Performance analysis of email systems under three types of attacks

Yang Wang\textsuperscript{a}, Chuang Lin\textsuperscript{b}, Quan-Lin Li\textsuperscript{c,}\textsuperscript{*}

\textsuperscript{a} Department of Computer Science and Technology, Peking University, Beijing 100871, PR China
\textsuperscript{b} Department of Computer Science and Technology, Tsinghua University, Beijing 100084, PR China
\textsuperscript{c} School of Economics and Management Sciences, Yanshan University, Qinhuangdao 066004, PR China

\textbf{ARTICLE INFO}

\textbf{Article history:}
Received 29 July 2008
Received in revised form 15 September 2009
Accepted 17 January 2010
Available online 10 February 2010

\textbf{Keywords:}
Performance analysis
Network security
Email attack
Queueing system
Information leakage probability

\textbf{ABSTRACT}

Email is a crucial part of our daily life, but email systems are facing increasing security threats such as attacks and spam. Because of this, new mechanisms are being provided to defend against the attacks and to reduce the amount of spam in email systems. Up to now, few available works have been reported on the performance evaluation of email systems under attack, which has become necessary for enhancing email security. In this paper, we propose a novel method to study the impact of three types of attacks on email systems. We construct a multiple queueing model to characterize three types of attacks integrally, and study the performance metrics of system security such as system availability, average queue length and information leakage probability. Numerical examples indicate that the approach of this paper is effective and efficient for dealing with the security analysis of email systems under attack. We believe that this work will open a new avenue for the performance evaluation of computer networks under email attack and other forms of attacks.

© 2010 Elsevier B.V. All rights reserved.

\section{Introduction}

With the fast development of Internet technology, emails have become a necessary part of our daily life. Today, email systems are facing increasing threats of various attacks and spam. In particular, email attacks have been causing economic and social issues worldwide. In its 2007 annual report, Microsoft reported the increase of email attacks, especially, a steep rise in specifically targeted email attacks. Widespread worms and viruses sent to millions of mailboxes are a main cause of destruction to email systems. Targeted attacks aiming at crucial affairs such as specific businesses, sensitive politics and military competitions are becoming nightmares. Therefore, it is urgent to study such system performance in order to enhance the security of email systems.

Some works have dealt with email security from several different aspects. Important examples include the following. Liao and Schwenk [1], Li and Somayaji [2], Xiong [3] and others focused on the anomaly detection of the attacks and spam. In particular, Kumaraguru, Rhee, Acquisti, Cranor, Hong and Nunge [4] discussed an embedded training email system that taught people about phishing during the normal use of emails. Wang and Abdel-Wahab [5] proposed an anomaly detection system on the layer correlation for an email system. Their detection engine was validated in an environment including normal email activities, simulated worm activities and attacks on an email server. Negnevitsky, Lim, Hartnett and Reznik [6] surveyed the analysis of complex networks and user behaviors related to the analysis of email communications, including anomaly detection in the behavior of email usage. Stolfo, Hershkop, Wang and Nimeskern [7] designed a behavior-based security system employing anomaly detection techniques, including a data analyzing system to assist email behavior.
that in reality, there are three types of attacks on email accounts. The first type is to crash the email system, based on different attacking targets, are generally classified as two types: the first is to crash the email system, and the other one is to take valuable information from the email account. The first type of email attack is a crucial case of Internet attacks. Such attackers try all sorts of ways to get control of the email server by means of, such as a sniffer, scanning, or exploiting vulnerabilities. In this case, the email server can be destroyed or exposed when the attacker breaks into the email system. Another type of email attacks lays more stress on obtaining valuable information from the email accounts. It is seen from Fig. 1 that in reality, there are three types of attacks on email accounts. The first type is to crash the password of the email account, and then get valuable information about the email account. The second type is to send malicious emails to some email accounts. These malicious emails may contain worms, viruses and Trojans as an attachment, and may contain images with clickable dangerous web links. A Trojan will be planted when a user clicks the malicious attachment. Therefore, the attackers will get remote control of the email systems. Based on this, valuable email information can be obtained by the attackers. The third type is to use an email bomb to send a large amount of emails to crash the victim. Because of this, the user cannot use the email account for a certain time period. Note that the email bomb is actually a type of DoS attack; see, e.g., [15].

Fig. 1 provides an intuitive interpretation for the three types of attacks that operate in an email system. In the rest of the paper, we provide a novel queueing model to study their impact on the security performance metrics of email systems.

3. Queueing analysis

In this section, we provide a novel queueing model to describe an email system under three types of attacks, as shown in Fig. 2. To do this, we describe the ordinary emails and the three types of email attacks as the input and output processes in a multiple queueing system. Furthermore, we also express the queueing model as an irreducible continuous-time QBD process with finitely many levels.
The ordinary emails in each email account can form a basic queue that is expressed through email information units. This email queue is drawn as the top one in Fig. 2. The ordinary email arrivals are a Poisson process with rate $\lambda > 0$, and each email information unit in the email account is dealt with for a time period which is exponential with mean $1/\mu > 0$. These email information units are processed according to the FCFS discipline. Assume that each email account has a capacity of $N$ email information units at most. This model can be easily extended to have multiple kinds of ordinary emails by introducing multiple parameters of exponentially distributed times. It will use the same mathematical method, only leaving the Markovian chain more complex. For presentation clarity, we will not give the extension here.

