2009 Third International Conference on Network and System Security

F-TAD: Traffic Anomaly Detection for Sub-Networks using Fisher Linear Discriminant

Hyunhee Park
Dept. of Electrical Engineering
Korea University
Seoul, Korea
hyunhee@widecomm.korea.ac.kr

Meejoung Kim
Research Institute for Information and Communication Technology
Korea University
Seoul, Korea
meejkim@korea.ac.kr

Chul-Hee Kang
Dept. of Electrical Engineering
Korea University
Seoul, Korea
chkang@korea.ac.kr

***Abstract*—Traffic anomaly detection is one of the most important technologies that should be considered in network security and administration. In this paper, we propose a traffic anomaly detection mechanism that includes traffic monitoring and traffic analysis. We develop an analytical system called WISE-Mon that inspects the traffic behavior by monitoring and analyzing the traffic. We establish a criterion for detecting abnormal traffic by analyzing training set of traffic and applying Fisher linear discriminant method. By using the properties of distributions such as chi-square distribution and normal distribution to the training set, we derive a hyperplane which enables to detect abnormal traffic. Since the trend of traffic can be changed as time passes, the hyperplane has to be updated periodically to reflect the changes. Accordingly, we consider the self-learning algorithm which reflects the trend of traffic and so enables to increase accuracy of detection. The proposed mechanism is reliable for traffic anomaly detection and compatible to real-time detection. For the numerical results, we use a traffic set collected from campus network. It shows that the proposed mechanism is reliable and accurate for detecting the abnormal traffic. Furthermore, it is observed that the proposed mechanism can categorize a set of abnormal traffic into various malicious traffic subsets.**

***Index Terms*—Anomaly detection; Adaptive defense system; Traffic analysis and measurement; Fisher linear discriminant**

I. Introduction

As traffic on the Internet is growing tremendously and becoming complex, traffic anomalies are tending to increase [1]-[3]. Traffic anomalies are characterized by its unusual behavior and causing significant changes in a network. Various problems such as network traffic overload, flash crowds, denial of service attack, propagation of malicious worm, and network intrusions, are caused by them. Changed access pattern and infrastructural problems caused by abnormal traffic may cause even malfunctioning of network devices [1]-[4]. One of approaches to deal with such problems is implementing network systems which are simple, robust, and scalable. Moreover, frequent abnormal traffic on backbone network requires more advanced technologies for monitoring and analyzing the network traffic. There are many researches on network security in the areas of detection, identification, and prevention of propagation of abnormal traffic.

For traffic anomaly detection, there are two main approaches: signature-based approach and measurement-based approach. The signature-based (or misuse-based) approach is applying previously established rules to the incoming traffic, while the measurement-based (or anomaly-based) approach is considering normal traffic characteristics such as traffic volume and the number of flows as well as link utilization, packet loss, the distribution of IP addresses, and port number for traffic anomaly detection [4].

The signature-based approach is detecting the known attacks when they occur. It uses predefined attack signatures and compares a current event against these signatures. Even though the approach shows high detection rate for well-known attacks, it is ineffective for novel attacks or slightly modified attacks whose signature is not available. To cope with novel attacks and unknown traffic pattern, it needs lots number of signatures and requires periodical updates with the latest rules. There are tools for this approach such as SNORT and Bro, and pattern matching is one of the well-known signature-based approaches [5]-[7]. On the other hand, the measurement-based approach is designed to identify a source that exhibits deviating behavior in a system. The construction of such a detector begins with developing a model for normal behavior. A detection system can learn the normal behavior by a training data set collected over a certain time period with no intrusions. Since this is using traffic characteristics that can be observed through network monitoring, it is more flexible and more sensitive than the signature-based approach, especially for detecting new anomaly traffic [8]. However this approach needs to keep per-connection or per-flow state over a single link or node. Therefore, they require a lot of computing resources making their cost unaffordable for many ISPs. Several tools are developed for this approach such as ADAM, SPADE, and NIDES [9].

Another crucial part of traffic anomaly detection is traffic monitoring. Therefore, it is necessary to develop an efficient traffic analysis system and define parameters which characterize traffic. There are two main techniques for traffic monitoring: active monitoring and passive monitoring. Active monitoring monitors the network layer metrics such as delay, jitter, loss, bottleneck point, and available bandwidth, by actively injecting probe packets into a network. Even though active monitoring may reduce system overhead by using small

978-0-7695-3838-9/09 $26.00 © 2009 IEEE
DOI 10.1109/NSS.2009.60

number of probe packets that have smaller sizes compare to real data packet, the performance measures may not be accurate for that reason. Ping, traceroute, and Netperf are examples of active monitoring. On the other hand, passive monitoring monitors up to application-layer that includes the user traffic condition such as the sizes of bandwidth, flow, and packet, by analyzing TCP packets. Since passive monitoring monitors a lot of data packets, it has system overhead problem. However, by that reason, its performance is more accurate and reliable than active monitoring. Many passive monitoring tools for advanced network are developed such as Ntop, CoralReef from CAIDA, PMA from NLANR, IP-Mon from Sprint, NG-Mon from POSTEC in Korea, and perfSONAR-PS of Internet2 [10]-[11]. Many monitoring systems, however, have common problems such as dealing with huge amount of traffic and various newly emerging applications. There are researches dealing with such kinds of problems [12]-[13].

There are many analytical researches on traffic anomaly detection [14]-[19]. One of the analytical detection methods is applying the seasonal auto-regressive integrated moving average (ARIMA) model. In [14]-[15], the authors considered the ARIMA models of the total traffic and the specific application such as HTTP of a network. Even though the ARIMA model is an effective time-series forecasting technique, it turns out that the anomalies cannot be well captured by this model [16]. Another detection method is Principal Component Analysis (PCA), which is the best-known statistical analysis technique for detecting network traffic anomalies [17]. It is a method of reducing a multi-dimensional data set to a lower dimensional subspace. As one of PCAs, the feature distributions of traffic are used to categorize the anomalies in systematic manner in [18]. The authors compared the performances of detection by using the traffic volume and traffic features such as src and dst addresses and ports. It shows that some attacks such as Alpha Flow are highly detected by the systematical method analyzing the traffic features, while the detection rate for scanning and small size DoS attacks is very low when only the traffic volume is used for detection. To manage the wide-network traffic effectively, the anomaly detection based on PCA is proposed in [19]. They tried to figure out the problems of using PCA for traffic anomaly detection such as the effectiveness of PCA under some specific conditions. In general, the PCA based approaches exhibit a lower detection rate compared to other approaches, which is partially owing to the abrupt change of a network and the complexity of the PCA scheme.

