Transformation Approach to Model Online Gaming Traffic

KwangSik Shin, Jinhyuk Kim, Kangmin Sohn, Chang Joon Park, and SangBang Choi

In this paper, we propose a transformation scheme used to analyze online gaming traffic properties and develop a traffic model. We analyze the packet size and the inter departure time distributions of a popular first-person shooter game (Left 4 Dead) and a massively multiplayer online role-playing game (World of Warcraft) in order to compare them to the existing scheme. Recent online gaming traffic is erratically distributed, so it is very difficult to analyze. Therefore, our research focuses on a transformation scheme to obtain new smooth patterns from a messy dataset. It extracts relatively heavy-weighted density data and then transforms them into a corresponding dataset domain to obtain a simplified graph. We compare the analytical model histogram, the chi-square statistic, and the quantile-quantile plot of the proposed scheme to an existing scheme. The results show that the proposed scheme demonstrates a good fit in all parts. The chi-square statistic of our scheme for the Left 4 Dead packet size distribution is less than one ninth of the existing one when dealing with erratic traffic.

Keywords: Gaming traffic, traffic analysis, analytical model, curve fitting, transformation approach.

Manuscript received Mar. 12, 2010; revised Sept. 30, 2010; accepted Oct. 25, 2010.

This work was supported by the Industrial Strategic Technology Development Program (KI002095, Online Game Quality Assurance Technology Development using Scenario Control of Massive Virtual User) funded by the Ministry of Knowledge Economy (MKE), Rep. of Korea.

KwangSik Shin (phone: +82 42 860 1045, email: kwangsik@etri.re.kr), Kangmin Sohn (email: sogarian@etri.re.kr), and Chang Joon Park (email: chjpark@etri.re.kr) are with the Contents Research Division, ETRI, Daejeon, Rep. of Korea.

Jinhyuk Kim (email: graybaily@inha.edu) and SangBang Choi (email: sangbang@inha.ac.kr) are with the Department of Electronic Engineering, Inha University, Incheon, Rep. of Korea.

doi:10.4218/etrij.11.1510.0087

I. Introduction

With the recent global explosion of online multiplayer gaming, it is becoming more important to understand its effect on network behavior and usage; games are expected to be major contributors to web traffic. Interactive game traffic will be much different from other conventional web traffic that prevails on the Internet today. In general, it tends to employ small, highly-periodic packets. It is small since the application requires extremely low latencies and is highly periodic as a result of a game's dynamic requirement of frequent, predictable state updates among clients and servers [1].

There have been many studies on packet size and inter-departure time (IDT) as they apply to the analysis and modeling of online gaming traffic [1]-[7]. Many researchers have analyzed online game traffic from these perspectives and have tried to find suitable statistical traffic models for packet IDT and size. Existing studies have not taken the modeling packet size distribution of gaming traffic seriously because it is quite deterministic [1], [6], [8]. However, the gaming traffic of recently released games has no simple model done by deterministic or well-known distribution functions. It is especially difficult to represent a packet size traffic model as found in certain well-known distributions since it depends on the game contents, whereas the IDT is modeled as well as ever, due to its traffic distribution being relatively smooth.

A number of researchers have studied analytical models for online gaming traffic. The traffic of some games was simply modeled with a single distribution function by early researchers. However, these models show a considerably high discrepancy between the analytical model represented by a single distribution function and the actual characteristic of the gaming traffic. Therefore, a composition method [7] (the same as the technique of splitting distributions [6]), which characterized gaming traffic as combination of several distribution functions, was proposed. This method divides a traffic distribution into certain parts according to their local behavior and then matches each part to a corresponding single distribution function. It is possible to split a distribution into as many parts as necessary, but more than two or three separations result in a cumbersome analytical model. In addition, repetitive separations of traffic distributions with a "spiky" dataset are impossible to perform without the intervention of a person since the dataset division depends only on the researcher's discretion without any algorithmic criterion. Through analyzing a certain amount of traffic for up-to-date recently released games, we found that it is difficult to separate the traffic of these games into only a few parts with well-known distributions. In other words, for erratic gaming traffic, the existing methods are not suitable for the development of an analytical model because some portions of the gaming traffic as well as the traffic in its entirety do not follow any established pattern. Thus, we decided to adopt a novel approach which divides the full dataset into several classified groups according to a filtering rule that requires each group have a patterned distribution in the filtered dataset domain. The proposed transformation scheme actively reconstitutes the erratic traffic into several patterned datasets using dataset domain transformation, whereas the existing composition method focuses on discovering a partial area with a typical pattern. Since, unlike the existing method, ours defines a filtering rule, a repetitive dataset separation is easily performed.

The rest of this paper is organized as follows. Section II provides a brief account of the characteristics of online gaming traffic, along with previous works on online gaming traffic modeling. Section III compares the online differences in gaming traffic between earlier and more recent online games. Section IV presents the results of a performance comparison with an existing modeling scheme. Finally, section V draws some conclusions.

II. Online Gaming Traffic Characteristics and Previous Models

With the current growing rate of multiplayer network games, such as first-person shooter (FPS) games and massively multiplayer online role-playing games (MMORPGs), networking researchers have been struggling to understand their network requirements through analysis of their traffic profiles [9], [10].

FPS games, such as Quake [11], provide large-scale gaming, and sometimes team-based combat, in a real-time virtual environment. This is done through games utilizing the character's point of view, transmitting data from the client to the server, and then immediately processing the data received at that time. Quake is one of the most popular multiplayer games found on the Internet; it imposes the hardest real-time requirements on a network. Because it is very sensitive to interactivity, this type of game requires low-latency point-to-point communication as well as directed broadcast channels in order to facilitate its real-time game logic. Therefore, packets are sent via UDP since clients need to send packets at intervals that are much shorter than the time it would take to retransmit lost packets [1].

MMORPGs, such as World of Warcraft (WoW) [12], are a genre of computer role-playing games in which a very large number of players interact with one another within a virtual world [2]. They focus on an accurate execution of the client inputs, which has an impact on the transport protocol used. In WoW and many other MMORPGs, TCP serves as the transport protocol. TCP is connection-oriented and offers reliability. This attribute is well suited for MMORPGs, preventing error propagation during long sessions. The round-trip time can be reduced by using small packets, even though it uses a relatively slow TCP connection.

