A User-Centric Framework for Comparing Applications' Network Robustness*

Hung-Hsuan Chen, Cheng-Chun Tu, and Kuan-Ta Chen
Institute of Information Science, Academia Sinica
sean@iis.sinica.edu.tw, williamt@iis.sinica.edu.tw, cychen@iis.sinica.edu.tw

ABSTRACT

In this paper, we compare real-time and interactive network applications in terms of *their ability to handle network errors*, i.e., network delay and loss. Our approach is based on session times, which indicate users' perceptions of the network performance. Using a regression model, we separate the effect of network error on users' departure decisions from an application's baseline departure rate, and define an index to quantify an application's network robustness. We demonstrate the effectiveness of our framework on real-life data traces.

Categories and Subject Descriptors: C.2.3 [Computer-Communication Networks]: Network Operations—*Network Monitoring*; G.3 [Mathematics of Computing]: Probability and Statistics—*Survival Analysis*; H.1.2 [Models and Principles]: User/Machine Systems—*Human Factors*

General Terms: Human Factors, Measurement, Performance

Keywords: Cox Model, Gaussian Mixture Model, Internet Measurement, Quality of Service, Survival Analysis, User Perception

1. MOTIVATION

Real-time data transmission is a constant challenge for network researchers. In packet-switched networks, such as the Internet, queuing delay and packet loss are inevitable. To reduce the impact of various types of delay, researchers have focused on rate control and the use of smoothing buffer; and to cope with network loss, various error detection and control mechanisms, such as ARQ and FEC, have been developed. As each solution involves one or more design decisions, the design of a network application is unavoidably *high-dimensional*. For example, an application that is effective for dealing with network delay may be ineffective in dealing with network loss, whereas another one may show opposite behavior. Furthermore, it is difficult to determine and interpret which performance dimension is most important to an applications's users. Therefore, quantifying and comparing the goodness of network design *across different applications* remains an open issue.

Because it is often difficult to justify design choices, we may find that different designs serve the same purpose. For example, some VoIP software programs adopt dynamic de-jitter buffers, while some others adopt static buffers; and some online RPG games adopt UDP as the message transport protocol, while others adopt TCP. As the success of an application depends on many non-technical issues, we cannot conclude that one design is better simply because an application that adopts it is successful in the marketplace. Thus, we are motivated by the question: How can we evaluate and compare two or more network applications in terms of their robustness to network errors?

*This work was supported in part by Taiwan Information Security Center (TWISC), National Science Council of the Republic of China under the grants NSC 96-2219-E-001-001, NSC 96-2219-E-011-008, and NSC 96-2628-E-001-027-MY3.

SIGCOMM'08, August 17–22, 2008, Seattle, Washington, USA.
ACM 978-1-60558-175-0/08/08.

2. METHODOLOGY

We adopt a *user-centric approach*, as we believe that users are the best judges of a network application's performance. In previous studies [1,2], we showed that users tend to leave an application early if they experience adverse network conditions. Based on this finding, we use the Cox model [3] to establish the relationship between user departure rates and the quality of the network path. Rather than model session times directly, we consider the cumulative hazard function, which denotes the probability of user departure events at time t. The function is defined as follows:

$$h(t|\mathbf{X}) = h_0(t)\exp(\sum_{i=1}^{p}\beta_i x_i),$$

where $h_0(t)$ is the baseline hazard function; $\mathbf{X} = x_1, \ldots, x_p$ are the network QoS factors, such as network delay and loss rate; and $\beta_i, \ldots, \beta_p$ denote the respective impact of each p network factor.

As shown by the equation, the model separates the overall cumulative hazard function into two components; one is application-specific and the other is not. The baseline hazard function, $h_0(t)$, refers to the application-specific part. It represents the "ideal" scenario in which there is no network impairment. The second component, the *risk score*, represents the magnifier that summarizes the impact of network impairment, which is comparable across different applications. To calculate the expected hazard function based on current network QoS, we multiply the score by the application-specific baseline hazard function. For example, if the risk score is 1.1, it means the current QoS setting allows a user to leave the system 1.1 times faster than in the ideal scenario. As the risk score has an identical meaning in different applications, we take it as an indicator of an application's network performance under different network conditions.

We define the *Network Robustness Index (NRI)* to quantify an application's overall performance in *a range of network scenarios*. To ensure that NRI is not significantlly affected by specific network configurations, we adopt a Gaussian-mixture model to capture the target network scenarios. We then compute the NRI as the reciprocal of the integral of the risk scores over the mixture model. The NRI is defined as:

$$(\int_{x_1,\ldots,x_p}\exp(\sum_{i=1}^{p}\beta_i x_i)\,dx_1,\ldots,x_p)^{-1},$$