In this paper, we propose an anomaly detection mechanism, which is a kind of PCA but using different method. As we observed from the previous researches, each detection mechanism contains some problems. For instance, in the known detection mechanisms, traffic characteristics such as amount of bandwidth and size of packets are used in a mutually exclusive way [20]. We conjectured that jointly using the traffic characteristics would give more reliable and accurate detection mechanism. With this point of view, we consider to use several traffic characteristics jointly in detection mechanism. In addition, there are many detection mechanisms including detecting malicious traffic [8]. However, most of them have to manipulate a lot of raw traffic to detect malicious traffic that causes severe overhead. To reduce such an overhead, we consider the problem of finding a criterion to distinguish abnormal traffic from the newly generated traffic in real-time. As the first step for that, in this paper, we consider a mechanism that separates the traffic into a normal group and an abnormal group. Furthermore, we consider the categorization of the abnormal traffic group into various subgroups of malicious traffic.

In this paper, we propose the *Fisher linear discriminant based traffic anomaly detection* (F-TAD) mechanism to detect the traffic anomalies. As a part of the mechanism, we develop a passive monitoring system called *Wide backbone network traffic Identification and Statistical Estimation-Monitoring* (WISE-Mon) for traffic monitoring and analysis. It is using the characteristics of traffic such as the amount of traffic in terms of bandwidth, flow, and packet at backbone router of a network. Based on the traffic monitoring through WISE-Mon, the characteristics of traffic are derived and the distribution of accumulated traffic is obtained. The proposed detection mechanism is a measurement-based approach performing mathematical methods with traffic volume. By using the Fisher linear discriminant method, we obtain a hyperplane which enables to separate the abnormal traffic from a set of traffic collection and detect abnormality in a newly generated traffic. In addition, we consider the self-learning algorithm which reflects the trend of traffic. It enables to increase the accuracy of anomaly detection. The rest of the paper is organized as follows. In section II, we explain the Fisher linear discriminant and the properties of distributions briefly. In section III, we explain the proposed F-TAD mechanism as well as the analytical system called WISE-Mon. Experimental environment and the procedure of traffic analysis is described in section IV. In section V, a hyperplane for detecting anomaly traffic is derived and the miss detection probability is presented. Finally, section VI concludes the paper.

II. Preliminaries

In this section, we briefly explain the basic concept of the Fisher linear discrimination analysis and the properties of distributions including the chi-square distribution. The Fisher linear discriminant (FLD) analysis is a method of classifying a set of data into several different categories by projecting high dimensional data onto a line [21]-[22].

Let $X = \{\mathbf{X}_i = (x_{i1}, \cdots, x_{in}) : \mathbf{X}_i \in \mathbf{R}^n, i = 1, \cdots, k\}$ be a set of vectors which consists of data with n components. We assume that there are m different categories of data and denote Y_j as a set of data in j categories. We assume that each element of X belongs to exactly one of m categories. Then $Y = \{Y_j\}_{j=1}^m$ is a collection of disjoint subsets of X satisfying $\sum_{j=1}^{m} |Y_j| = k$, where $|A|$ is the number of elements in a set A. Let μ and μ_j be the averages of X and Y_j, respectively. Then they are given by $\mu = (\mu_1, \cdots, \mu_n)^t$ and

$\mu_j = (\mu_{j1}, \cdots, \mu_{jn})^t$ with the components $\mu_l = \sum_{i=1}^{n} x_{il}/n$ and $\mu_{jl} = \sum_{\mathbf{X}_i \in Y_j} x_{il}/|Y_j|$, respectively. Here t denotes the transpose of a matrix.

Now we introduce a direction vector and two matrices called the scatter matrices of within class and between classes. The direction vector $\mathbf{w} = (w_1, \cdots, w_n)$ is a vector which is perpendicular to a hyperplane $\mathbf{w}^t\mathbf{X} + d = 0$, where $\mathbf{X} = (x_1, \cdots, x_n)$ and d is the distance from the origin to the hyperplane. We denote the scatter matrices of within class and between classes by $\mathbf{S}_{intra}$ and $\mathbf{S}_{inter}$, respectively. To define the matrices, let $\mathbf{S}_j$ be the scatter matrix of category j which is given by

$$\mathbf{S}_j = \sum_{\mathbf{X}_i \in Y_j} (\mathbf{X}_i - \mu_j)(\mathbf{X}_i - \mu_j)^t.$$

Then $\mathbf{S}_{intra}$ and $\mathbf{S}_{inter}$ are defined by

$$\mathbf{S}_{intra} = \sum_{j=1}^{m} \mathbf{S}_j \text{ and } \mathbf{S}_{inter} = \sum_{j=1}^{m} (\mu_j - \mu)(\mu_j - \mu)^t,$$

respectively. We define the sum of variance of each category around its average in the projected space and the sum of square distances between μ_j and μ as

$$\frac{1}{||\mathbf{w}||}\mathbf{w}^t\mathbf{S}_{intra}\mathbf{w} \text{ and } \frac{1}{||\mathbf{w}||}\mathbf{w}^t\mathbf{S}_{inter}\mathbf{w},$$

respectively, where $||\mathbf{w}||$ is the Euclidean norm and $\mathbf{w}^t\mathbf{S}_{inter}\mathbf{w}$ is defined by $\sum_{j=1}^{m} (\mathbf{w}^t(\mu_j - \mu))^2$.