An attack involving cracking a password is abstracted as an input to the second queue, as shown in Fig. 2. By cracking the password of the email account, the attacker can get some useful and valuable email information for either interests of himself, a business or other purposes. The events to successfully get the password are treated as an attacking process or a customer arrival process that is assumed to be a Poisson process with rate $\alpha_c > 0$. Such a customer remains in the email account with a time period $x_c$ that is of phase type with irreducible representation $(\gamma_c, S_c)$, and $S_c^0 = -S_c e$ is the absorbing
rate vector. The phase type distribution is a generalization of the exponential distribution. In detail, it characterizes the distribution of the time that a finite-state Markov process takes to reach a given state from its initial state distribution, where $S_c$ is the Markov chain and $\gamma$ is the initial vector of the Markov chain $S_c$. The phase type distribution can be used to describe a more general case with an algorithmic solution. More information about phase type distribution can be found in [14]. Since the attackers are only interested in some useful and valuable email information, we assume that this type of email attack can grab the email information with a probability $P_1(\alpha_k)$ when it remains in the email account.

An attack involving sending malicious emails with a Trojan virus attached is modeled as an input to the third process, described as the third queue in Fig. 2. The Trojan Horse is planted if the email user clicks the attachment. Based on this, such attackers could get into the email system to read the email account’s private and valuable information. Malicious emails are regarded as a customer arrival process that is a Poisson process with rate $\alpha_m > 0$. Sometimes, the attached Trojan virus is defended by a firewall or it is ignored by the email user. Let $P_m$ denote the probability that the malicious attachment is clicked on by the email user, which means that the attacker gets access into the email account. This type of attack remains in the email system for a time period $x_m$, which is of phase type with irreducible representation $(\gamma_m, S_m)$, and $S_m^0 = -S_m e$ is the absorbing rate vector. Let $P_m(x_m)$ be the probability that the attacker is able to obtain the email information.

The fourth input process in Fig. 2 is an attack involving email bombs. For ease of description, we only treat the successfully deployed email bombs as customers, whose arrivals are a Poisson process with rate $\alpha_b > 0$. The email bombs usually cause severe damage whenever they happen. Thus, once an email bomb arrives, the email account is assumed to be crashed for a time period whose length is exponential with mean $1/\beta_b > 0$.

It is worth noting that attacks involving cracking passwords and sending malicious emails will not change the ordinary email queue. However, attacks involving email bombs can change the behavior of the ordinary email queue.

Based on the above description, the排队模型 for the email system under attack has four types of waiting lines: one waiting line is for the ordinary email information units while the other lines are for the three types of attacks. We assume that these different arrival processes and all the above time periods are independent of each other. It is noted that each customer in the attack queues does not require any actual service, so that it will directly enter the email account without any waiting time, while it can remain in each email account for a random time period in order to take valuable information, crash the system, shut down the email services or have other purposes. Thus, we integrate the four queues into a novel multiple queueing model, which will be analyzed by the QBD process as follows.

Now, we express the multiple queue system as an irreducible continuous-time QBD process with finitely many levels, which is necessary for performance analysis of this system.

Let $n(t)$, $s(t)$ and $r(t)$ denote the number of email information units, the state of the server (i.e., email account) and the types of attacks in the email system at time $t$, respectively, where $0 \leq n(t) \leq N$ is the maximum number of email information units in the email account, $s(t) \in \{I, W, F\}$ and $r(t) \in \{C, M, CM\}$. We provide a simple interpretation for the elements $I$, $W$, $F$, $C$, $M$ and $CM$. First, $I$, $W$ and $F$ stand for server status Idle, Working and Fail, respectively. Second, $C$, $M$ and $CM$ describe the attacks involving cracking passwords, the attacks involving malicious email, and the co-existence of attacks involving cracking passwords and malicious emails, respectively. Furthermore, let $I_1(t)$ and $I_2(t)$ be the phase numbers of the two PH (phase type) distributions $(\gamma_c, S_c)$ and $(\gamma_m, S_m)$ at time $t \geq 0$, respectively, where $1 \leq I_1(t) \leq q_c$ and $1 \leq I_2(t) \leq q_m$, $q_c$ and $q_m$ are the orders of the two PH distributions. Thus $\{n(t), s(t), r(t), I_1(t), I_2(t) : t \geq 0\}$ is a continuous-time Markov chain whose state space $\Omega$ is given by

$$\Omega = \bigcup_{k=0}^{N} L_k,$$

where

$$L_0 = \left\{ (0, j_1, j_2, I), (0^C, j_1, j_2, I), (0^M, j_1, j_2, I), (0^{CM}, j_1, j_2, I), (0, j_1, j_2, F) : 1 \leq j_1 \leq q_c, 1 \leq j_2 \leq q_m \right\},$$

and, for $1 \leq i \leq N$,

$$L_i = \left\{ (i, j_1, j_2, W), (i^C, j_1, j_2, W), (i^M, j_1, j_2, W), (i^{CM}, j_1, j_2, W), (i, j_1, j_2, F) : 1 \leq j_1 \leq q_c, 1 \leq j_2 \leq q_m \right\}.$$