While there are some differences between game genres, online gaming traffic generally tends to employ small, highly-periodic packets. An in-depth knowledge of their traffic behavior will certainly assist network game developers and publishers to provide better design and service for online games. There are many studies being conducted to present traffic measurements of popular network games and to provide their traffic characterization.

Some researchers have developed a traffic model for packet IDT and size to evaluate the impact of the loss rate and delay on the client usability [3]-[5]. They used a single probability density function to characterize their traffic model for Counter-Strike [13], Half-Life [14], and Quake 3 [11]. However, this model cannot fully characterize network gaming traffic, although it is easy to show a rough trend with a reasonable execution speed of the simulation model.

There have been several studies that classify the datasets into a few parts to characterize separate distribution functions [6], [7]. Some have proposed source models for the popular FPS game, Quake, and other researchers have studied Second Life, the most popular virtual world, where the IDTs and packet sizes were modeled as having extreme, exponential, deterministic, or other types of distributions. A quantile-quantile (Q-Q) plot is used to compare the discrepancies between two probability distributions. They designed the dataset with a split distribution that models one part using one type of distribution and the rest with another. It is possible to split the distribution as many times as necessary, but more than two or three separations result in a cumbersome analytical model, and it is therefore difficult to apply to messy data.

III. Analytical Model for Online Gaming Traffic

In this section, we introduce statistical distributions of recent online gaming traffic and our proposed traffic modeling method. Unlike the web or other conventional Internet traffic, most online games have delay-sensitive and low-bitrate requirements due to their intrinsic characteristic of sending small packets frequently. Therefore, we tracked a lot of network traffic from several popular online games since we know online gaming traffic has different characteristics according to the type, as mentioned in the previous section. This research is a preliminary study used to develop an online gaming traffic generator to test the reliability of an online game server. Since the accuracy of the gaming traffic generator depends on the accuracy of the online gaming traffic analytical model, this study is fundamental research used to improve the test quality of the online game server. However, we only focused on the outbound data packets from the clients, except for TCP ACK, because we are interested in getting the analytical model for the client's gaming traffic.

1. Erratic Distribution of Recent Online Gaming Traffic

Previous works argued that the game traffic has a large affect in determining the characteristics in regards to packet size. Therefore, that kind of traffic can be analyzed without much effort. For the Quake series (1, 2, and 3) and the other popular games analyzed in previous works, it was not difficult to analyze the traffic. However, the packet structure and the game protocols of recent games have become much more complicated to keep pace with gamers' expectations. Thus, the traffic of recent games is not able to be modeled easily by specific distributions anymore. In order to demonstrate the variations of gaming traffic distributions, we statistically analyzed these popular games: Quake 3, WoW, Quake 4 [11], and Left 4 Dead (L4D) [15]. Figure 1 shows the packet size distributions of the analyzed gaming traffic. For Quake 3, which is an older game, every packet is located at one of a few peaks. In other words, since most packets are represented by one of only several peaks, it can be well-modeled as deterministic. However, since the traffic of recently released games has so many "spiky" datasets and does not follow any specific patterns, it does not seem to fit into any well-known analytical deterministic distributions. For example, even though Quake 4 is a sequel of Quake 3, its traffic distribution is quite different and considerably intricate. In addition, L4D, which is the latest game among these examples, is far more erratic than the others. Therefore, from these figures, we know that recent gaming traffic has become more and more erratic and is too irregular to design an analytical model for it using the existing methods.

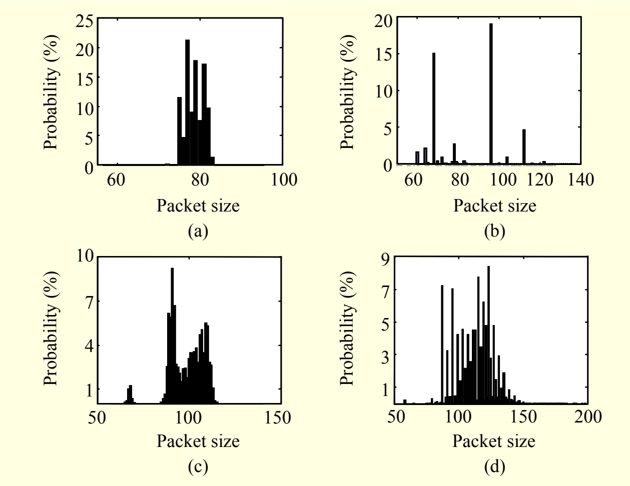

Fig. 1. Packet size distribution of gaming traffic according to release year: (a) Quake 3 (1999), (b) WoW (2004), (c) Quake 4 (2005), and (d) L4D (2008).

2. Proposed Modeling Scheme for Online Gaming Traffic

As previously mentioned, it is very difficult to analyze recent online gaming traffic. Thus, in advance of the analysis, online gaming traffic needs to be transformed into new datasets which are feasible to analyze. In this paper, we propose an analytical model designing method for gaming traffic. It is a four-step process: statistical analysis of measured packets, transformation of packet distribution, curve fitting, and traffic regeneration. These are shown in Fig. 2. The first step is similar to conventional statistical analysis, and the third one is achieved by iterative work based on some fitting algorithms. Therefore, the second and fourth are the key procedures of the proposed scheme. We particularly focus on the transformation of packet distribution since it is the most critical of the methods used to simplify a complex dataset to derive an analytical model.