To separate different categories conveniently, data within a category have to be aggregated while data between categories have to be separated. In other words, the distances of data between categories have to be maximized while the distances of data within a category have to be minimized. Therefore, classical Fisher criterion function is generally defined by the following objective formula:

(OPT)
$$\max_{\mathbf{w} \in D} OPT(\mathbf{w}) = \max_{\mathbf{w} \in D} \frac{\mathbf{w}^t\mathbf{S}_{inter}\mathbf{w}}{\mathbf{w}^t\mathbf{S}_{intra}\mathbf{w}},$$

where $D = \{\mathbf{w} : \mathbf{w}^t\mathbf{X} + d = 0, \mathbf{X} \in X\}$. It is known that solving (OPT) is equivalent to the following generalized eigenvalue problem:

$$\mathbf{S}_{inter}\mathbf{w} = \lambda \cdot \mathbf{S}_{intra}\mathbf{w}. \quad (1)$$

Furthermore, it turns out that the eigenvector corresponding to the maximal eigenvalue of Eq. (1) is the direction vector $\mathbf{w}$. Due to the space limitation, we omit the proof of these in this paper. It is observed that some characteristics such as the amount of bandwidth and the size of flows of daily traffic seem to be normally distributed. Furthermore, the superposition of the squared traffic of each day seems to follow chi-square distribution. The following theorems are representing the relation between normal distribution and chi-square distribution and a well-known central limit theorem [23], which will be used in section IV.

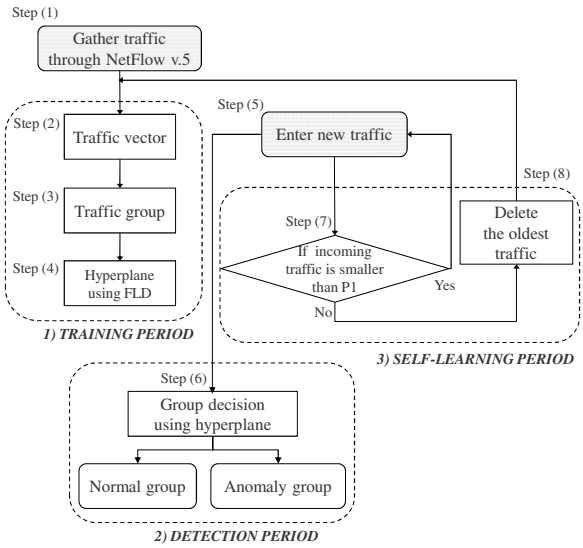

Fig. 1. Proposed detection mechanism.

Theorem 1: Let $Z_1, \cdots, Z_k$ be the independent and identically distributed *(i.i.d)* random variables with $N(0, 1)$, i.e., normal distribution with mean 0 and variance 1. Then the random variable $Z_1^2 + Z_2^2 + \cdots + Z_k^2$ follows chi-square distribution with k degree of freedom. We denote the chi-square distribution as $Z_1^2 + Z_2^2 + \cdots + Z_k^2 \sim \chi^2(k)$.

Theorem 2: Central Limit Theorem

Let $\{X_i\}_{i=1}^n$ be the *i.i.d* random variables with finite mean $E[X_i] = \mu$ and finite variance $E[X_i] = \mu$. Let $S_n = \sum_{i=1}^{n} X_i$ and define a random variable Z_n as $Z_n = (S_n - n\mu)/\sigma\sqrt{n}$. Then Z_n approaches to $N(0, 1)$, as $n \to \infty$.

III. The Proposed F-TAD Mechanism

In this section, we propose the F-TAD mechanism. The F-TAD mechanism consists of three periods: training period, detection period, and self-learning period. Before explaining the detection mechanism in detail, we define some terminologies used in this paper.

Definition 1: Let $\mathbf{X} = (x_1, \cdots, x_n)$ be a vector whose components are composed of traffic information. Then we call $\mathbf{X}$ as a *traffic vector*.

Definition 2: We call a category Y_j which is a subset of traffic vector $X = \{\mathbf{X}_i\}_{i=1}^k$ as a *traffic group*.

The components of a traffic vector can be anything observed on a backbone router such as source (src) IP, destination (dst) IP, src port, dst port, protocol, the number of flows, etc. For example, (amount of bandwidth, size of flows, size of packets) and (src IP, dst IP, src port, dst port, protocol, size of flows, size of packet) can be traffic vectors. Since a traffic group is a collection of traffic vectors having similar characteristics, we may consider several traffic groups such as normal group, abnormal group, attack group, worm group, traffic congestion group, etc. Furthermore, an abnormal group can be more

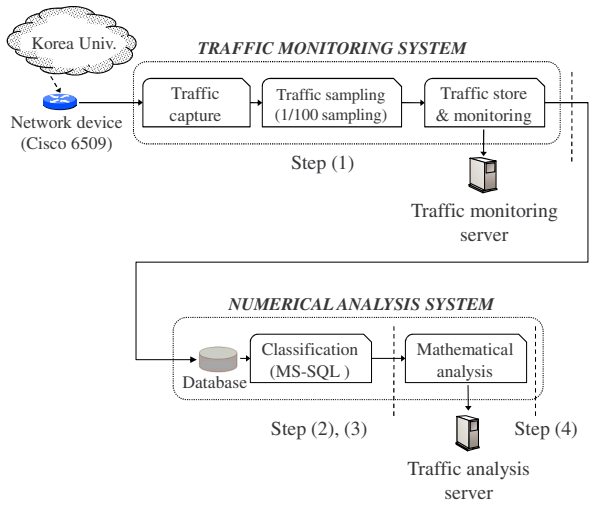

Fig. 2. Architecture of the WISE-Mon.

specified such as network operation anomaly group, flash crowd anomaly group, and network abuse anomaly group [24] by defining different traffic vectors. For example, the scanning attack can be represented as (src IP, dst port, dst IP, packet length) [25].