Each $L_k$, $0 \leq k \leq N$, denotes one level. Here, we provide a simple interpretation for the elements in $L_k$ for $k \geq 0$. For example, in $(i, j_1, j_2, I)$, $(i, j_1, j_2, W)$, $(i, j_1, j_2, F)$, $i$ represents that there are $i$ ordinary email information units; in $(i^C, j_1, j_2, I)$, $(i^C, j_1, j_2, W)$, $(i^M, j_1, j_2, W)$, $(i^{CM}, j_1, j_2, W)$, $(i, j_1, j_2, F)$, $i$ and $j_1, j_2$ denote that the attacks involving cracking passwords are in the email system. In a similar way, we can easily understand $(i^M, j_1, j_2, I)$, $(i^M, j_1, j_2, W)$, $(i^{CM}, j_1, j_2, I)$ and $(i^{CM}, j_1, j_2, W)$.

According to the state space $\Omega$, we give the state transitions of the Markov chain as shown in Fig. 3. There we have transitions among states in one level $L_k$ $(0 \leq k \leq N)$ or between different levels. States in one level have the same number of email information units; thus there is no transition for the arrival or departure of ordinary emails among the different states in the same level, where the transitions are only caused by the email attacks’ actions. Transitions among states of different levels are for ordinary emails’ arrival and departures.
Based on the state transitions, it is easy to see that the Markov chain is a QBD process whose infinitesimal generator is given by

$$Q = \begin{pmatrix}
A_1^{(0)} & A_0^{(0)} \\
A_1^{(1)} & A_0^{(1)} \\
& \ldots \\
A_1^{(N)} & A_0^{(N)}
\end{pmatrix},$$

where $A_i^{(i)}$ denotes the transition rate matrix among different phases in level $l_i$ for $0 \leq i \leq N$, and $A_j^{(j)}$ and $A_k^{(k)}$ represent the transition rate matrices between two different levels. More concretely, we can write $A_i^{(i)}$ for $0 \leq i \leq N$, $A_j^{(j)}$ for $1 \leq j \leq N$ and $A_k^{(k)}$ for $0 \leq k \leq N - 1$ as follows:

$$A_i^{(i)} = \begin{pmatrix}
-(\alpha_c + \alpha_m + \alpha_b)l_i \otimes l_m \\
\alpha_c l_i \otimes l_m \\
\alpha_m l_i \otimes l_m \\
0 \\
\alpha_b l_i \otimes l_m
\end{pmatrix},$$

for $1 \leq i \leq N - 1$, and

$$A_1^{(0)} = \begin{pmatrix}
-(\alpha_c + \alpha_m + \alpha_b)l_1 \otimes l_m \\
-\lambda l_1 \otimes l_m \\
0 \\
0 \\
0
\end{pmatrix},$$

$$A_1^{(N)} = \begin{pmatrix}
-(\alpha_c + \alpha_m + \alpha_b)l_1 \otimes l_m \\
\mu l_1 \otimes l_m \\
0 \\
0 \\
0
\end{pmatrix}.$$
and the information leakage probability, which is used to describe the information leakage from the email account due to bomb attacks on the email system; (2) the average queue length, which denotes the email information units of ordinary emails; and (3) the information leakage probability, which is used to describe the information leakage from the email accounts due to the two types of attacks: attacks involving cracking passwords and attacks involving sending malicious emails.

4. Performance metrics

In this section, we first provide an RG-factorization solution to the QBD process, and then give some security performance metrics for email system under attack.

Because of the finite buffer size of the email system, the QBD process $Q$ had finitely many levels, so its state space is finite. Note that $Q$ is irreducible; thus this QBD process is positive recurrent. Let $\pi = (\pi_0, \pi_1, \ldots, \pi_N)$ be the stationary probability vector of the QBD process $Q$, partitioned according to levels. We write $\pi_i = (\pi_{i,1}, \pi_{i,2}, \ldots, \pi_{i,N})$ for $0 \leq i \leq N$.

Since $\pi$ is the stationary probability vector of the QBD process $Q$, we have

$$\begin{aligned}
\pi Q &= 0, \\
\pi e &= 1.
\end{aligned}$$

(2)

Applying an LU-type RG-factorization, given in [13] or [14], we provide an RG-factorization solution to Eq. (2). RG-factorization is a kind of factorization approach for QBD processes. It is iteratively constructed in terms of the R-measures and G-measures. The LU-type RG-factorization is a special type of RG-factorization. For more detailed information about RG-factorization, refer to [14].

