논문 13-38B-11-01

한국통신학회논문지(J-KICS) '13-11 Vol.38B No.11
http://dx.doi.org/10.7840/kics.2013.38B.11.861

실시간스캔과 배치스캔을 갖춘 안티바이러스시스템의 운영 분석

양 원 석◆, 김 태 성°

Analysis on Operation of Anti-Virus Systems with Real-Time Scan and Batch Scan

Won Seok Yang◆, Tae-Sung Kim°

요 약

본 논문에서는 정보시스템에 바이러스가 λ의 비율을 갖는 포아송 프로세스를 따라 도착한다고 가정한다. 정보시스템에는 바이러스를 검출하고 치료하기 위해 실시간스캔과 배치스캔의 두가지 방식으로 안티바이러스시스템을 운용하고 있다. 실시간스캔 방식에서는 바이러스가 시스템에 도착하자마자 스캔하게 되어 무한 용량의 안티바이러스 시스템을 보유한 것과 같은 효과가 있다. 스캔과 치료에 소요되는 시간은 일반분포를 따르는 것으로 가정한다. 배치스캔 방식에서는 시스템 관리자가 일정한 시간 간격마다 정기적으로 시스템을 스캔하여 시스템에 존재하는 바이러스들을 동시에 치료한다. 본 논문에서는 안티바이러스시스템의 동작을 확률적으로 모형화하고 경제적으로 최적 운용정책이 달성되는 조건을 유도한다. 비용 요소를 고려하여 실제적인 운용 환경에서의 시사점을 제시할 수 있는 수치예제도 제시한다.

Key Words : anti-virus system, real-time scan, batch scan, economic analysis, probability model

ABSTRACT

We consider an information system where viruses arrive according to a Poisson process with rate λ. The information system has two types of anti-virus operation policies including 'real-time scan' and 'batch scan.' In the real-time scan policy, a virus is assumed to be scanned immediately after its arrival. Consequently, the real-time scan policy assumes infinite number of anti-viruses. We assume that the time for scanning and curing a virus follows a general distribution. In the batch scan policy, a system manager operates an anti-virus every deterministic time interval and scan and cure all the viruses remaining in the system simultaneously. In this paper we suggest a probability model for the operation of anti-virus software. We derive a condition under which the operating policy is achieved. Some numerical examples with various cost structure are given to illustrate the results.

Ⅰ. INTRODUCTION

As the side effects of information society, for example, virus, unauthorized access, theft of proprietary information, denial of access, etc., diffuse, the information security becomes one of

※ 이 논문은 2013년도 한남대학교 교비학술연구조성비 지원에 의하여 연구되었음.
◆ First Author : 한남대학교 경영학과, wonsyang@hnu.kr, 정회원
° Corresponding Author : 충북대학교 경영정보학과, kimts@cbnu.ac.kr, 종신회원
논문번호 : KICS2013-10-426, 접수일자 : 2013년 10월 2일, 최종논문접수일자 : 2013년 11월 4일

the most important issues for organizations. According to a survey by CSI/FBI, total losses resulted by security attacks or misuse amount to about $130 million, and virus attacks alone resulted in about $43 million to 700 organizations[1]. In response to these security threats, organizations have implemented several security counter-measures such as firewalls, anti-virus software, intrusion detection systems (IDS), encryption (of data in transit and of files), smart cards, etc. A major growing concern for organizations is to evaluate and compare the performance of portfolios consisting of security counter-measures.

Recently many researchers have studied on economic aspects of information systems security. The first group of the literature is on assessment of the economic benefits of information security investments. Gordon and Loeb provided an economic modeling framework for assessing the optimal amount to invest in information security to protect a given set of information[2]. Yang et. al. presented economic analysis models to evaluate information security investment portfolios[3,4]. The second group is on the value of individual security technologies. Cavusoglu et al. assessed the value of IDS in a firm's information technology security architecture. The third group is on guidelines for information security investment decision making[5]. Cavusoglu et al. presented an analytic model in an attempt to facilitate decisions regarding security investments[6]. Bodin et al. showed how a chief information security officer can apply the analytic hierarchy process (AHP) to determine the best way to spend a limited information security budget[7]. Kong et al. evaluated information security investments from a BSC perspective[8].