Now we explain the proposed F-TAD mechanism. Fig. 1 describes the operation of the proposed mechanism. In the figure, step (1)-step (4) is a training period consisting of procedures of collecting and analyzing the traffic. For this period, we developed a system called WISE-Mon. The WISE-Mon is a real-time Internet traffic monitoring and analysis system for Internet backbone networks. Fig. 2 illustrates the overall architecture of the WISE-Mon. As we see in the figure, the WISE-Mon is divided into two parts: traffic monitoring system and numerical analysis system. Traffic monitoring system consists of three phases such as traffic capture, traffic sampling, and traffic store and monitoring. While numerical analysis system consists of traffic classification and mathematical analysis. We evaluate the proposed mechanism with the traffic obtained in Korea University. Even though we are using the traffic which does not represent the traffic generated from the Internet backbone network, the mechanism can be adapted to the traffic from the larger network.

From now on, we explain the mechanism with network devices of the University and our system configuration. The raw traffic generated from in science network of Korea University is gathered in Cisco 6509 backbone router. Cisco 6509 backbone router of the University which is complied with NetFlow v.5 export formats contains 34 interfaces. The WISE-Mon captures raw packets from the backbone router and then selects samples with 1/100 rate of the captured packets. Then the sampled packets are stored in storage which is connected to the traffic monitoring server[1]. The monitoring server classifies the stored traffic according to the traffic bandwidth, flow and

[1]Through web page in traffic monitoring server, the incoming and the gathered traffic from Cisco 6509 router can be seen.
See http://163.152.31.171:8080/

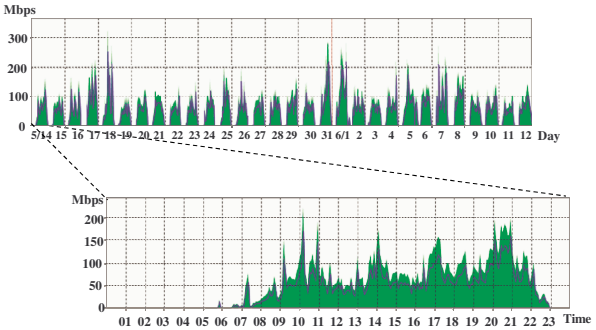

Fig. 3. 24-hour seasonal phenomenon of bandwidth.

packet information, in hour basis, interface basis, IP address basis, etc. These processes are step (1) of Fig. 1 and traffic monitoring system part in Fig. 2. Step (2)-step (3) in Fig. 1 are performed through the numerical system part in Fig. 2.

In a database of the system, traffic is collected every five minutes and is saved as an excel-file format per day. By using characteristics of the collected traffics, we decide the components of traffic vector and find a criterion of traffic group for each component of traffic vector. Based on the criterions and using some prescribed commends in MS-SQL query, the traffic are divided into several traffic groups. In the final phase in Fig. 2, the criterion for traffic detection, which is represented as a hyperplane, is derived through FLD and the properties of distributions, as described in section II. This is the end of the training period.

The next step is the detection period. When a new traffic is generated in step (5), it has to be determined which group it belongs to based on the derived hyperplane, which is step (6). In addition, we consider the dynamical update of the hyperplane. The procedure is performed by self-learning period. The purpose of performing self-learning period is to reflect the trend of newly generated traffic. With P1 as an update cycle, the P1 oldest traffic are replaced to the newly generated P1 traffic. It is step (7)-step (8). Then the procedure is repeated from step (2) for each cycle.

IV. Traffic Analysis

In this section, we analyze the traffic collected from Cisco 6509 backbone router in science campus network at Korea University. Traffic is collected for 30 days from midnight of May 14 to 23:55 of June 12, 2008. The 30 days are the ordinary one month in which no special event occurred in the campus. The traffic is collected for every five minutes, which gives 288 viewpoints per day. Therefore, a training set consists of 8640 sample traffic. In the traffic analysis, we considered the amount of bandwidth (b), the size of flows (f), and the size of packets (p) as the components of traffic vector. In other words, the training set is given by $X = \{\mathbf{X}_i = (x_{ib}, x_{if}, x_{ip}) : i = 1, \cdots, 8640\}$. In addition, we considered the two traffic groups which are a normal group Y_n and an abnormal group Y_{an}. We analyzed the trend of each component of traffic vectors $\mathbf{X}_i$ for a day and for 30

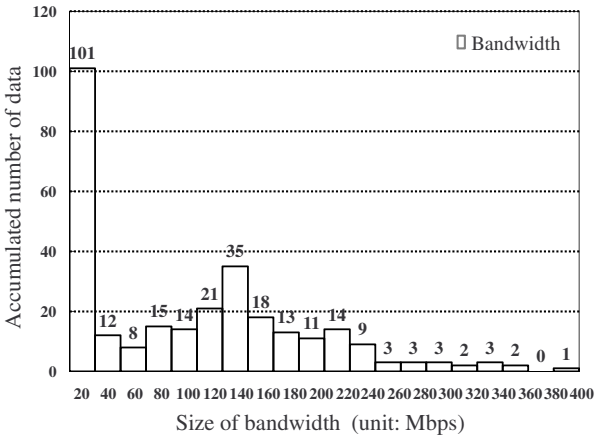

Fig. 4. Distribution of the amount of bandwidth for one day traffic.

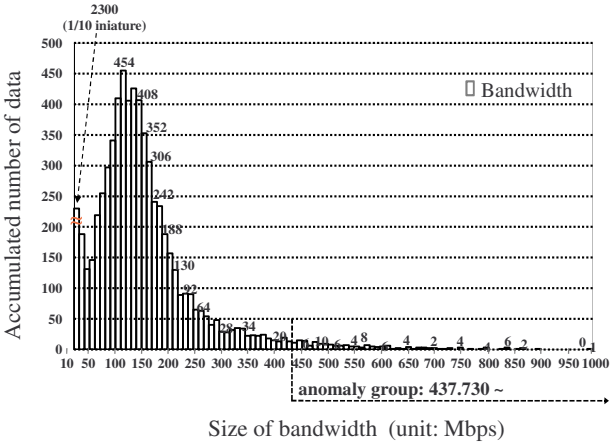

Fig. 5. Distribution of the amount of squared bandwidth for 30 days traffic.