The statistical analysis of measured packets consists of the traffic capture, statistical analysis of captured packets, and distribution of analyzed packets modules. The first module collects the online gaming packets. After tabulating the captured packets, we graphically display them as a result of the second module. We draw a probability density function (pdf) from the statistically analyzed packets in the remaining module since we analyze the gaming traffic using a graphical approach. Often, after finishing the first stage, the measured pdf shows too many spikes to be modeled by the common distributions

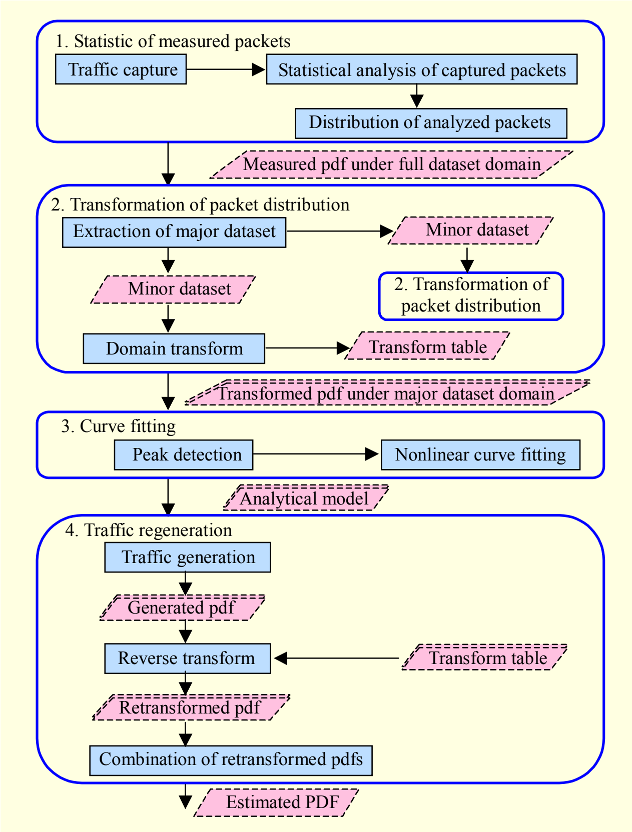

Fig. 2. Process used for modeling online gaming traffic.

because the recent online gaming traffic is so erratic. Therefore, it is required to convert it into a more simplified shape that can be mapped into an analytical model.

In the second stage, in order to obtain a simplified shape, we extract a major dataset, which is a set of the relatively heavy-weighted density data, and then transform the packet distribution from the full dataset domain into the major dataset domain. A moving average scheme is applied for the extraction of a major dataset. A moving average is not a single number, but a set of numbers, each of which is the average of the corresponding subset of a larger set of data points. When a value is bigger than or equal to the average of the corresponding subset, it is classified as a major dataset point. When it is not, it belongs to the minor dataset and becomes an input of the second procedure again. This routine is repeated until the proportion of minor datasets is less than a chosen threshold since the transformation of packet distribution is a recursive procedure. The stricter the threshold we choose, the more complicated the modeling process becomes, and the closer the analytical model becomes to the measured pdf, which is the result of the distribution of analyzed packets. After obtaining the major dataset, it is fully changed into a transformed pdf under the major dataset domain, and the transform table is generated for the reverse transform in the traffic regeneration procedure.

Even though the transformed pdf under the major dataset domain is much smoother than the distribution under the full dataset domain, it still has multiple peaks. We know that a multiple peak curve is not easily modeled on any general distribution function. Therefore, we resolve the problem by modeling the transformed pdf with an aggregate of multiple distribution functions. At the third stage, we specify the number of peaks in the peak detection and then fit the corresponding distributions around each peak using the nonlinear curve fitting. When the curve fitting procedure is over, we obtain an analytical model for each transformed pdf.

At the last stage, a generated pdf is produced from the analytical models in the traffic generation module. However, gaming packets cannot be immediately generated from a generated pdf since the distribution domain of the generated pdf is inconsistent with the original gaming traffic distribution. Therefore, the distribution domain is converted into a full dataset domain through reverse transform according to the transform table. Finally, we obtain an estimated pdf for gaming packet generation by combining the retransformed pdfs.

3. Example of Analytical Modeling

In this paper, as other researchers have done, we analyzed the packet size and IDT for various games of different genres since online gaming traffic has different characteristics according to the type of game. Among them, we chose two recent popular online FPS games, L4D and WoW, for use in designing the analytical model. We introduce a modeling example for the packet size distribution of L4D, which is the most complex distribution among all of the traffic properties we analyzed. We utilized the MATLAB [16] system over the whole modeling process except during the packet capture.

First, we captured L4D gaming packets and then tabulated them according to their packet size. Figure 3(a) shows a histogram that illustrates the ratio of the packet appearance to the packet size. In this example, the histogram displays the average number of packets that are captured 10 times. We redrew the normalized density of the traffic distribution from the histogram because we analyzed the gaming traffic using a graphical approach. Figure 3(b) illustrates the probability density function. This is an example of the output of the statistic of measured packet procedure used in Fig. 2.

For the second step, we performed a dataset extraction and transformation to obtain a simplified distribution. We applied a weighted moving average scheme to Fig. 3(b) to discriminate the heavy-weighted density data points. The result is shown in Fig. 3(c). A moving average scheme is commonly used with time series data to smooth out the short-term fluctuations