According to the survey of Korean government [9], the most popular information security system for businesses in the country as of December 2011 is anti-virus software (Fig. 1). Followed by firewall, anti-virus software is widely adopted by businesses.

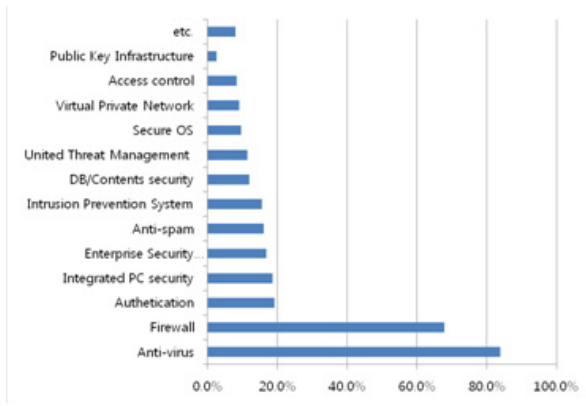

Fig. 1. Information security systems implemented by businesses

In this paper we suggest a probability model for the operation of anti-virus software. We derive a condition under which the operating policy is achieved. Some numerical examples with various cost structures are given to illustrate the results.

The rest of this paper is structured as follows. In section 2, we review previous research. In section 3, we describe a stochastic model with some notations. In section 4, we analyze the stochastic model and obtain the optimal time interval for the batch scan policy to minimize the average operating cost during the time interval for the batch scan. In section 5, we show numerical examples. Finally, we conclude this study in section 6.

Ⅱ. LITERATURE REVIEW

The idea of a batch scan policy was motivated by a clearing mechanism in queueing systems[10]. If a clearing signal occurs to a system, it removes all the customers in the system. The clearing signal resembles a batch scan in this paper, which clears all the viruses from an information system. In the literature related to queueing systems, a clearing signal is usually referred to as called disaster[11-17], and also called mass exodus[18], queue flushing[19], catastrophes[20,21], and stochastic clearing system[10,22]. Note that in the abovementioned research, clearing signals arrive at a system according to a Poisson process. Their inter-arrival

time follows an exponential distribution. In this paper, however, the inter-arrival time of batch scans is deterministic.

Ⅲ. MODEL DESCRIPTION

We consider an information system where viruses arrive according to a Poisson process with rate λ. The information system has two types of anti-virus operation policies including 'real-time scan' and 'batch scan'. In the real-time scan policy, a virus is assumed to be scanned immediately after its arrival. Consequently, the real-time scan policy assumes infinite number of anti-viruses. We assume that the time for scanning and curing a virus follows a general distribution. In the batch scan policy, a system manager operates an anti-virus every deterministic time interval d and scan and cure all the viruses remaining in the system simultaneously. The scale of d depends on the operation scheme of the system, which can be a second, a minute, an hour, a day, and so on.

The anti-virus operation policy considered in this paper is applicable to an information system with hundreds of personal computers connected through a network, for example, local area networks (LAN). Today, in an information system, every user has a copy of an anti-virus software in his or her PC and operates the anti-virus whenever he or she wants. The behavior of operating an anti-virus is different from person to person. That is, how often a user runs an anti-virus or how many files he or she scans might have different personal patterns. Therefore, a virus's survival time becomes stochastic. This corresponds to the real-time scan policy with stochastic operating time. On the other hand, system managers upgrade or update their anti-virus software periodically to deal with new types of viruses. This corresponds to the batch scan policy. After updating the anti-virus, they scan all the viruses in the system in order to protect their system from the damage caused by new viruses. The periodic update of anti-virus softwares corresponds to the deterministic batch scan interval d assumed in this paper. According to the survey of Korean government[9], nearly 40% of internet users update anti-virus software manually.