Now, we first write the U-measures as

$$U_0 = A_1^{(0)}$$

and

$$U_k = A_1^{(k)} + A_2^{(k)}(-U_{k-1}^{-1})A_0^{(k-1)}, \quad 1 \leq k \leq N.$$  

Since the QBD process is irreducible, the matrix $U_k$ is invertible for $0 \leq k \leq N - 1$ while the matrix $U_N$ is singular. Based on the matrices $U_k$ for $1 \leq k \leq N$, the R-measures and G-measures are defined as

$$R_k = A_2^{(k)}(-U_{k-1}^{-1}), \quad 1 \leq k \leq N,$$

and

$$G_k = (-U_{k-1}^{-1})A_0^{(k)}, \quad 0 \leq k \leq N - 1.$$  

Using the matrix sequences $\{U_k, 0 \leq k \leq N\}, \{R_k, 1 \leq k \leq N\}$ and $\{G_k, 0 \leq k \leq N - 1\}$, the LU-type RG-factorization for matrix $Q$ is given by

$$Q = (I - R_L)U_D(I - G_U),$$

(3)

where

$$U_D = \text{diag}(U_0, U_1, \ldots, U_{N-1}, U_N),$$

$$R_L = \begin{pmatrix}
0 & 0 & \cdots & 0 \\
R_1 & 0 & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots \\
R_{N-1} & 0 & \cdots & 0 \\
R_N & 0 & \cdots & 0
\end{pmatrix},$$

$$G_U = \begin{pmatrix}
0 & G_0 & \cdots & \cdots \\
0 & G_1 & \cdots & \cdots \\
\cdots & \cdots & \cdots & \cdots \\
0 & \cdots & \cdots & G_{N-1}
\end{pmatrix}.$$  

By means of the LU-type RG-factorization, Li and Cao [13] and Li [14] provide the expression for the stationary probability vector $\pi$ as follows:

$$\pi_k = \kappa \nu_N R_{N} R_{N-1} \cdots R_{k+1} \nu_{k+1}, \quad 0 \leq k \leq N - 1,$$

$$\pi_N = \kappa \nu_N$$

where $\kappa$ is a normalization constant and $\nu_N$ is the stationary probability vector of the censored chain $U_N = A_1^{(N)} + R_N A_0^{(N-1)}$ to level $N$, which is positive recurrent according to the censoring properties.

Based on the stationary probability vector $\pi$, we can compute the security metrics for the email systems. In what follows, we focus on three security performance metrics: (1) system availability, which expresses the impact of the attacks of email bombs on the email system; (2) the average queue length, which denotes the email information units of ordinary emails; and (3) the information leakage probability, which is used to describe the information leakage from the email accounts due to the two types of attacks: attacks involving cracking passwords and attacks involving sending malicious emails.
Now, we define the stationary security performance metrics as follows.

(1) System availability

The system availability is the quantification of the alternation between proper and improper operations of the email system. The email system is suspended by attacks involving email bombs; thus the system becomes unavailable. In general, the stationary system availability, denoted as $A$, is expressed as the time fraction in steady state as follows:

$$ A = \sum_{i=0}^{N} \sum_{j=1}^{4} \pi_{i,j} = 1 - \sum_{i=0}^{N} \pi_{i,5}. \quad (4) $$

(2) Average queue length

The probability that there are $i$ ordinary email information units is given by

$$ P_i = \sum_{j=1}^{5} \pi_{i,j}. \quad (5) $$

Thus the stationary average queue length, denoted as $AQL$, is defined as the stationary average number of ordinary email information units when the system is in steady state. Thus we obtain

$$ AQL = \sum_{i=1}^{N} iP_i. \quad (6) $$

Now, we provide a new interpretation for the $AQL$. Since attacks involving cracking passwords and attacks involving malicious emails do not affect the ordinary email information unit number in this system, we will not consider either of them in the computation of $AQL$. Based on the attacks involving email bombs, we may define a generalized processing time, denoted as $y$, of each ordinary email information unit by the email account: the generalized processing time is the time length from the beginning of the processing time of an ordinary email to its completed time after being processed by the email account.

Fig. 4 shows the structure of the generalized processing time. It is clear from Fig. 4 that the generalized processing time is of phase type with irreducible representation $(\theta, T)$, where

$$ \theta = (1, 0) $$

and

$$ T = \begin{pmatrix} -\mu + \alpha_b & \alpha_b \\ -\beta_b & -\beta_b \end{pmatrix}. $$

It is clear from Chapter 2 of Neuts [16] that $E[y] = -\theta T^{-1} e$.

Using the generalized processing time, the email system for computing $AQL$ can be described as an $M/PH/1/N$ queue, which can be used to derive a new expression for the $AQL$ in terms of the QBD process with finitely many levels, the detailed solving procedures of which are listed in [13]. Therefore, we omit the computation details here.