days. Fig. 3 shows the trend for the amount of bandwidth of collected traffic for a day and for 30 days. It shows that the amount of bandwidth of collected traffic has typical 24-hour seasonal phenomenon. We have noticed that there is few traffic from 11 p.m. to 6 a.m. It is obvious since all reading room in science campus is closed during that time. Even though we omit the figures showing the trends of the sizes of flows and packets, we noticed that they also have typical 24-hour seasonal phenomenon as similar as that of bandwidth. The averages of these three components are calculated as 103,709,872.7bps, 5,247fps, and 18,336pps, respectively, as described in Table II. From the training set, we investigate the distribution of collected traffic to make a grouping criterion. Fig. 4 and Fig. 5 show the distribution of the amount of bandwidth for one day and the distribution of the amount of squared bandwidth for 30 days traffic, respectively. As shown in Fig. 4, even though it is not the continuous case, one day distribution of the amount of bandwidth looks following the normal distribution with an exception with 0∼20Mbps. The exception is caused by few traffic generations from 11 p.m. to 6 a.m. Since the small amount of traffic does not affect the system in general, it could be excluded in the analysis.

Now we introduce the well-known terminologies:

Definition 3: Traffic is *true positive (TP)* and *true negative (TN)* if it is detected correctly. In other words, normal is detected as normal and abnormal is detected as abnormal. Traffic is *false positive (FP)* and *false negative (FN)* if it is detected incorrectly. In other words, abnormal is detected as normal and normal is detected as abnormal.

Since the distribution of the amount of bandwidth for one day traffic looks like a normal, the distribution in Fig. 5 can be interpreted as a chi-square by Theorem 1 and it looks as it is. Since the traffic is collected for 30 days, we may assume that it has 30 degree of freedom. Even though the collected traffic is already filtered by the firewall of the University, we examined the the ratio of abnormal traffic by exhaustively searching from 1% to 9% as shown in Table I. Based on the table, we assumed

TABLE I
EXHAUSTIVELY SEARCHED VALUE FOR FINDING AN OPTIMAL GROUPING CRITERION.

%	Results	Probabilities
9%	TP: 106, FN: 2, TN: 148, FP: 32	P_{cor} : 0.882, P_{mis} : 0.118
8%	TP: 103, FN: 5, TN: 162, FP: 18	P_{cor} : 0.921, P_{mis} : 0.079
7%	TP: 101, FN: 7, TN: 169, FP: 11	P_{cor} : 0.937, P_{mis} : 0.063
6%	TP: 98, FN: 10, TN: 176, FP: 4	P_{cor} : 0.952, P_{mis} : 0.048
5%	TP: 101, FN: 7, TN: 179, FP: 1	P_{cor} : 0.972, P_{mis} : 0.028
4%	TP: 97, FN: 11, TN: 178, FP: 2	P_{cor} : 0.954, P_{mis} : 0.046
3%	TP: 92, FN: 16, TN: 178, FP: 2	P_{cor} : 0.937, P_{mis} : 0.063
2%	TP: 80, FN: 28, TN: 179, FP: 1	P_{cor} : 0.899, P_{mis} : 0.100
1%	TP: 76, FN: 32, TN: 180, FP: 0	P_{cor} : 0.888, P_{mis} : 0.112

that 5% of total traffic as abnormal traffic.

We define the *the correct detection probability* P_{cor} and *the miss detection probability* P_{mis} as follows:

Definition 4:

$$P_{cor} = \frac{|\text{TP}| + |\text{TN}|}{|\text{Totaltraffic}|} \text{ and } P_{mis} = \frac{|FP| + |FN|}{|\text{Totaltraffic}|},$$

respectively, where $|FP|$, $|FN|$, $|TP|$, and $|TN|$ are the numbers of false positive, false negative, true positive, and true negative, respectively.

From the table of chi-square distribution (Table C.5, [23]), we obtained $\chi^2_{30;0.05,b} = 43.7730$ which corresponds 437.730 in Fig. 5. In other words, the traffic with bandwidth larger than 437.730Mbps can be considered as abnormal traffic with this training set. This value is variable according to the training set and can be changed through self-learning period.

As a result, 8468 traffic out of 8640 are assumed to be normal while the remaining 172 traffic are assumed to be abnormal. Even though we present the bandwidth only among three components in this paper, the other two components of traffic vector are dealt similarly. Assuming 5% abnormality,

TABLE II
AVERAGE OF EACH GROUP AND AVERAGE OF TOTAL TRAFFIC.

Average	Bandwidth(bps)	Flow(fps)	Packet(pps)
Average for normal traffic μ_n	$\mu_{nb} = 87,945,088.43$	$\mu_{nf} = 5,220.758403$	$\mu_{np} = 12,951.99879$
Average for abnormal traffic μ_{an}	$\mu_{anb} = 429,317,861.4$	$\mu_{anf} = 5,789.884712$	$\mu_{anp} = 129,538.0727$
Average for total traffic μ_{total}	$\mu_{total,b} = 103,709,872.7$	$\mu_{total,f} = 5,247.040972$	$\mu_{total,p} = 18,336.00845$

the number of normal traffic for the remaining components, flow and packet, are given by 8617 and 8263, respectively. Then we applied MS-SQL query to define the criterion for classification. Commend used in the query is OR condition which is given as follows: If at least one of three components is turned out to be abnormal, the traffic vector which contains the component is regarded as abnormal traffic. In other words, only the traffic vector which has all normal components is regarded as a normal traffic. With this criterion, we obtained the training set $Y = \{Y_n, Y_{an}\}$ with $|Y_n| = 8241$ and $|Y_{an}| = 399$. As shown in Fig. 4, one day traffic collected from campus network could be assumed to follow normal distribution. However, in a larger network, it may not be true. In that case, we may use the sufficient amount of traffic as a training set. According to Theorem 2, the accumulation of the sufficient amount of traffic could be assumed to follow normal distribution. Therefore, in that case, the mechanism could be adapted just replacing the chi-square distribution to the normal distribution.