Ⅳ. COST ANALYSIS

We consider a virus and an anti-virus for the real-time scan policy as a customer and a server in a queueing system, respectively. Note that the real-time scan policy assumes an infinite number of anti-viruses. Accordingly, the stochastic behavior of the model we consider is identical to the time dependent M/G/∞ queue until a system manager operates batch scan.

Let $G(x)$ be the cumulative density function of the operation time for the real-time scan and $Y(t)$ be the number of viruses scanned and cured by the real-time scan policy during $(0,t]$. From Gross and Harris, we have the distribution function of $Y(t)$[23],

$$P[Y(t)=n]=\frac{[\lambda(1-q)t]^n e^{-\lambda(1-q)t}}{n!}, \quad (1)$$

for $n=0,1,2,\cdots$. From (1), we have

$$E[Y(t)]=\lambda(1-q)t, \quad (2)$$

where q is given by

$$q=\frac{\int_0^t [1-G(x)]dx}{t}.$$

We seek to obtain the optimal time interval for the batch scan policy. So we focus on the cost analysis during the time interval d shown in Fig. 2.

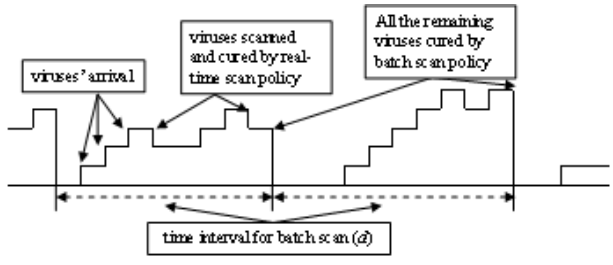

Fig. 2. A sample path of the model

Let c_R and c_B be the operation cost for the real-time scan and that for the batch scan per unit

time, respectively. Let $C(d)$ be the long-run average operating cost during the time interval for the batch scan. Then, we have

$$C(d) = \frac{c_B}{d} + c_R E[Y(d)]. \quad (3)$$

Substituting (2) into (3) results in

$$C(d) = \frac{c_B}{d} + \lambda c_R \left\{ d - \int_0^d [1 - G(x)]dx \right\}. \quad (4)$$

Let d^* be the optimal interval for batch scan which minimizes $C(d)$ in (4). Since $C(d)$ is a continuous and concave function, d^* is a value which makes the derivative of $C(d)$ zero. Then, d^* is obtained by solving the following equation.

$$\frac{c_B}{c_R} = \lambda x^2 G(x). \quad (5)$$

Remark 1. c_B/d^* stands for the average cost during d^*. In addition, $G(d^*)$ is the probability of finishing the real-time scan by d^*. Then, $c_R \lambda d^* G(d^*)$ implies the average cost by the real-time scan. This confirms (5).

Remark 2. The equation (5) shows that we do not need to estimate c_R and c_B separately if we know their ratio.

V. NUMERICAL EXAMPLES

We present numerical examples assuming the distribution of the operation time for the real-time scan. We consider two distributions: an exponential distribution and a uniform distribution.

5.1. Exponential Distribution

It is assumed that the operation time for the real-time scan follows an exponential distribution with rate μ. Then, the cumulative density function of the operation time for the real-time scan results in $G(x) = 1 - e^{-\mu x}$, for $x > 0$. Then, we have

$$C(d) = \frac{c_B}{d} + \lambda c_R d - \frac{\lambda c_R}{\mu}(1 - e^{-\mu d}). \quad (6)$$

$$\frac{c_B}{c_R} = \lambda x^2 [1 - e^{-\mu x}]. \quad (7)$$

First, we present the shape of the average cost in (6) over the interval for the batch scan. It is assumed that $\lambda = 1$, $\mu = 2$, $c_R = 1$, and $c_B = 1$ in Fig. 3. The average cost shows a continuous and concave function and has one global optimal solution. Solving (7) numerically, the optimal value d^* is 3.16.