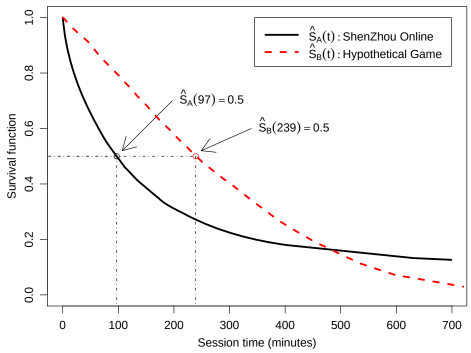

Figure 1: Survival curves of ShenZhou Online and the hypothetical game

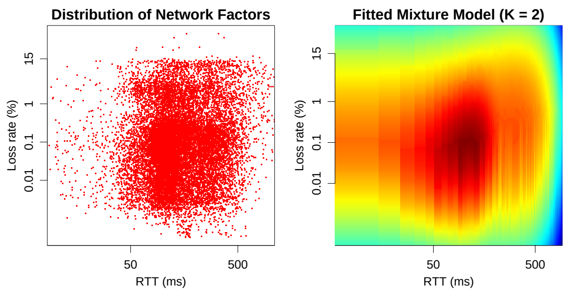

Figure 2: Scatter plots of the target network scenarios and the fitted mixture model

where $(x_1, ..., x_p) \sim \sum_{k=1}^{K} p_k \cdot \mathcal{N}(\mathbf{x}; \mu_k, \sigma_k)$. Here, p_k is the mixing probability of the k-th Gaussian component, and $\mathcal{N}(\mathbf{x}; \mu_k, \sigma_k)$ is a multivariate Gaussian density function with mean μ_k and standard deviation σ_k. The number of mixture components, K, can be specified by ourselves. The parameters p_k, μ_k, and σ_k are inferred by the Expectation-Maximization algorithm from the target network scenarios.

3. PRELIMINARY RESULTS

To demonstrate how our model works, we compare a real-life online game and a hypothetical game. Currently, we do not have a trace for another game, so we use a trace from ShenZhou Online (SZ) [2] and compile a hypothetical trace based on the SZ trace. The median session times of SZ and the hypothetical game are 97 and 239 minutes respectively. We plot the survival curves for both games in Fig. 1.

We regress the network factors to the Cox model and observe that the RTT and packet loss rate have the greatest effect on users' departure decisions. Therefore, we fit the target network scenarios with a bivariate (RTT and packet loss rate) Gaussian mixture model based on the traces. The number of mixture components, K, is set to 2. The scatter plot of the target network scenarios and the fitted model are shown in Fig. 2. We plot the distributions of the risk scores for both games in Fig. 3 and the risk score differences between the two games in Fig. 4. The solid red line in Fig. 4 indicates the set of scenarios in which the risk scores of the two games are equal, where a point represents an observed user in the traces.

As shown in Fig. 3, the risk scores in SZ are generally higher than that of the other game in the same network scenarios, which implies that SZ performs worse than the other game in most scenarios. However, most of users in our traces experienced network scenarios located to the left of the equivalence line, where the the risk scores of the hypothetical game are higher than that of SZ. Thus, even though SZ exhibits less network robustness with most network configurations, it may yield higher NRI than the other game.

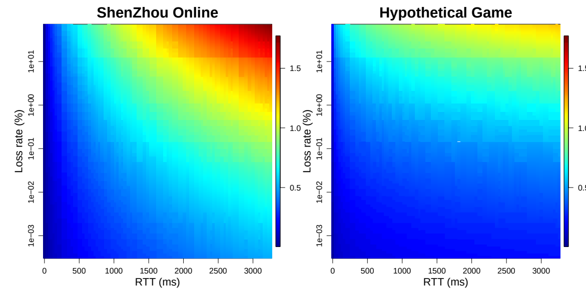

Figure 3: Risk score distributions of ShenZhou Online and the hypothetical game

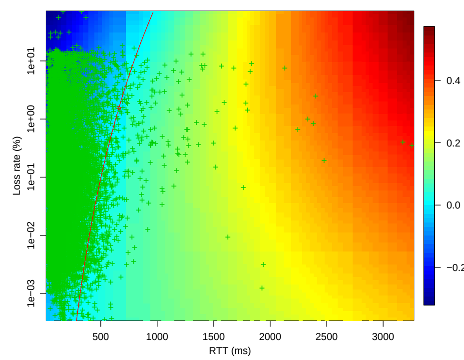

Figure 4: Risk score differences between ShenZhou Online and the hypothetical game.

Our calculations show that the NRI of SZ is 6.58 and that of the hypothetical game is 4.07, which suggests that, on average, SZ's network performance is better than the hypothetical game from the perspective of the users in our traces.

4. CONCLUSIONS AND FUTURE WORK

In this paper, we propose a framework for quantifying an application's ability to handle network impairment. It enables us to evaluate and compare two or more applications in terms of their robustness to network errors. As this is an ongoing work, in the future, we will extend the framework in the following ways. 1) We will further validate our methodology. We need support from the networking perspective to demonstrate the validity of NRI. For example, the computed NRIs of VoIP applications may be justified by the PESQ measurement of the transmitted sound signals. 2) Even though the Cox model is non-parametric, i.e., it is assumed that the distribution of session times does not follow a particular statistical distribution, the model still requires hazard functions under different network scenarios to be in proportion and the effects of network QoS factors are additive. We will further investigate the cases when the model does not provide a good fit and extend it to a time-dependent model if necessary.

5. REFERENCES

[1] K.-T. Chen, C.-Y. Huang, P. Huang, and C.-L. Lei. Quantifying skype user satisfaction. In *Proceedings of ACM SIGCOMM 2006*, Pisa, Itlay, Sep 2006.

[2] K.-T. Chen, P. Huang, G.-S. Wang, C.-Y. Huang, and C.-L. Lei. On the sensitivity of online game playing time to network qos. In *Proceedings of IEEE INFOCOM 06*, Barcelona, Spain, APR 2006.

[3] D. W. Hosmer and S. Lemeshow. *Applied Survival Analysis: Regression Modeling of Time to Event Data.* John Wiley & Sons, Inc., New York, NY, USA, 1999.