient connected NS.

Fixed parameters of Type A are $\alpha = 29.526, \beta = 0.518$ and $\gamma = 50.019$

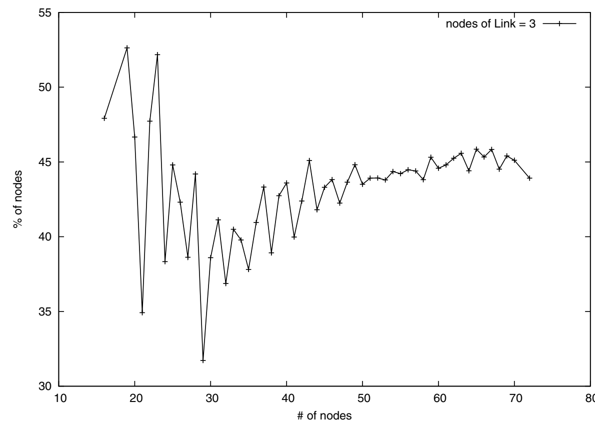

Figure 6. Avarage number of nodes which has L=3

	Avg.	Avg.(> 35)
Type A	0.0212	0.0143
Type B-1	0.0268	0.0182
Type B-2	0.0317	0.0238
Type B-3	0.0269	0.0173

Table 3. S value average in fixed parameters

5. Measurement method for access control in NS

$S = 0$ means the degree distribution completely fit to approximation function. If NS is well managed from access control point of view, its S value is close to 0. At first, We survey S values of worse design case in this section, whether S value is suitable for measurement of well access controlled NS.

S values of NS which L3 and L4 network are designed loosely, that is, each nodes in same segment are all communicable, are shown in Fig. 8. In well managed cases, values of S are nearby 0.01. On the other hand, values of S are mostly around 0.15.

Since it is hard to understand that $S = 0.15$ indicates "Not Good" for design, we propose the new value V as follows:

$$\begin{aligned} V &= 1 - \epsilon S \\ &= 1 - \epsilon \frac{\sum_i (y_i - f(x_i))^2}{\sum_i f(x_i)^2} \end{aligned} \tag{9}$$

Where ϵ is inverse of the average of S in loose cases.

$$\epsilon = \frac{M}{\sum_i S_i} = 6.386 \tag{10}$$

$V = 1.0$ means the degree distribution completely fit to approximation function. If NS are loose designed, V value

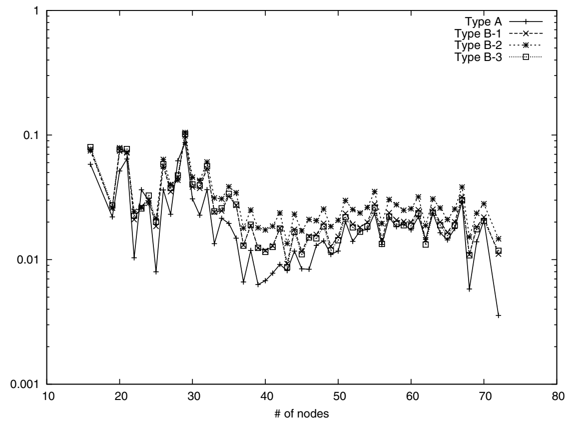

Figure 7. S value in fixed parameters

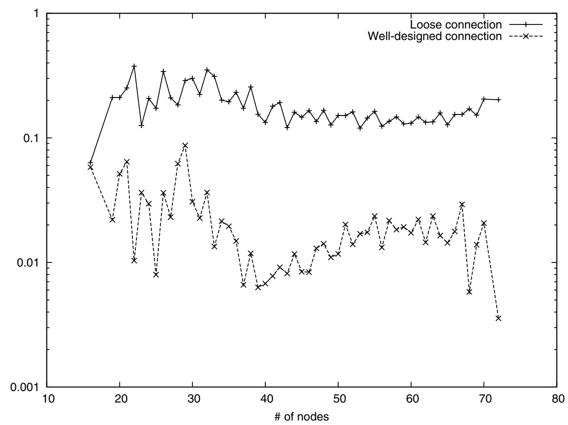

Figure 8. S value between well maneged and loose case

is pointed around 0. The state of access control in NS is measured by proposed value V.

6. Conclusion

In this paper, the approximation function of degree distributions on well managed NS using the model by A. Kanaoka et al. [8], is obtained. The obtained function is $f(x) = \Theta x^{\alpha}/(1 + \beta x)^{\gamma}$ where $\alpha = 29.526, \beta = 0.518$ and $\gamma = 50.019$. Such the fact that the function $f(x)$ is reasonably obtained suggests a well managed NS has common characteristics in degree distribution.

Moreover, the NS measurement method and its value V from access control point of view, is proposed using $f(x)$. Such measurement of NS which includes several network devices and communicates on several layers, is not realized before. Using proposed measurement method as a design constraints, we can discuss NS design problem and its complexity.

Building an optimum method to design using the proposed measurement V, finding other common characteristics in NS, and building further integrated design method including security , cost and scalability , are conceivable future works.

References

[1] A. Hayrapetyan, C. Swamy, E. Tardos: Network Design for Information Networks. Proc. of 16th annual ACM-SIAM symposium on Descrete Algorithms, pp.933-942, 2005

[2] N. Sadagopan, M. Singh, B. Krishnamachari: Decentralized Utility-based Sensor Network Design. Mobile Networks and Applications 11, pp.341-350, 2006

[3] C. Chekuri: Routing and Network Design with Robustness to Changing or Uncertain Traffic Demands. ASM SIGACT News, pp.106-129, 2007

[4] L.C. Lau, J. Naor, M. R. Salavatipour, M. Singh: Survivable Network Design with Degree or Order Constraints. STOC'07, 2007

[5] V. Melkonian, E. Tardos: Algorithms for a Network Design Problem with Crossing Supermodular Demands. Networks, Vol.43, Issue 4, pp.256-265, 2004

[6] E. Rosenberg: Hierarchical Topological Network Design. IEEE/ACM Trans. on Networking, val.13, no.6, pp.1402-1409, 2005

[7] P. Bellotti, F. Malucelli, L. Brunetta: Multicommodity Network Design with Discrete Node Costs. Networks, Vol.49, Issue 1, pp.90-99, 2007

[8] A. Kanaoka, M. Kato, N. Todo, E. Okamoto: Networked System Modeling and its Access Control Characteristic Analysis, Proceedings of World Academy of Science, Engineering and Technology (WASET), Vol. 35, pp.125-133 (2008)

[9] M. Faloutsos, P. Faloutsos, and C. Faloutsos: On power-law relationships of the Internet topology, In Proceedings of the Conference on Applications, Technologies, Architectures, and Protocols For Computer Communication, SIGCOMM '99, pp.251-262, (1999)