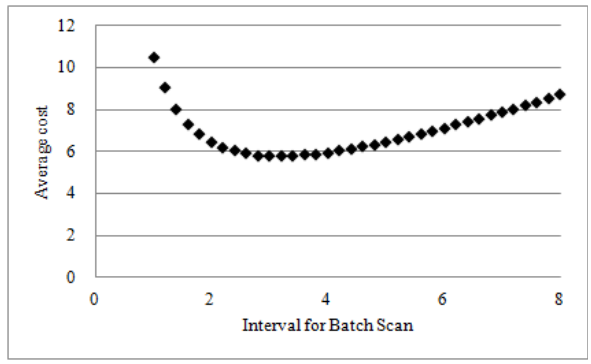

Fig. 3. Average cost

Next, we examine the average costs where the ratio c_B/c_R stays constant. We analyze four cases having the cost structure shown in Table 1 with different values for c_R and c_B the same value for the ratio c_B/c_R. It is assumed that we assumed that $\lambda = 1$ and $\mu = 2$.

Table 1. Cost structure 1

Item	Case 1	Case 2	Case 3	Case 4
c_B	20	15	10	5
c_R	4	3	2	1
c_B/c_R	5	5	5	5

Fig. 4 gives the analysis results of the above four cases in Table 1. Fig. 4 shows that d^* has the same value for each case. This result confirms Remark 1. Solving (7) numerically, d^* is 2.25.

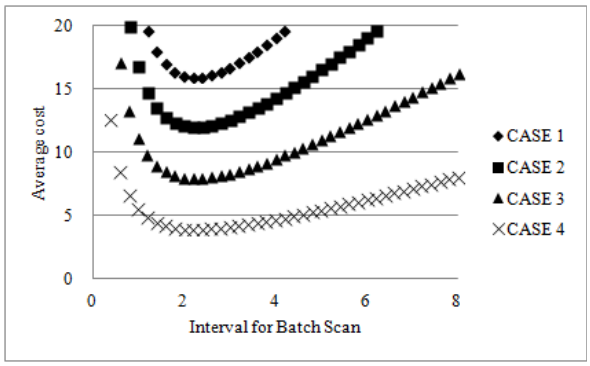

Fig. 4. Average cost with the same value of c_B/c_R

Next, we analyze a numerical example when the ratio c_B/c_R changes as shown in Table 2.

Table 2. Cost structure 2

Item	Case 1	Case 2	Case 3	Case 4
c_B	20	10	5	2
c_R	1	1	1	1
c_B/c_R	20	10	5	2

Fig. 5 shows the average cost $C(d)$ with different values of c_B/c_R in Table 2. Fig. 5 shows that d^* increases as the ratio c_B/c_R increases.

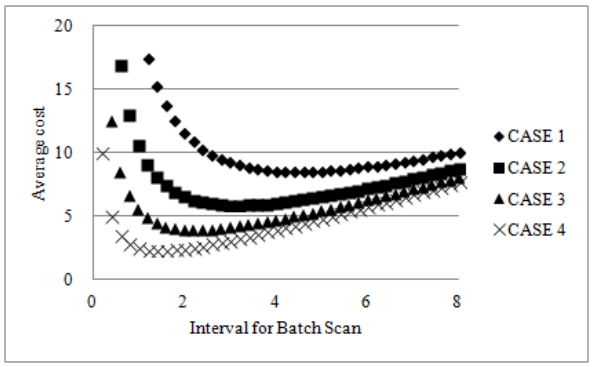

Fig. 5. Average cost with different values of c_B/c_R

We solve the equation in (7) numerically for various c_B/c_R to verify the observation in Fig. 5. The numerical result is summarized in Fig. 6, which confirms that d^* increases as the cost ratio c_B/c_R increases.