(3) Information leakage probability

We quantify the information leakage probability ($ILP$) of each email account. We first express the average number of leaked email information units as follows:

$$ ILN = \sum_{i=1}^{N} i(\pi_{i,2}P_c + \pi_{i,3}P_m + \pi_{i,4}P_{cm}). \quad (7) $$
where \( P_c \) and \( P_m \) are two information obtaining probabilities by the attackers through the attacks involving cracking password and the attacks involving malicious emails, respectively. Note that \( P_{cm} \) is the probability to obtain the email information units when the both types of attacks simultaneously exist in the system.

Based on the \( ILN \) and \( AQL \), we provide the average information leakage probability as follows:

\[
ILP = \frac{ILN}{AQL}.
\]

Now, we study the impact of the cracking password attacks and malicious email attacks on the email system, respectively. At the same time, we present the bounds for the ratio of attack impact to the total information leakage. The impact ratio \( ILR \) of the attacks involving cracking passwords can be estimated by the upper and lower bounds as follows:

\[
ILR^{(i)}_c \leq ILR_c \leq ILR^{(ii)}_c,
\]

where

\[
ILR^{(i)}_c = \frac{\sum_{i=1}^{N} i\pi_{i,2}P_c}{\sum_{i=1}^{N} (i\pi_{i,2}P_c + \pi_{i,3}P_m + \pi_{i,4}P_{cm})}
\]

and

\[
ILR^{(ii)}_c = \frac{\sum_{i=1}^{N} (i\pi_{i,2}P_c + \pi_{i,4}P_{cm})}{\sum_{i=1}^{N} (i\pi_{i,2}P_c + \pi_{i,3}P_m + \pi_{i,4}P_{cm})}.
\]

Similarly, the impact ratio \( ILR_m \) of the attacks involving malicious emails to the total email information leakage is given by

\[
ILR^{(i)}_m \leq ILR_m \leq ILR^{(ii)}_m,
\]

where

\[
ILR^{(i)}_m = \frac{\sum_{i=1}^{N} i\pi_{i,3}P_m}{\sum_{i=1}^{N} (i\pi_{i,2}P_c + \pi_{i,3}P_m + \pi_{i,4}P_{cm})}
\]

and

\[
ILR^{(ii)}_m = \frac{\sum_{i=1}^{N} (i\pi_{i,3}P_m + \pi_{i,4}P_{cm})}{\sum_{i=1}^{N} (i\pi_{i,2}P_c + \pi_{i,3}P_m + \pi_{i,4}P_{cm})}.
\]

Through the above three metrics, we can study the email security from different aspects. The system availability of the email system shows the security from the system level; the information leakage probability characterizes the information level security; and the average queue length denotes the system load when facing the three types of attacks, especially attacks involving email bombs. The three security metrics are useful for revealing the security issues of email systems; for example, we can gain some insights into the three security metrics through numerical examples in the next section.

5. Numerical examples

In this section, we provide some numerical examples to observe the impact of the three types of attacks on the performance metrics of an email system.

In the first example, we consider the system availability with respect to the arrival rate and the recovery rate for email bombs. Note that the email bombs can cause the email system to go out of service. Assume that \( N = 100, \lambda = 10, \mu = 12; \alpha_c = 0.1, \gamma_c = (0.4, 0.6), S_c = \begin{pmatrix} -0.6 & 0.4 \\ 0.3 & -0.5 \end{pmatrix}, \alpha_m = 1, \gamma_m = (0.5, 0.5), S_m = \begin{pmatrix} -1.8 & 0.6 \\ 0.7 & -1.9 \end{pmatrix}; P_{in} = 0.7, p_c = 0.5, P_m = 0.8 \) and \( P_{cm} = 0.9 \). As depicted in Fig. 5, the system availability, denoted as \( A \), is decreasing with the email bomb arrival rate \( \alpha_b \) while is increasing with the system recovery rate \( \beta_b \).

We numerically study the average queue length \( AQL \) where all the parameters in Fig. 5 are kept except that we set \( \alpha_b = 0.01 \) and \( \beta_b = 1 \). The \( AQL \) is tightly related to the email arrival rate \( \lambda \) and the service rate \( \mu \). Assume that \( \mu = 12 \), and \( \lambda \) is taken from 0 to 30. From Fig. 6, we can observe that the queue is saturated when \( \lambda \) becomes larger than \( \mu \). This is the same as our expectation in an ordinary queueing system.
Now, we conduct a numerical analysis for the email information leakage. Fig. 7 shows how the information leakage number $ILN$ depends on the email arrival rate $\lambda$. It is seen from Fig. 7 that $ILN$ is increasing with the email arrival rate $\lambda$, until the email queue of an account is saturated. At the same time, Fig. 8 denotes how the information leakage probability $ILP$ depend on $\alpha_b$ and $\beta_b$. It is observed that the $ILP$ decreases with the arrival rate $\alpha_b$ of the email bombs but increases with the processing rate $\beta_b$ of the email bombs. This is due to the fact that the email system is totally down once it is suffering from email bombs, while no information is leaked during the active period of the email bombs. Thus, less information is leaked when the attacks of the email bombs last for a longer time. Meanwhile, Fig. 9 depicts how the $ILP$ is affected by $P_c$. It is noted that the $ILP$ is linearly increasing with $P_c$ for different $\alpha_a$ or $\alpha_m$, while the $ILP$ deceases with $\alpha_b$ and increases with $\alpha_m$.