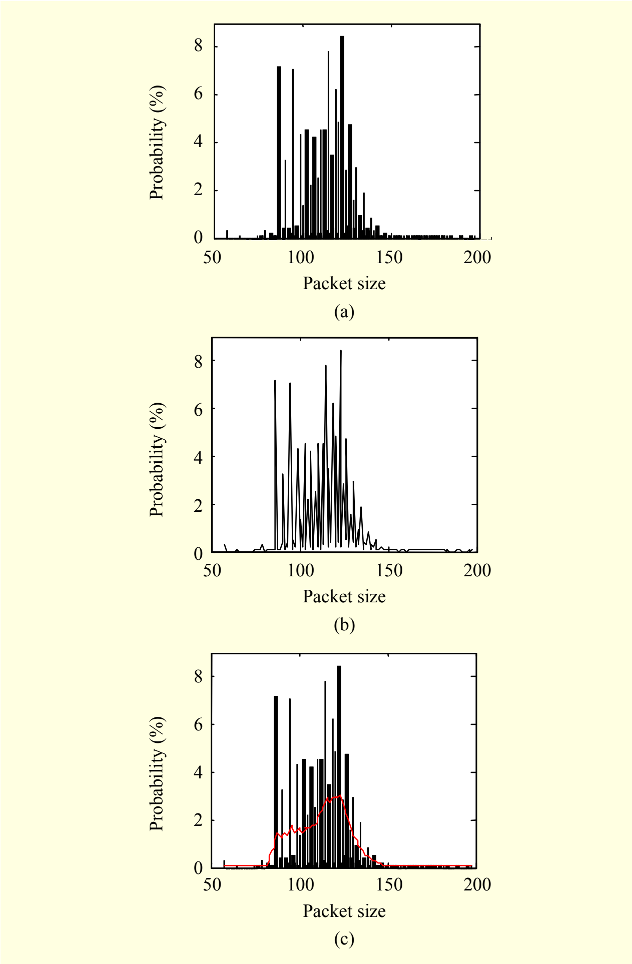

Fig. 3. Traffic distribution according to L4D packet size: (a) histogram, (b) pdf, and (c) moving average (red line).

and highlight the longer-term trends or cycles. It is similar to a low-pass filter used in signal processing to remove noise, represented by fast moving peaks. However, in the figure, the peaks are not noise; they are critical values as they symbolize the more frequent packet sizes compared to the other nearby ones. For the packet size of L4D, we fixed the window size in the weighted moving average scheme to 11 points; the number of the dataset points is 150, ranging from 50 to 200. A weighted average is any average that has multiplying factors to give different weights to different data points. We put more weight on the center of window than on the outside. The moving average value in each data point is calculated by

$$M_t = \frac{W_{\lfloor \frac{n}{2} \rfloor} \times W_{t-\lfloor \frac{n}{2} \rfloor} + \cdots W_0 \times P_t + \cdots + W_{\lfloor \frac{n}{2} \rfloor} \times W_{t+\lfloor \frac{n}{2} \rfloor}}{n \times \sum W_n}, \qquad (1)$$

$$W_0 = \left\lceil \frac{n}{2} \right\rceil, \quad W_{t+1} = W_t - 1,$$

where M_t is the moving average value in the t-th data point, n is the window size, P_t is the value in the t-th data point, and W is the weighting factor.

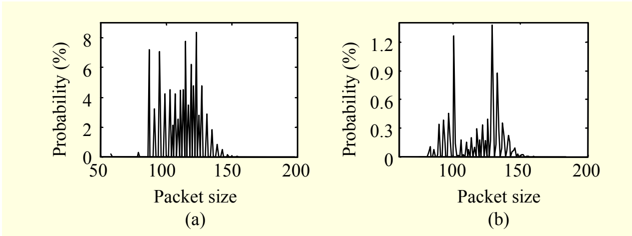

Fig. 4. Distributions classified by weighted moving average: (a) heavy-weighted density data and (b) light-weighted density data.

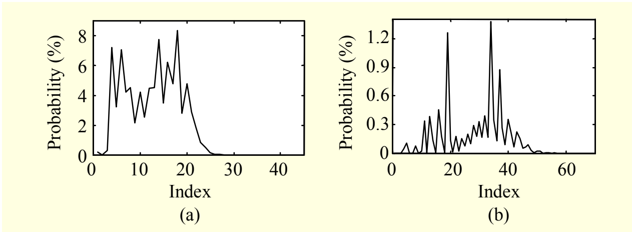

Fig. 5. Transformed pdfs after performing transformation of packet distribution for (a) heavy-weighted density data and (b) light-weighted density data.

We divided the dataset into major and minor parts according to the result of the comparison with the weighted moving average represented by the red line in Fig. 3(c). If P_t is bigger than M_t, the t-th data becomes a member of the major dataset; if not, it becomes a member of the minor dataset. Figures 4(a) and 4(b) show the distributions of the major dataset and the minor dataset, respectively. The transformation of the packet distribution procedure was performed only once in this example, even though it is a recursive procedure. The major dataset and minor dataset were thereby identified as the heavy-weighted density data and the light-weighted density data for the major dataset and the domain transform, respectively.

However, there are a number of spiky shapes in the figures due to the many invalid data points that have probability densities of zero. These types of graphs are difficult to fit into any distribution due to their erratic appearance. Thus, we obtained a new graph under another dataset domain using the domain transform module. Figures 5(a) and 5(b) show the transformed pdf redrawn for the heavy-weighted and light-weighted density datasets, respectively.

In the third step, we fit the transformed pdf into nonlinear curves. The aim of the nonlinear fitting is to estimate the parameter values that best describe the data. The standard way of finding the best fit is to choose the parameters that would minimize the deviations of the theoretical curve from the experimental points. This method is called chi-square minimization and is defined as

Table 1. Parameter definitions for chi-square minimization.

Parameters	Definitions
N	Number of data points
σ_i	Variance related to the measurement error for Y_i
Y_i	Observed mean
$f(x_i)$	Expected value

$$\chi^2 = \sum_{i=1}^{n} \left[Y_i - f\left(x_i^{'}, \theta\right) \right]^2, \tag{2}$$

where $x_i^{'}$ is the row vector for the i-th observation, i=1, 2,..., n, and θ are the parameters we need to compute. To estimate the θ value with the least square method, we need to solve the normal equations, which are set at zero for the partial derivatives of χ^2 with respect to each θ_p:

$$\frac{\partial \chi^2}{\partial \theta_p} = -2\sum_{i=1}^{n} \frac{1}{\sigma_i^2} \left[Y_i - f\left(x_i^{'}, \theta\right) \right] \left[\frac{\partial f\left(x_i^{'}, \theta\right)}{\partial \theta_p} \right] = 0, \tag{3}$$

$$\chi^2 = \sum_{N} \frac{1}{\sigma_i^2} \left[Y_i - f\left(x_i\right) \right]^2 \quad (0 < \chi^2 < \infty). \tag{4}$$