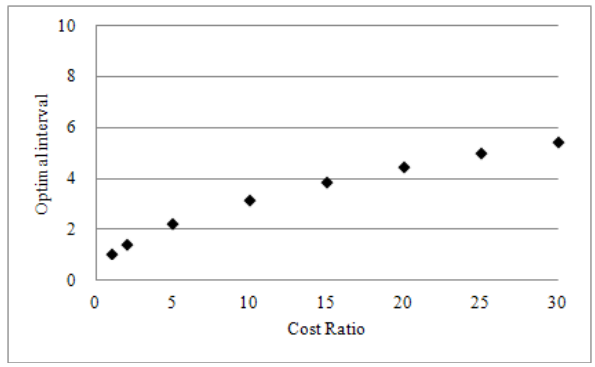

Fig. 6. The optimal interval of batch scan over c_B/c_R

Now, we examine the average cost over the arrival rate of viruses λ with the fixed value of μ, the operation rate of the real-time scan, shown in Table 3. It is assumed that $c_B = 20$ and $c_R = 1$.

Table 3. Arrival rate of viruses

Item	Case 1	Case 2	Case 3	Case 4
λ	8	6	4	2
μ	10	10	10	10

Fig. 7 shows that the average cost increases as the value of λ increases. It is obvious that the number of the real-time scan increases as the occurrence of viruses increases. In addition, d^* seems to decrease as λ increases in Fig. 7.

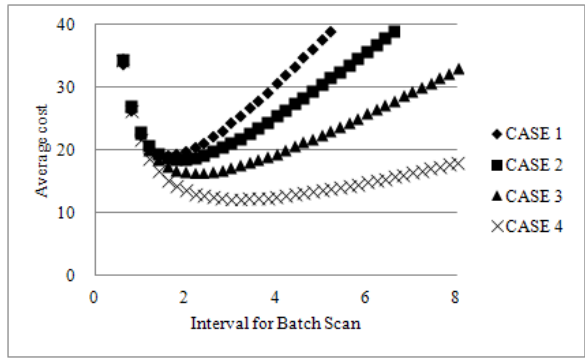

Fig. 7. Average cost over the arrival rates of viruses

Solving (7) for various values of λ results in Fig. 8. It is assumed that μ has the fixed value of 10 similar to Fig. 7, $c_B = 20$ and $c_R = 1$. Fig. 8 confirms that d^* decreases as λ increases. It implies

that we have to perform a batch scan frequently to minimize the average cost as viruses that occur in a system increase.

In numerical examples, we assume that $c_B > c_R$ since usually the cost for a batch scan is higher that a real-time scan. In addition, the observations in the above numerical examples hold good even though $c_B < c_R$.

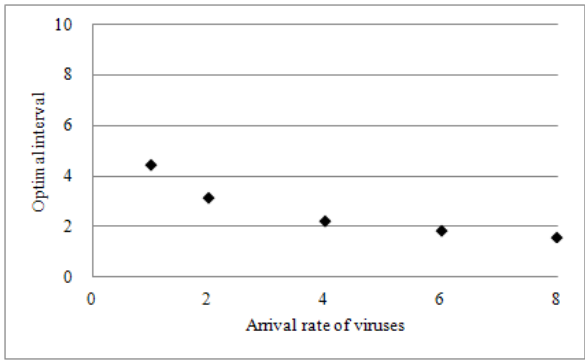

Fig. 8. The optimal interval of batch scan over λ

5.2. Uniform Distribution

It is assumed that the operation time for the real-time scan follows a uniform distribution which is defined in [0,1]. Then, the cumulative density function of the operation time for the real-time scan $G(x) = x$, for $0 \le x \le 1$. Then, we have

$$C(d) = \frac{c_B}{d} + \frac{\lambda c_R d^2}{2}. \quad (8)$$

$$d^* = \sqrt[3]{\frac{c_B}{\lambda c_R}}. \quad (9)$$

The expression in (9) shows that d^* depends on the cost ratio c_B/c_R and increases as the cost ratio increases.

For convenience, it is assumed that $\lambda = 1$, which is the same value used in the case of an exponential distribution. First, we consider the cost structure 1 in Table 1. Fig. 9 shows the average cost with the cost structure 1 in Table 1. Solving (9), we obtain that $d^* = 1.71$ for four cases in Table 1.