V. NUMERICAL RESULTS

In this section, we present the hyperplane which is obtained by FLD method. As we already knew from section II, we have to find the direction vector $\mathbf{w}$. To do it, we need the average of each group and average of total traffic, which are denoted by $\mu_n = (\mu_{nb}, \mu_{nf}, \mu_{np})^t$, $\mu_{an} = (\mu_{anb}, \mu_{anf}, \mu_{anp})^t$, and $\mu_{total} = (\mu_{total,b}, \mu_{total,f}, \mu_{tatal,p})^t$, respectively. The values are presented in Table II.

As shown in Table II, the difference of averages of two groups of flow component is less than those of other two components. Since small differences make it difficult to distinguish groups, the size of flow may the component that cause incorrect detection. It is obvious that if we delete any of the components from traffic vector, the exactness of detection may decreases. Conversely, as more components considered, the mechanism becomes more reliable.

With the training set, we obtained the $\mathbf{S}_{intra}$ and $\mathbf{S}_{inter}$ as follows:

$$\mathbf{S}_{intra} = \begin{pmatrix} 1.7643e^{+010} & 4.618e^{+013} & -1.0985e^{+010} \\ 4.618e^{+013} & 5.7556e^{+019} & -3.2343e^{+013} \\ -1.0985e^{+010} & -3.2343e^{+013} & 4.0687e^{+014} \end{pmatrix},$$

$$\mathbf{S}_{inter} = \begin{pmatrix} 295510.0992 & 1.7717e^{+011} & 60506850.02 \\ 1.7717e^{+011} & 1.0627e^{+017} & 3.6293e^{+013} \\ 60506850.02 & 3.6293e^{+013} & 1.2395e^{+010} \end{pmatrix}.$$

Since it is suffices to solve the eigenvalue problem Eq. (1) to find $\mathbf{w}$, we solved the characteristic equation

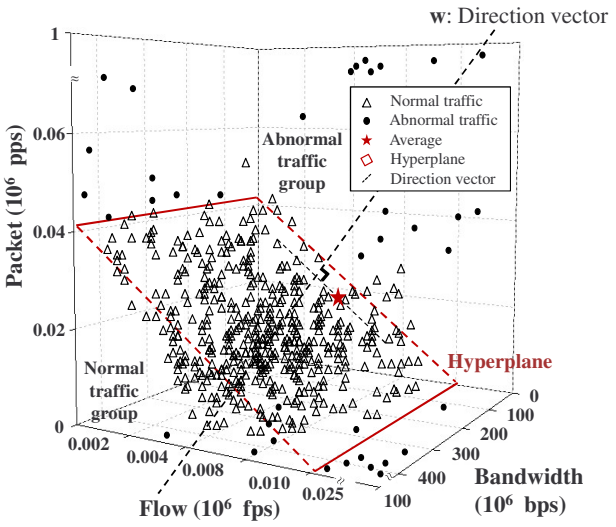

Fig. 6. Direction vector, hyperplane, the two groups, and traffic vectors.

$\left|\mathbf{S}_{intra}^{-1}\mathbf{S}_{inter} - \lambda \mathrm{I}\right| = 0$ to find the eigenvalues. From the equation, we obtained the three distinct eigenvalues and the corresponding eigenvectors as follows:

$$\lambda_1 = 0.1882e^{-002}, \lambda_2 = -0.1083e^{-14}, \text{ and } \lambda_3 = 0.7932e^{-008},$$
$$\mathbf{w}_1 = (-28.6341, -0.9995e^{-002}, -0.4927)^t,$$
$$\mathbf{w}_2 = (0.2597e^{-003}, -0.3358e^{-002}, 9.8352)^t, \text{ and}$$
$$\mathbf{w}_3 = (995.6493, -0.1654e^{-002}, -0.1577e^{-001})^t.$$

Furthermore, since the direction vector is an eigenvector corresponding to the maximal eigenvalue, the direction vector $\mathbf{w}$ is $\mathbf{w}_1$. Therefore, we derive a hyperplane which is a plane perpendicular to $\mathbf{w}_1$ and passing through the point μ_{total}. In fact, equation of the hyperplane is given by

$$28.63(x - 103,709,872.7) + 0.99e^{-002}(y - 5,247.040) + 0.49(z - 18,336.008) = 0. \quad (2)$$

Fig. 6 shows the direction vector $\mathbf{w}$, the hyperplane, and the two groups as well as some of the traffic vectors. In the figure, the spotted traffic vectors are 10% of randomly chosen traffic from each original training group. It is obvious that the area under hyperplane is normal group while area above the hyperplane is abnormal group.

Now we investigate the reliability of the derived hyperplane with a special day, July 2nd, 2008. Since the grade of Spring

abnormal traffic pattern	abnormal traffic type	dangerous level	amount of traffic	duration	interface of backbone router in Korea univ. network
	Jump (TrafficPps)	Critical	403.958	2008-07-02 21:37:00~ 2008-07-02 21:37:00	163.152.16.17 - HanaPark_16.17 (HanaPark_16.17) 163.152.16.47 - Vl217
	Jump (TrafficPps)	Critical	1160.620	2008-07-02 21:14:00~ 2008-07-02 21:35:00	163.152.16.17 - HanaPark_16.17 (HanaPark_16.17) 163.152.16.47 - Vl217
	Jump (TrafficPps)	Critical	901.967	2008-07-02 21:14:00~ 2008-07-02 21:35:00	163.152.16.17 - HanaPark_16.17 (HanaPark_16.17) 163.152.16.17 - Vl216
	Jump (TrafficPps)	Critical	1402.500	2008-07-02 20:12:00~ 2008-07-02 21:11:00	163.152.16.17 - HanaPark_16.17 (HanaPark_16.17) 163.152.16.17 - Vl216
	Jump (TrafficPps)	Critical	1200.720	2008-07-02 20:12:00~ 2008-07-02 21:11:00	163.152.16.17 - HanaPark_16.17 (HanaPark_16.17) 163.152.16.47 - Vl217
	Profile (HighPps)	Critical	1.696	2008-07-02 21:10:00~ 2008-07-02 21:10:00	163.152.16.17 - HanaPark_16.17 (HanaPark_16.17) 163.152.16.47 - Vl217

Fig. 7. A part of abnormalities provided by the University.