Since there are no explicit solutions to the normal equations, we employed an iterative strategy, the Levenberg-Marquardt algorithm, which combines the Gauss-Newton method and steepest descent method, to estimate the parameter values. This process starts with some initial values θ_0. With each iteration, a single χ^2 value is computed, and the parameter values are then adjusted so as to reduce the value. When the χ^2 values computed in two successive iterations are small enough (compared with the tolerance), the fitting procedure has converged [17]. In this example, for the nonlinear curve fitting, we used the Levenberg-Marquardt algorithm based on the Gaussian distribution function which is provided in MATLAB's curve fitting toolbox. It specifies the number of peaks using a second derivative method revealing peaks as the local minima and then fits them into the sum of each corresponding distribution. In this way, we designed the analytical model as a synthesis of the multiple approximated Gaussian distribution functions. Equations (5) and (6) are a Gaussian distribution function and its approximation.

$$y = y_0 + \sum \frac{a_i}{c_i \times \sqrt{\pi/2}} \times e^{-2((x-b_i)^2/c_i^2)}, \tag{5}$$

$$y = y_0 + \sum a_i^{'} \times e^{-((x-b_i)/c_i)^2}, \tag{6}$$

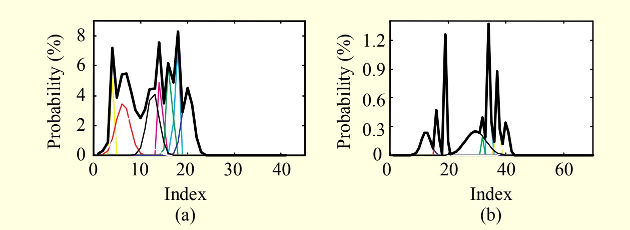

Fig. 6. Results of curve fitting for each transformed pdf: analytical models of (a) transformed heavy-weighted density data and (b) transformed light-weighted density data.

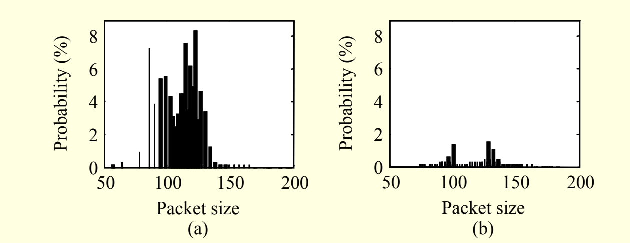

Fig. 7. Regenerated distribution derived from analytical model under full dataset domain: (a) heavy-weighted density data and (b) light-weighted density data.

where, $a_i^{'} = \frac{e^{-2}}{\sqrt{\pi/2}} \times \frac{a_i}{c_i}$.

For the L4D packet size distribution, the analytical model of the relatively heavy-weighted density data is defined as

$$\begin{aligned} y = {} & 0.3581 * e^{-((x-17.57)/0.3422)^2} + 0.9853 * e^{-((x-14.47)/0.2711)^2} \\ & + 0.05244 * e^{-((x-3.955)/0.434)^2} + 0.03492 * e^{-((x-6.289)/2.131)^2} \\ & + 0.06685 * e^{-((x-16.25)/0.754)^2} + 0.04549 * e^{-((x-20.11)/1.642)^2} \\ & + 0.02488 * e^{-((x-7.856)/3.827)^2} + 0.04214 * e^{-((x-12.68)/1.831)^2}. \end{aligned} \tag{7}$$

Figures 6(a) and 6(b) display the diagram forms of the modeled distribution functions for the heavy-weighted and light-weighted data. Each distribution of the multiple peaks and their sums are illustrated by the normal and bold lines in the figures, respectively.

In the final step, we regenerated the traffic distributions from previous analytical models. A generated pdf, which was derived from each analytical model, became a retransformed pdf according to the transform table through the reverse transform module. Figures 7(a) and 7(b) show the retransformed pdf for each dataset of the analytical model. In Fig. 8(a), the estimated pdf is shown as its aggregate. Finally, in Fig. 8(b), we show the difference between our estimated distribution and the measured traffic distribution. Even though there is some discrepancy, it can be reduced to smaller value by choosing a lower threshold.

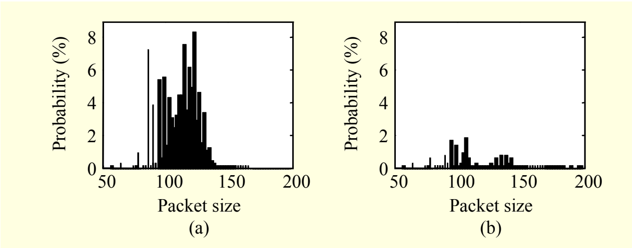

Fig. 8. Comparison between measured and estimated histograms: (a) estimated histogram and (b) difference between both distributions.

IV. Evaluation

In this section, we demonstrate the superiority of the proposed scheme (the transformation scheme) through a comparison to an existing one (the composition scheme). We analyzed the packet size and the IDT distributions of L4D and WoW because there are different traffic characteristics between FPS games and MMORPGs, as previously mentioned. To collect the analysis data, we captured gaming packets using Wireshark [18], which is a network protocol analyzer, during approximately 15 minutes for each game; each capture is iterated 10 times. We took the median values for designing the analytical model. The average number of packets captured for L4D and WoW were 26,216 and 4,441, respectively.

We converted the frequencies of each data point into a percentage, as shown in Figs. 9 through 12 and then analyzed the normalized values. In this section, we divided traffic into 2 or 3 distributions according to the number of repetitions for transformation of packet distribution procedure. For IDT, the procedure was only performed once, and then the major dataset and minor dataset were thereby identified as the heavy-weighted density and the light-weighted density. For packet size, meanwhile, it was performed one more time since packet size distribution is more complicated. The major dataset of the first repetition became the heavy-weighted density. After the second repetition, the major dataset and minor dataset were expressed as the medium-weighted density and the light-weighted density, respectively. We evaluated the two methods in three ways, as follows.

First, we show an intuitive comparison among the measured and estimated histograms in Figs. 9 through 12. The figures show that the packet size and the IDT of WoW and the IDT of L4D are relatively well modeled by the existing scheme, as well as in the proposed one, since they are modeled by simple distributions and deterministic. However, the measured histogram for the packet size of L4D is difficult to represent based on its composition of several distributions; these distributions are particularly complicated. Figures 12(b) and 12(c) show that the existing method only represents sketchy behaviors but does not regenerate them in detail, while our method simulates the measured distribution analogously.