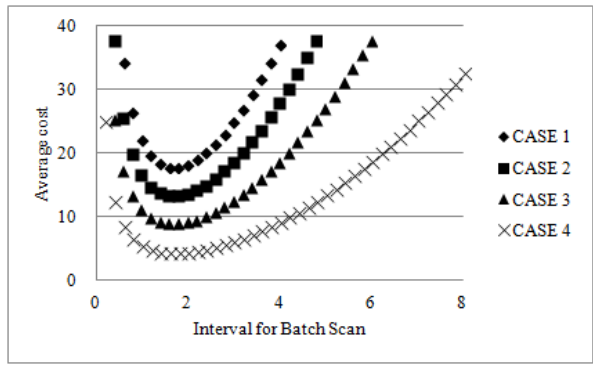

Fig. 9. Average cost with the same value of c_B/c_R

Summarizing the numerical results of the uniform distribution, similar to those of the exponential distribution in Figs. 5-8, results in Figs. 10-13. Fig. 10 shows the average cost in (8). In Fig. 10, the values of c_R and c_B follow the cost structure in Table 2.

Fig. 11 shows that d^* increases as the ratio c_B/c_R increases. Fig. 12 shows the average cost over the arrival rates of viruses given in Table 4. From (8), it is obvious that the average cost increases as the arrival rate of viruses increases as shown in Fig. 12. Finally, Fig. 13 shows that d^* decreases as λ increases. It is obvious since d^* is inversely proportional to $\sqrt[3]{\lambda}$ as shown in (9).