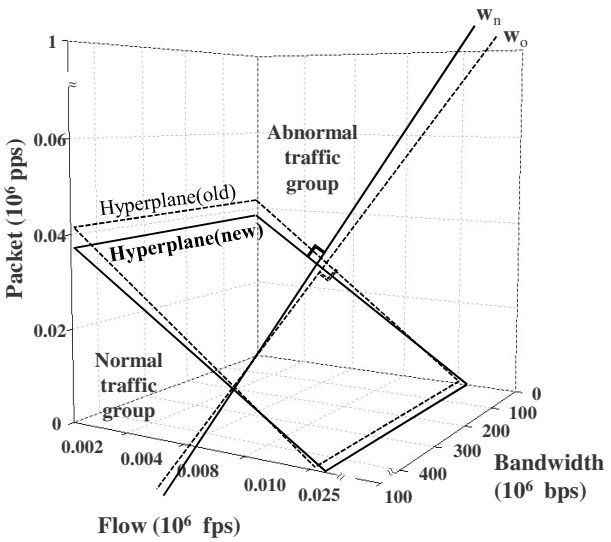

Fig. 8. Comparison of the original hyperplane and the updated hyperplane.

semester is announced at the day through the University portal system (http://portal.korea.ac.kr), heavy traffic which may considered as abnormal is generated at the day. Actually, the computer users who want to join the portal system are suffered from jamming network situation.

Traffic monitoring through Cisco 6509 backbone router also shows heavy traffic and many abnormality of network. Fig. 7 illustrates a part of table that verifies abnormality provided by the University. It shows the anomaly traffic pattern, type, and occurring point. The special day has the jump traffic type of irregular pattern and rapidly increasing pattern.

On the special day, it turns out that 108 traffic out of total 288 traffic belong to normal group while 180 traffic belong to abnormal group based on the derived hyperplane. Based on the information described in Fig. 7, 103 traffic as normal while 185 traffic as abnormal. By investigating the 288 traffic and comparing them with the information from the University, 101 traffic among 108 normal traffic lie below the hyperplane while 7 normal traffic lie above the hyperplane. In addition, 179 traffic among 180 abnormal traffic lie above the hyperplane while 1 abnormal traffic lies below the hyperplane. Table III summaries the detection information of the special day including miss detection probability.

Furthermore, we classified the scanning attack that can be represented as (src IP, dst port, dst IP, packet length) among 185 abnormal traffic. A host scanning and a port scanning are found to be 2 and 6, respectively. In other words, the scanning attack occupies 4.3% of abnormal traffic in our experimental environment. The remaining set of abnormal traffic can be categorized into another malicious traffic classes by defining traffic vectors differently, which describes the characteristics of the classes. Since the proposed detection mechanism is reliable up to 97%, we may conclude that the categorization of abnormal traffic is also reliable.

In [26], various detection mechanisms are compared for detecting typical attacks such as DoS by using DARPA '98 [27]. They obtained approximately 70∼85% of detection rate. Even though the used traffic sets are different so that the comparison of the results may inadequate in this stage of work, the detection rate of our mechanism is relatively high compared to those results.

In [28], the authors used a traffic set that is as similar as ours for anomaly detection. They collected the campus network traffic from Tunisian National University for forty five days and developed a Anomaly Detection System (ADS). Even though the experimental data set is different from ours, it is observed that the trends of the traffic from two universities are similar. In their experiment, the detection rate turns out to be 90% which is 7% less than our result.

One experimental study reports that the Sasser worm located in a vulnerable PC disables the machine in less than five minutes from when the machine was connected to the Internet [29]. Another frequently cited example is the Slammer worm, which is the fastest computer worm in history. It is known that the Slammer worm broke into the majority of vulnerable hosts on the Internet in less than ten minutes, congested many networks, and then left many hosts infected [30]. Based on these previously observed phenomena, using five minutes as update period seems to relevant to detect such worms. However, five minutes update period could cause severe overhead for some network situations and therefore it may not be the optimal value for update period. Since finding the optimal value of the update period is beyond the scope of this paper, we used five minutes update period in our experiment. We observed that the change of the hyperplane with the update period is negligible during ordinary time and day.

To compare the original hyperplane and the updated hyperplane, we used one week as the update period. Fig. 8 compares the two hyperplanes: old and updated. The updated hyperplane is obtained by replacing the traffic during May 14, 2008∼May 20, 2008 to the traffic during June 13, 2008∼June 19, 2008. As shown in the figure, the two hyperplane have similar trend with

TABLE III
SUMMARY DETECTION INFORMATION.

Categories	Normal group	Abnormal group
Information provided by the univ.	103	185
Using the hyperplane	108	180
Observed results	TP: 101, FP: 7, TN: 179, FN: 1	
Probabilities	$P_{cor} = 0.972$, $P_{mis} = 0.028$	

a slight difference of slopes. It is obvious that the difference of slopes reflects the change of traffic trend.

VI. CONCLUSION

In this paper, we proposed a traffic anomaly detection mechanism. For the proposed mechanism, we developed a novel analytical monitoring system and used Fisher linear discriminant method. Numerical results demonstrate that the proposed mechanism is excellent tool for detecting the abnormal traffic and provides criterions for anomaly detection and traffic trend of campus network. The most crucial advantage of the mechanism is it enables to detect abnormal traffic in real-time. In other words, by using the hyperplane which is derived based on a set of traffic generated in past, the newly generated traffic can be examined without any delay. Moreover, the mechanism includes self-learning algorithm which reflects the trend of traffic by updating the training set periodically. Obviously, periodical update of criterion increases accuracy of detection. We observed that it is possible to classify abnormal traffic more specifically by using different traffic vectors which characterize the specific attacks. In other words, it is possible to distinguish abnormal traffic into more specific groups such as network operation anomaly group, network attack group, and network abuse anomaly group. Therefore, future work will contain the subdivision of abnormal traffic into more specific groups which may require adding new components and/or considering totally new components in a traffic vector. In addition, experimental environment will be extended to cover heavy abnormal traffic including DARPA data set.