Second, we show the quantitative values found in the results of our comparison between both schemes. Table 2 summarizes the complete analytical models and the chi-square statistics (χ^2) for each scheme. The chi-square test is widely used to determine the goodness of fit of a distribution to the set of the experimental data [19]; its values are calculated by adding together all the squares of differences between the density values from the estimated and measured histograms. In the table, the chi-square statistic (χ^2) of the proposed scheme is less than one ninth compared to the existing scheme for the L4D packet size distribution. Notably, our scheme does not exceed the threshold for the desired 5% significance level, unlike the existing scheme. The standard level of significance used to justify a claim of a statistically significant effect is 0.05 [20]. This means that the proposed scheme is more suitable than the old one for complicated traffic, although the existing method is also good at modeling deterministic traffic patterns. From this, we know that the accuracy of the fit of our scheme is much better than the older scheme when applied to erratic traffic.

Finally, we did a qualitative analysis. A chi-square value provides the degree of difference, but it does not provide a graphical view of how the properties, such as location, scale, and skew, are similar or different in the two distributions. Therefore, we display the Q-Q plots of each model for a more detailed comparison. This is a meaningful assessment in that the existing method found its best fit model using a Q-Q plot. It is commonly useful to examine whether a dataset accurately fits a distribution using the estimated parameters. On this plot, the corresponding quartiles of both the empirical and the analytical distributions are graphed against each other, so that the deviations may be easily identified. If the two distributions being compared are similar, the points in the Q-Q plot will approximately lie on the line $y=x$. Figures 13 through 16 show the Q-Q plots for the packet size and the IDT models of L4D and WoW. They show in detail which parts fit together accurately in the model distributions. Overall, the analytical model of the proposed method shows better tail characteristics. The Q-Q plots show a pronounced performance difference between the proposed and existing schemes. In Fig. 13(a), for the most complicated traffic, some parts of the analytical model do not fit well. Thus, from Figs. 13(a) and 13(b), we know that the analytical model using the existing method does not fit well in some areas, while our method shows a more accurate fit in all parts of the measured traffic distribution.

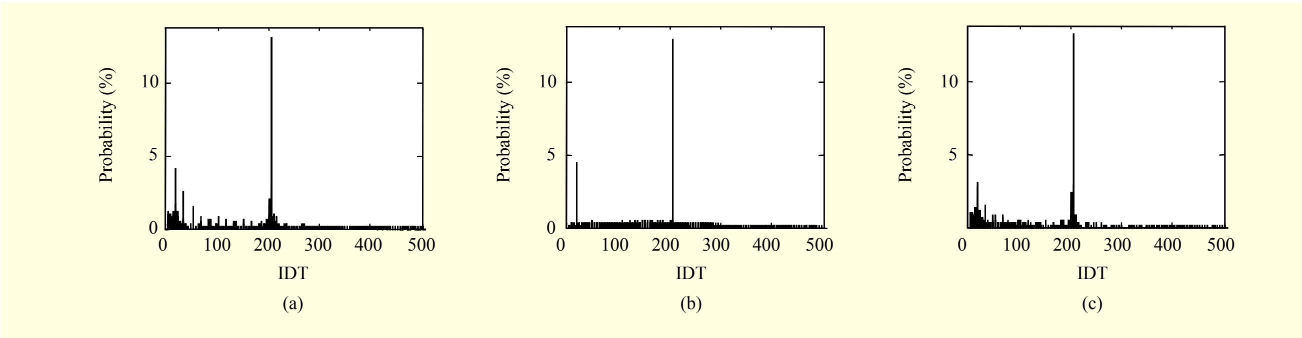

Fig. 9. WoW IDT histograms: (a) measured histogram, (b) estimated histogram of the composition scheme, and (c) estimated histogram of transformation scheme.

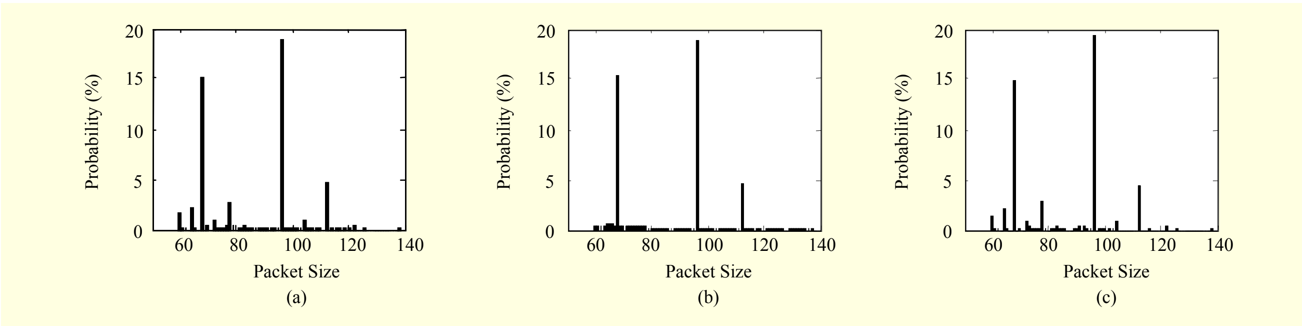

Fig. 10. WoW packet size histograms: (a) measured histogram, (b) estimated histogram of composition scheme, and (c) estimated histogram of transformation scheme.