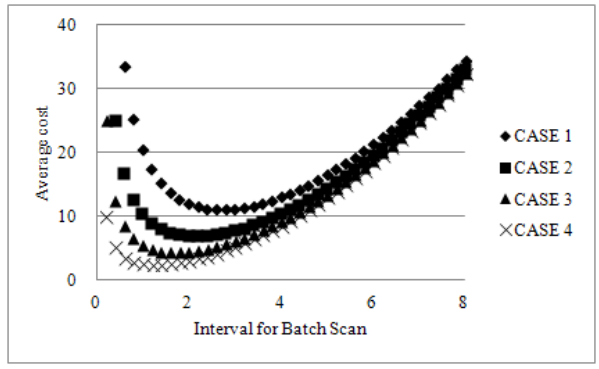

Fig. 10. Average cost with different values of c_B/c_R

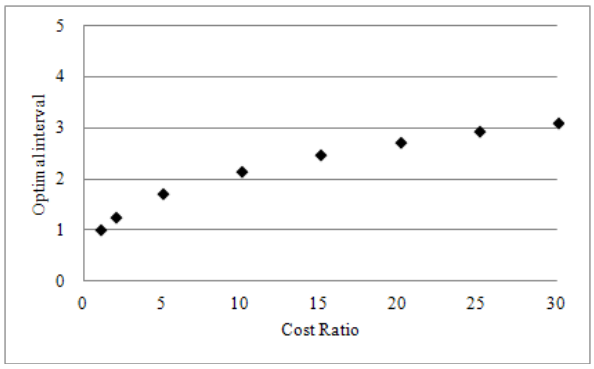

Fig. 11. The optimal interval of batch scan over c_B/c_R

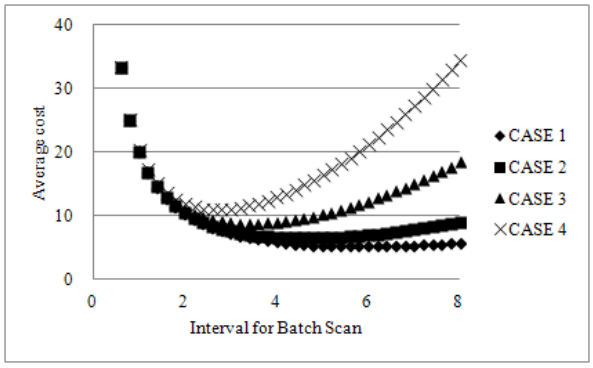

Fig. 12. Average cost over the arrival rates of viruses

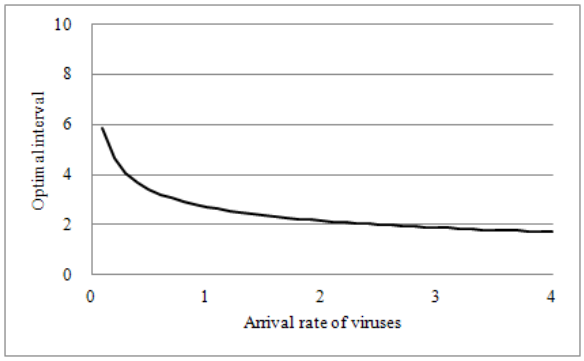

Fig. 13. The optimal interval of batch scan over λ

Table 4. Arrival rate of viruses

Item	Case 1	Case 2	Case 3	Case 4
λ	0.1	0.2	0.5	1.0

Ⅵ. CONCLUSION

We analyzed the operation of anti-virus software, and derived the optimal batch scan interval. With numerical examples, various cost structures were compared on the basis of the long-run average operating cost. Optimal batch scan interval depends on not only cost factors such as operation costs of real-time scan and batch scan, and ratio of two operation costs, but also arrival rate of viruses and operation time of real-time scan. Relative cost of batch scan over real-time scan and virus arrivals have effect on economic operation of anti-virus software.

Finally, we briefly discuss the future research directions. We expect the analysis results and the numerical examples to help the system managers decide the operating policy. Analyses for the economic operation of other major information security counter-measures such as firewalls, IDS, encryption, smart cards, etc can be suggested for the future research. Another research direction may include economic and managerial issues related with security management of critical infrastructure such as smart grid, power plant, transportation, harbor, airport, gas/oil pipeline network. The estimation of the cost parameters including accounting separation of the common cost is also important to improve practical use.

References

[1] Computer Security Institute, *Computer Crime and Security Survey*, Jun. 2011.

[2] L. A. Gordon and M. P. Loeb, "The economics of information security investment," *ACM Trans. Inform. Syst. Security*, vol. 5, no. 4, pp. 438-457, Nov. 2002.

[3] W. S. Yang, T. S. Kim, and H. M. Park, "Probabilistic modeling for evaluation of information security investment portfolios," *J. Korean Operations Research Management Sci. Soc.*, vol. 34, no. 3, pp. 155-163, Sep. 2009.

[4] W. S. Yang, T. S. Kim, and H. M. Park, "Considering system throughput to evaluate information security investment portfolios," *J. Korea Inst. Inform. Security Cryptology*, vol. 20, no. 2, pp. 109-116, Apr. 2010.

[5] H. Cavusoglu, B. Mishra, and S. Raghunathan,

"The value of intrusion detection systems in information technology security architecture," *Inform. Syst. Research*, vol. 16, no. 1, pp. 28-46, Mar. 2005.

[6] H. Cavusoglu, B. Mishra, and S. Raghunathan, "A model for evaluating IT security investments," *Commun. ACM*, vol. 47, no. 7, pp. 87-92, July 2004.

[7] L. D. Bodin, L. A. Gordon, and M. P. Loeb, "Evaluating information security investments using the analytic hierarchy process," *Commun. ACM*, vol. 48, no. 2, pp. 79-83, Feb. 2005.

[8] H. K. Kong, T. S. Kim, and J. Kim, "An analysis on effects of information security investments: a BSC perspective," *J. Intell. Manufacturing*, vol. 23, no. 4, pp. 941-953, Aug. 2012.

[9] Korea Communication Commission (KCC) an d Korea Internet & Security Agency (KISA), *Information Security Survey-Businesses*, Mar. 2012.