We further study the impact ratios $ILR_c$ and $ILR_m$ of the two different types of attacks on email information leakage, respectively. Fig. 10 shows the bounds of $ILR_c$ and $ILR_m$ which are changing with $\alpha_c$. It is seen that the impact ratio of the
attacks involving cracking passwords is increasing with its arrival rate $\alpha_c$, while the impact ratio of the attacks involving malicious email is decreasing with $\alpha_c$. Fig. 11 shows the effect of $\alpha_b$ on $ILR_c$ and $ILR_m$; it is noticed that $ILR_c$ is slightly decreasing with $\alpha_b$, while $ILR_m$ is increasing with $\alpha_b$. Furthermore, attacks involving malicious emails have more severe impact on the email system than attacks involving cracking passwords. Fig. 12 depicts how the impact ratios of the two types of attacks depend on $P_c$ and $P_m$, respectively. It is seen that both impact ratios are increasing with $P_c$ and $P_m$ within similar upper and lower bounds.

Based on the above analysis, we summarize that attacks involving email bombs create rather serious damage for email systems; thus the whole service in the email system can become crashed. In contrast, the other two types of attacks, cracking passwords and malicious emails, can cause a large amount of email information to be leaked. Obviously both of them are prevalent and serious threats to email security which is necessary in our daily life and work.
6. Sensitivity analysis

In this section, we provide a sensitivity analysis for the performance metrics with respect to some crucial parameters of an email system. Such a sensitive analysis is necessary for dealing with security and stability of an email system under attack.

It is seen from (1) that a sensitivity analysis of such an email system can be regarded as that of a QBD process. Thus, we need to consider a perturbed continuous-time QBD process $\{x(t), t \geq 0\}$, whose infinitesimal generator is given by

$$Q(\epsilon) = Q_1 + \epsilon Q_2,$$

where $\epsilon$ is a sufficiently small positive number, and

$$Q_1 = \begin{pmatrix}
\tilde{A}_1^{(0)} & \tilde{A}_0^{(0)} \\
\tilde{A}_1^{(1)} & \tilde{A}_0^{(1)} \\
& \ddots \ddots \\
\tilde{A}_2^{(N-1)} & \tilde{A}_1^{(N-1)} & \tilde{A}_0^{(N-1)} \\
\tilde{A}_2^{(N)} & \tilde{A}_1^{(N)} & \tilde{A}_0^{(N)}
\end{pmatrix},$$

where

$$\tilde{A}_1^{(i)} = \begin{pmatrix}
-(\alpha_c + \alpha_m + \alpha_b) & \alpha_c & \alpha_m & 0 & \alpha_b \\
-(\lambda + \mu) & -(\alpha_m + \alpha_b) & 0 & \alpha_m & \alpha_b \\
\beta_c & -(\alpha_m + \alpha_b) & -(\lambda + \mu + \beta_c) & \alpha_c & \alpha_b \\
\beta_m & 0 & -(\alpha_c + \alpha_b) & -(\lambda + \mu + \beta_m) & \alpha_c & \alpha_b \\
0 & \beta_m & \beta_c & -(\alpha_b + \lambda + \mu) & -\beta_b & \alpha_b \\
0 & \beta_b & 0 & 0 & -\beta_b & \beta_b
\end{pmatrix},$$
for \( 1 \leq i \leq N - 1 \),

\[
\tilde{A}_1^{(0)} = \begin{pmatrix}
-(\alpha_c + \alpha_m + \alpha_b)
& \alpha_c
& \alpha_m
& 0
& \alpha_b \\
-\lambda
& -\alpha_m + \alpha_b
& 0
& \alpha_m
& \alpha_b \\
\beta_c
& -\alpha_m + \alpha_b
& 0
& \alpha_m
& \alpha_b \\
-\lambda + \beta_c
& 0
& -(\alpha_c + \alpha_b)
& \alpha_c
& \alpha_b \\
0
& \beta_m
& \beta_c
& -(\beta_c + \lambda)
& \alpha_b \\
\beta_b
& 0
& 0
& 0
& -\beta_b
\end{pmatrix},
\]