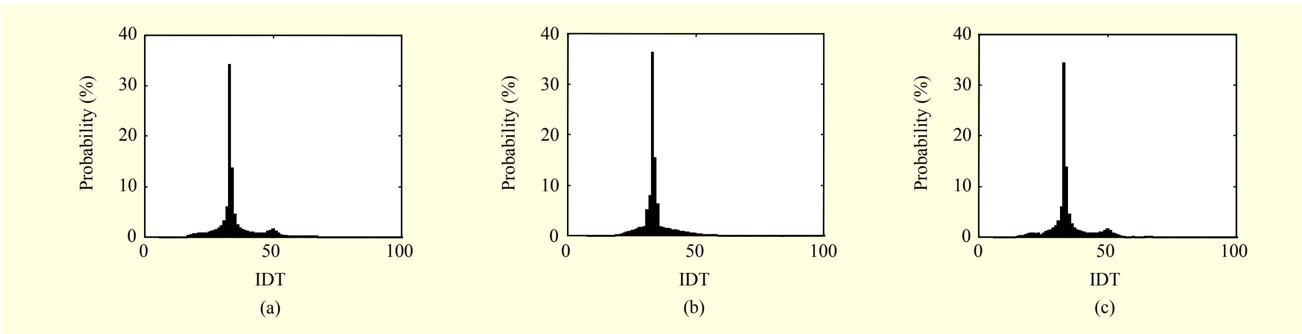

Fig. 11. L4D IDT histograms: (a) measured histogram, (b) estimated histogram of composition scheme, and (c) estimated histogram of transformation scheme.

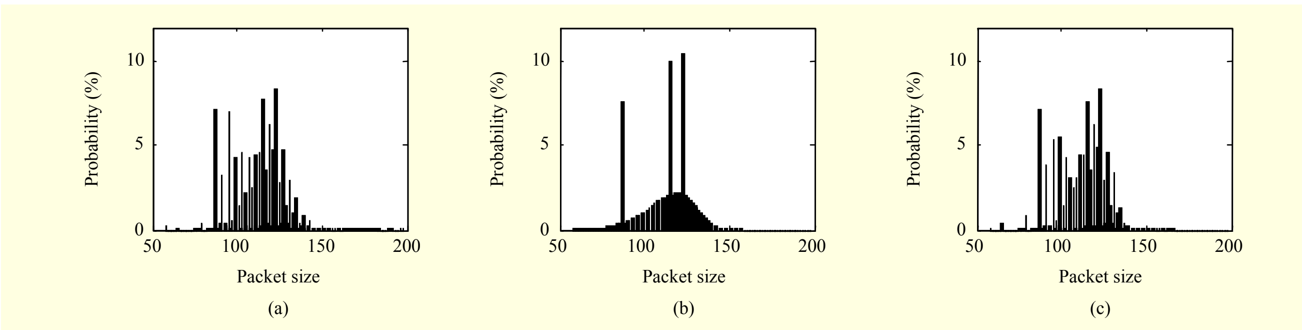

Fig. 12. L4D packet size histograms: (a) measured histogram, (b) estimated histogram of composition scheme, and (c) estimated histogram of transformation scheme.

Table 2. Chi-square statistic and proportion of fitting distributions between both methods for each traffic property.

Properties	Game	Scheme (no. of distributions)	Distribution	Chi-square statistic (χ^2)	Degree of freedom	Significance level (5%)
IDT	WoW	Composition (3)	Deterministic (17.29%) Deterministic (4.12%) Normal (78.59%)	30.43	*104	130
		Transformation (2)	Heavy-weighted density (50.5%) Light-weighted density (49.5%)	5.37		
	L4D	Composition (5)	Deterministic (3.20%) Deterministic (5.91%) Normal (47.9%) Deterministic (4.47%) Lognormal (38.53%)	9.20	95	120
		Transformation (2)	Heavy-weighted density (44.18%) Light-weighted density (55.82%)	3.23		
Packet size	WoW	Composition (4)	Deterministic (30.02%) Deterministic (37.92%) Deterministic (9.22%) Generalized extreme (22.84%)	78.62	52	71
		Transformation (3)	Heavy-weighted density (95.20%) Medium-weighted density (4.16%) Light-weighted density (0.64%)	44.40		
	L4D	Composition (4)	Deterministic (7.2%), Deterministic (7.75%) Deterministic (8.36%), Weibull (76.69%)	1289.65	112	138
		Transformation (3)	Heavy-weighted density (89.21%) Medium-weighted density (8.19%) Light-weighted density (2.78%)	134.64		

* Omitting infrequently occurring categories (less than 0.15%)

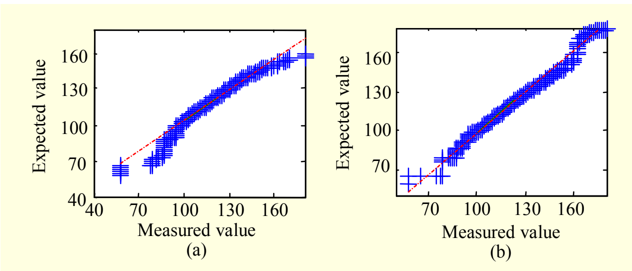

Fig. 13. Q-Q plots of L4D packet size models: (a) composition and (b) transformation schemes.

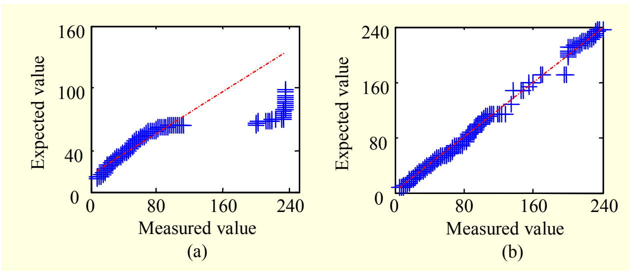

Fig. 15. Q-Q plots of L4D IDT models: (a) composition and (b) transformation schemes.

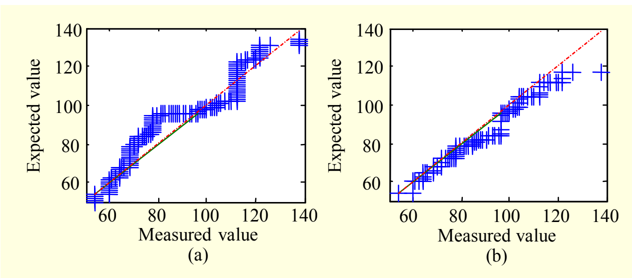

Fig. 14. Q-Q plots of WoW packet size models: (a) composition and (b) transformation schemes.