[10] W. S. Yang, J. D. Kim, and K. C. Chae, "Analysis of M/G/1 stochastic clearing systems," *Stochastic Anal. Applicat.*, vol. 20, no. 5, pp. 1083-1100, Oct. 2002.

[11] G. Jain and K. Sigman, "A Pollaczek-Khintchine formula for M/G/1 queues with disasters," *J. Applied Probability*, vol. 33, no. 4, pp. 1191-1200, Dec. 1996.

[12] I. Atencia and P. Moreno, "The discrete-time Geo/Geo/1 queue with negative customers and disasters," *Comput. Operations Research*, vol. 31, no. 9, pp. 1537-1548, Aug. 2004.

[13] A. Gomez-Corral, "On a finite-buffer bulk-service queue with disasters," *Math. Methods Operations Research*, vol. 61, no. 1, pp. 57-84, Mar. 2005.

[14] F. Jolai, S. M. Asadzadeh, and M. R. Taghizadeh, "Performance estimation of an Email contact center by a finite source discrete time Geo/Geo/1 queue with disasters," *Comput. Ind. Eng.*, vol. 55, no. 3, pp. 543-556, Oct. 2008.

[15] X. W. Yi, J. D. Kim, D. W. Choi, and K. C. Chae, "The Geo/G/1 queue with disasters and multiple working vacations," *Stochastic Models*, vol. 23, no. 4, pp. 21-31, Nov. 2007.

[16] H. M. Park, W. S. Yang, and K .C. Chae, "Analysis of the GI/Geo/1 queue with disasters," *Stochastic Anal. Applicat.*, vol. 28, no. 1, pp. 44-53, Jan. 2010.

[17] D. H. Lee, W. S. Yang, and H. M. Park, "Geo/G/1 queues with disasters and general repair times," *Applied Math. Modelling*, vol. 35, no. 4, pp. 1561-1570, Apr. 2011.

[18] A. Chen and E. Renshaw, "The M/M/1 queue with mass exodus and mass arrivals when empty," *J. Applied Probability*, vol. 34, no. 1, pp. 192-207, Mar. 1997.

[19] D. Towsley and S. K. Tripathi, "A single server priority queue with server failures and queue flushing," *Operations Research Lett.*, vol. 10, no. 6, pp. 353-362, Aug. 1991.

[20] E. G. Kyriakidis and A. Abakuks, "Optimal pest control through catastrophes," *J. Applied Probability*, vol. 27, no. 4, pp. 873-879, Dec. 1989.

[21] X. Chao, "A queueing network model with catastrophes and product form solution," *Operations Research Lett.*, vol. 18, no. 2, pp. 75-79, Sep. 1995.

[22] J. R. Artalejo and A. Gomez-Corral, "Analysis of a stochastic clearing system with repeated attempts," *Stochastic Models*, vol. 14, no. 3, pp. 623-645, Jun. 1998.

[23] D. Gross and G. M. Harris, *Fundamentals of Queueing Theory*, John Wiley & Sons, 1974.

양 원 석 (Won Seok Yang)

1993년 2월 KAIST 경영과학과 학사
1995년 2월 KAIST 경영과학과 석사
2000년 2월 KAIST 산업공학과 박사
2000년 2월~2007년 1월 LG U+ 차장
2007년 2월~2010년 2월 ETRI 선임연구원
2010년 3월~현재 한남대학교 경영학과 교수
<관심분야> 확률모형, 대기행렬 이론, 생산관리, 통신정책, 통신망 성능분석, 기술경제성, 보안 경제성

김 태 성 (Tae-Sung Kim)

1997년 2월 KAIST 산업경영 박사
1997년 2월~2000년 8월 ETRI 선임연구원
2005년 1월~2006년 2월 U of North Carolina at Charlotte 방문교수
2010년 7월~2012년 7월 Arizona State University 방문연구원
2000년 9월~현재 충북대학교 경영정보학과 교수, 대학원 정보보호경영학과 주임교수
<관심분야> 통신 및 보안 분야의 경영 및 정책 의사결정