\[
\tilde{A}_1^{(N)} = \begin{pmatrix}
-(\alpha_c + \alpha_m + \alpha_b)
& \alpha_c
& \alpha_m
& 0
& \alpha_b \\
-\mu
& -\alpha_m + \alpha_b
& 0
& \alpha_m
& \alpha_b \\
\beta_c
& -(\alpha_m + \alpha_b)
& 0
& \alpha_m
& \alpha_b \\
-\mu + \beta_c
& 0
& -(\alpha_c + \alpha_b)
& \alpha_c
& \alpha_b \\
0
& \beta_m
& \beta_c
& -(\beta_c + \mu)
& \alpha_b \\
\beta_b
& 0
& 0
& 0
& -\beta_b
\end{pmatrix},
\]

\[
\tilde{A}_2^{(j)} = \begin{pmatrix}
\mu
& \mu
& \mu
& 0
\end{pmatrix}, \quad 1 \leq j \leq N,
\]

\[
A_0^{(k)} = \begin{pmatrix}
\lambda
& \lambda
& 0
\end{pmatrix}, \quad 0 \leq k \leq N - 1.
\]

Note that \( Q_1 \) is the infinitesimal generator of the QBD process for the email attack model, as seen in (1). For simplicity of description, \( Q_1 \) is a special case for the matrix given in (1) when we use exponential distributions instead of phase type distributions. We give the detailed description of these exponential distributions as follows. For an attack involving cracking the password of an email account, a customer remains in the email account with a period \( x_c \) that is exponential with mean \( 1/\beta_c \); and the attack type involving sending malicious emails remains in the email system for a time length \( x_m \) which is exponential with mean \( 1/\beta_m \). \( Q_2 \) is the perturbed matrix. In what follows, we describe how to construct the matrix \( Q_2 \) according to the assumption of being in a practical email system. For example, let \( \lambda(\epsilon) = \lambda + \epsilon, \epsilon > 0 \), be the perturbed parameter. It is easy to check that

\[
Q_2 = \begin{pmatrix}
B_0^{(0)} & B_0^{(1)} & \cdots & B_0^{(N-1)} \\
0 & B_1^{(0)} & \cdots & B_1^{(N-1)} \\
\vdots & \vdots & \ddots & \vdots \\
0 & B_{N-1}^{(0)} & \cdots & B_{N-1}^{(N-1)} \\
0 & 0 & \cdots & 0
\end{pmatrix},
\]

(13)

where,

\[
B_0^{(k)} = B_1^{(k)} = \begin{pmatrix}
-1 & -1 \\
-1 & -1 \\
0 & 0
\end{pmatrix}, \quad 0 \leq k \leq N - 1.
\]

We use the algorithmic approach for the sensitivity analysis provided in [17]. Note that this algorithmic approach is based on LU-type RG-factorization. To this end, let \( \eta(\epsilon) \) be a stationary performance metric of the perturbed QBD process as follows:

\[
\eta(\epsilon) = \sum_{i=0}^N \pi_i(\epsilon) f_i,
\]
where \( \pi(\epsilon) = (\pi_0(\epsilon), \pi_1(\epsilon), \pi_2(\epsilon), \ldots, \pi_N(\epsilon)) \), partitioned according to the levels, is the stationary probability vector of the perturbed QBD process \( Q(\epsilon) \), and \( f' = [f'_0, f'_1, f'_2, \ldots, f'_N]^T \); the superscript \( T \) denotes matrix transpose and \( f_i \) is a real column vector of size \( m_i \) for \( i \geq 0 \).

The performance metric \( \eta(\epsilon) \) is a general index of the email system. Note that \( m_i = 5 \) for \( 0 \leq i \leq N \) in this model. We consider some interesting examples. When \( f_i = (1, 1, 1, 0, 0), 0 \leq i \leq N \), \( \eta(\epsilon) \) is the performance metric of system availability \( A \) as defined in [4]. When \( f_i = (1, i, i, i, i) \) for \( 0 \leq i \leq N \), \( \eta(\epsilon) \) is the average queue length \( AQL \) as given in (6).

We are interested in the sensitivity analysis metric

\[
\eta^{(1)} = \frac{d}{d\epsilon} \eta(\epsilon)|_{\epsilon=0},
\]

From Theorem 6 in [17, pp. 377], we have

\[
\eta^{(1)} = \pi(0)Q_2(-Q_1^{-1})f' = \pi(0)Q_2(I - \widetilde{G}_U)^{-1}(-\widetilde{U}_D)(I - \widetilde{R}_k)^{-1}f',
\]

where \( \widetilde{G}_U \) and \( \widetilde{U}_D \) are the \( RU \)-factorizations of matrix \( Q_1 \), and \( \pi(0) \) is the stationary probability vector of \( Q_1 \).