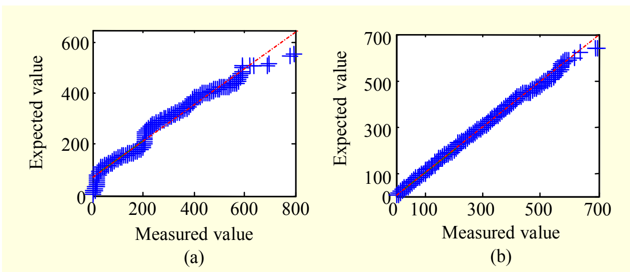

Fig. 16. Q-Q plots of WoW IDT models: (a) composition and (b) transformation schemes.

V. Conclusion

In this paper, we proposed a new modeling scheme for online gaming traffic. We described the characteristics found in recent online gaming traffic. The modeling is accomplished by successive statistic of measured packets, transformation of packet distribution, curve fitting, and traffic regeneration procedures. Our research focuses more on the transformation of packet distribution procedure since it plays the most critical role in simplifying a complex dataset used for deriving the analytical model.

We analyzed the packet size and the IDT distributions of L4D and WoW to compare our method with an existing scheme. We showed intuitive, quantitative, and qualitative comparisons to demonstrate the superiority of the proposed transformation scheme. The intuitive comparison for the packet size of L4D showed that the existing method results in sketchy behaviors and does not regenerate them in detail, while our method simulates the measured distribution analogously. The quantitative and qualitative comparisons clearly show a performance difference between the proposed and existing schemes. The chi-square statistic (χ^2) of the proposed method for most erratic traffic is less than one ninth of the old one. From these results, we know that the proposed transformation scheme demonstrates an accurate fit in all parts of the measured traffic distribution.

References

[1] W.C. Feng et al., "A Traffic Characterization of Popular On-line Games," *IEEE/ACM Trans. Networking*, vol. 13, no. 3, June 2005, pp. 488-499.

[2] B. Hariri, S. Shirmohammadi, and M.R. Pakravan, "A Hierarchical HMM Model for Online Gaming Traffic Patterns," *IEEE Int. Instrum. Meas. Tech. Conf.*, May 2008.

[3] J. Färber, "Network Game Traffic Modeling," *Proc. 1st Workshop Netw. Syst. Support for Games*, 2002, pp. 53-57.

[4] T. Lang et al., "A Synthetic Traffic Model for Half-Life," *Proc. Australian Netw. Telecommun. Conf.*, 2003.

[5] T. Lang, P. Branch, and G. Armitage, "A Synthetic Traffic Model for Quake3," *Proc. 2004 ACM SIGCHI Int. Conf. Adv. Computer Entertainment Tech.*, vol. 74, 2004, pp. 233-238.

[6] M.S. Borella, "Source Models of Network Game Traffic," *Computer Commun.*, vol. 23, no. 4, Feb. 2000, pp. 403-410.

[7] R. Antonello et al., "Traffic Analysis and Synthetic Models of Second Life," *Springer Multimedia Syst. J.*, vol. 15, no. 1, Feb. 2009, pp. 33-47.

[8] Y.T. Han and H.S. Park, "Game Traffic Classification Using Statistical Characteristics at the Transport Layer," *ETRI J.*, vol. 32, no. 1, Feb. 2010, pp. 22-32.

[9] P. Bettner and M. Terrano, "1500 Archers on a 28.8: Network Programming in Age of Empires and Beyond,"*Gamasutra*, Mar. 2001. http://www.gamasutra.com

[10] Y.S. Ng, "Designing Fast-Action Games for the Internet," *Gamasutra*, May 1997. http://www.gamasutra.com

[11] Quake. http://www.idsoftware.com/games/quake

[12] World of Warcraft. http://www.worldofwarcraft.com

[13] Counter Strike. http://www.counter-strike.net

[14] Half-Life. http://orange.half-life2.com/

[15] Left 4 Dead. http://www.l4d.com

[16] MATLAB. http://www.mathworks.com

[17] M. Ledvij, "Curve Fitting Made Easy," *The Ind. Physicist*, Apr. 2003, pp. 24-27.

[18] Wireshark. http://www.wireshark.org

[19] A. Leon-Garcia, "Probability and Random Processes for Eelectrical Engineering 2/E," Addison-Wesley, 1994.

[20] R.A. Fisher, "Statistical Methods for Research Workers," Edinburgh: Oliver and Boyd, 1925.

KwangSik Shin received the BS, MS, and PhD in electronic engineering from Inha University, Incheon, Korea, in 2001, 2003, and 2008, respectively. From 2008 to 2010, he worked for ETRI, Korea, as a post doc, and he is currently working as a senior researcher at ETRI, Korea. His research interests include online game technology, traffic modeling, multimedia communication, wireless communication, computer networks, and parallel and distributed computing.

Jinhyuk Kim received the BS in electronic engineering from Inha University, Incheon, Korea, in 2009. He is currently working on his MS at the same university. His recent interests include online game technology, multimedia communication, wireless communication, computer networks, and distributed computing.

Kangmin Sohn received the BS and MS in electronic engineering from Inha University, Incheon, Korea, in 1995 and 1997, respectively. From 1998 to 2001, he worked for Unitech, Seoul, Korea, as a researcher. He is currently working as a senior researcher at the ETRI, Korea. His recent interests include online game technology, online game QA, and artificial intelligence.

Chang Joon Park received the BS, MS, and PhD in electronic engineering from Kyungbook National University, Daegu, Korea, in 1994, 1996, and 2000, respectively. From 1998 to 2010, he worked for ETRI, Daejeon, Korea, as a senior member of engineering staff, and he is currently working as a head of the Smart Game Research Team of ETRI. His research interests include online game technology, online game QA, computer vision technology, and parallel computing.

SangBang Choi earned the MS and PhD in electrical engineering from the University of Washington, Seattle, in 1988 and 1990, respectively. He is currently a professor of electronic engineering at Inha University, Incheon, Korea. His research interests include computer architecture, computer networks, wireless communications, parallel and distributed systems, and online game technology. He is a member of the IEEE and IEEE Computer Society.