REFERENCES

[1] M. Thottan and C. Ji, "Anomaly Detection in IP Networks," *IEEE Transaction on Signal Processing*, vol. 52, no. 8, pp. 2191-2204, August 2003.

[2] A. Lakhina, M. Crovella and C. Diot, "Diagnosing Network-Wide Traffic Anomalies," in *Proc. ACM SIGCOMM 2004*, pp. 219-230, September 2004.

[3] S. H. Lee, H. J. Kim, J. C. Na, and J. S. Jang, "Abnormal Traffic Detection and Its Implementation," in *Proc. IEEE ICACT 2005*, pp. 246-250, February 2005.

[4] S. S. Kim and A. L. N. Reddy, "Statistical Techniques for Detecting Traffic Anomalies Through Packet Header Data," *IEEE/ACM Transaction on Networking*, vol. 16, no. 3, pp. 562-575, January 2008.

[5] R. Ahmed and R. Boutaba, "Distributed Pattern Matching: A Key to Flexible and Efficient P2P Search," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 1, pp. 73-83, January 2007.

[6] M. Roesch, "Snort-lightweight intrusion detection for networks," in *Proc. USENIX LISA 1999*, pp.229-238, November 1999.

[7] V. Paxson, "Bro: A System for Detecting Network Intruders in Real-Time," in *Proc. USENIX Security Symposium*, January 1998.

[8] G. Androulidakis and S. Papavassiliou, "Intelligent Flow-Based Sampling for Effective Network Anomaly Detection," in *Proc. IEEE GLOBECOM 2007*, pp. 1948-1953, November 2007.

[9] R. Sekar, M. Bendre, D. Dhurjati, and P. Bollineni, "A Fast Automaton-based Method for Detecting Anomalous Program Behaviors," in *Proc. IEEE Symposium on Security and Privacy*, May 2001.

[10] CAIDA Traffic measurement tool (CoralReef). [Online]. Available: http://www.caida.org/tools/measurement/coralreef/index.xml.

[11] M. S. Kim, Y. J. Won, and James W. Hong, "Characteristic analysis of internet traffic from the perspective of flows," *Computer Communications*, vol. 29, Issue 10, pp. 1639-1652, June 2006.

[12] A. Sridharan and T. Ye, "Implementing Real Time Port Scan Detection for the IP Backbone," in *Proc. LSAD workshop with SIGCOMM 2007*, Kyoto, Japan, August 2007.

[13] H. Hajji, "Statistical analysis of network traffic for adaptive faults detection," *IEEE Transactions on Neural Networks*, vol. 16, pp. 1053-1063, September 2005.

[14] Y. Zhang, Z. Ge, A. Greenberg, and M. Roughan, "Network Anomography," in *Proc. USENIX IMC 2005*, pp. 317-330, October 2005.

[15] Y. W. Chen, "Traffic behavior analysis and modeling of sub-networks," *International Journal of Network Management*, vol. 12, pp. 323-330, September 2002.

[16] H. Moayedi and M. Masnadi-Shirazi, "Arima model for network traffic prediction and anomaly detection," in *Proc. ITSim 2008*, pp. 1-6, August 2008.

[17] L. Huang, X. Nguyen, M. Garofalakis, M. Jordan, A. Joseph, and N. Taft, "In-Network PCA and Anomaly Detection," in *Proc. NIPS 2007*, pp. 617-624, December 2007.

[18] A. Lakhina, M. Crovella, and C. Diot, "Mining Anomalies Using Traffic Feature Distributions," in *Proc. ACM SIGCOMM 2005*, pp. 217-228, August 2005.

[19] H. Ringberg, A. Soule, J. Rexford, and C. Diot, "Sensitivity of PCA for Traffic Anomaly Detection," in *Proc. ACM SIGMETICS 2007*, pp. 109-120, June 2007.

[20] M. V. Mahoney, "Network Traffic Anomaly Detection Based on Packet Bytes," in *Proc. ACM symposium on Applied computing*, March 2003.

[21] A. Shashua, "On the Relationship Between the Support Vector Machine for Classification and Sparsified Fisher's Linear Discriminant," *Neural Processing Letters*, vol. 9, pp. 129-139, April 1999.

[22] R. Johnson, and D. Wichern, *Applied Multivariate Statistical Analysis*, 6th ed., Prentice-Hall, 2007, pp. 576-593, 623-633.

[23] K. Trivedi, *Probability and Statistics with Reliability, Queuing, and Computer Science Applications*, 2nd ed., John Wiley and Sons, 2002, pp. 658-664.

[24] P. Barford and D. Plonka, "Characteristics of Network Traffic Flow Anomalies," in *Proc. ACM SIGCOMM 2001*, pp. 69-73, August 2001.

[25] H. Choi and H. J. Lee, "PCAV: Internet Attack Visualization on Parallel Coordinates," in *Proc. ICICS 2005*, December 2005.

[26] A. Lazarevic, L. Ertoz, V. Kumar, A. Ozgur, and J. Srivastava, "A Comparative Study of Anomaly Detection Schemes in Network Intrusion Detection," in *Proc. SIAM International Conference on Data Mining 2003*, April 2003.

[27] MIT Lincon Laboratory, DARPA Intrusion Detection Data Sets, [Online]. Available: http://www.ll.mit.edu/mission/communications/ist/corpora/ideval/data/index.html.

[28] K. H. Ramah1, H. Ayari, and F. Kamoun, "Traffic Anomaly Detection and Characterization in the Tunisian National University Network," in *Proc. NETWORKING 2006*, February, 2006.

[29] USA TODAY. Unprotected PCs can be Hijacked in Minutes. [Online]. Available: http://www.usatoday.com/money/industries/technology.

[30] D. Moore, V. Paxson, S. Savage, C. Shannon, S. Staniford, and N. Weaver, "Inside the Slammer worm," *IEEE Security and Privacy Magazine*, vol. 1, pp.33-39, July 2003.