Based on the \( LU \)-type R-measure \( \{\tilde{g}_k, 1 \leq k \leq N\} \) and \( G \)-measure \( \{G_k, 0 \leq k \leq N - 1\} \), we write

\[
X^{(l)}_k = \tilde{g}_k \tilde{g}_{k-1} \tilde{g}_{k-2} \cdots \tilde{g}_{k-k+1}, \quad 1 \leq k \leq l \leq N,
\]

and

\[
Y^{(l)}_k = G_k G_{k-1} G_{k-2} \cdots G_{k-k-1}, \quad 1 \leq k + l \leq N, \quad 0 \leq k \leq N - 1.
\]

Thus, it follows from Theorem 5 in [17] that

\[
-Q_1^{-1} = (I - \widetilde{G}_U)^{-1}(-\widetilde{U}_D)(I - \widetilde{R}_k)^{-1} = (\psi_{m,n})_{0 \leq m,n \leq N}
\]

where

\[
\psi_{m,n} = \begin{cases} 
U_{m,n}^{-1}, & m = n = N, \\
U_{m,n}^{-1}X^{(N)}_{N-n}, & m = N, 0 \leq n \leq N - 1, \\
Y^{(m)}_{n-m} U_{m,n}^{-1} Y^{(m)^{\dagger}}_{1+m}, & 0 \leq m \leq N - 1, m + 1 \leq n \leq N, \\
U_{m,n}^{-1} + \sum_{i=1}^{N-m} Y^{(m)}_{i} U_{i,m} Y^{(i+m)}_{1+i+m}, & 0 \leq m = n \leq N - 1, \\
U_{m,n}^{-1} X^{(m)}_{n-m} + \sum_{i=1}^{N-m} Y^{(m)}_{i} U_{i,m} X^{(i+m)}_{1+i+m-n}, & 0 \leq m \leq N - 1, 0 \leq n \leq m - 1.
\end{cases}
\]

Now, we provide numerical examples for the sensitivity analysis of the email attack model. We illustrate the how the derivatives of the performance metrics depend on the parameters of the system, for example, \( \lambda \). We discuss three cases: when \( f_i = (1, 1, 1, 1, 0), (i, i, i, i, 0) \) or \( (0, P_C, P_M, P_{CM}, 0) \), \( 0 \leq i \leq N \). In these three cases, \( \eta(\epsilon) \) stands for the performance metric Availability \( A(\epsilon) \), Average Queue Length \( AQL(\epsilon) \) and Information Leakage Number \( ILN(\epsilon) \), respectively. Based on the definition of Information Leakage Probability \( ILP(\epsilon) \) in (8), we can derive the first derivative of \( ILP(\epsilon) \):

\[
\frac{d}{d\epsilon} ILP(\epsilon)|_{\epsilon=0} = \frac{d}{d\epsilon} ILN(\epsilon)|_{\epsilon=0} \cdot AQL(0) - ILN(0) \cdot \frac{d}{d\epsilon} AQL(\epsilon)|_{\epsilon=0},
\]

where \( AQL(0) \) and \( ILN(0) \) are the performance metrics of the QBD process \( Q_0 \).

Fig. 13 shows the first-order derivatives for the three performance metrics, \( A(\epsilon), AQL(\epsilon), ILP(\epsilon) \) depending on \( \lambda \). Let \( N = 10, \alpha = 0.1, \beta = 0.2, \gamma = 0.2, \rho = 0.3, \omega = 0.01, \beta = 0.05, P_C = 0.8, P_M = 0.5 \) and \( P_{CM} = 0.9 \). We study \( \frac{d}{d\epsilon} ILP(\epsilon)|_{\epsilon=0}, \frac{d}{d\epsilon} A(\epsilon)|_{\epsilon=0} \) and \( \frac{d}{d\epsilon} AQL(\epsilon)|_{\epsilon=0} \) with different values of \( \mu \). From the figure, we can see that \( A(\epsilon) \) and \( AQL(\epsilon) \) are greatly affected by the perturbed \( \lambda \), especially when \( \mu \) is getting larger. And \( ILP(\epsilon) \) is not sensitive in this model.

7. Concluding remarks

In this paper, we have provided a novel queuing model to study the security attributes for email systems under attack. We use four parallel but different queues in a composite queuing system to model the ordinary email traffic and three types of email attacks. By using the \( RU \)-factorization given in [13, 14] to solve the corresponding Markov chain which is an irreducible continuous-time QBD process, we calculate the security performance metrics for the impact of the three types of attack on an email system. Meanwhile, we provide some numerical examples to indicate that the method described in this paper is effective and efficient for analyzing email system security.
Fig. 13. The first-order derivatives of performance metrics depending on $\lambda$.

We hope that the research insights gained in this paper will lead to more theoretical works on email security, which is an increasingly crucial issue in our daily life. The method proposed in this paper can be used to deal with other types of attacks, such as denial of service attack, man-in-the-middle attack and so on, and also spam in network systems. There are still some interesting issues to be further considered. First, the model could be extended with more general arrival processes and processing times that are closer to reality. And we would like to use practical data to refine the model. Second, the importance of different email information needs to be well considered, because the leakage of more crucial information could lead to more serious damage of email systems. Furthermore, email accounts and the attackers are actually regarded as some form of games; thus game theory models should be introduced in studies of email system security. Therefore, it is necessary to show the Nash strategies among the email account and the different attackers.

Acknowledgements

The constructive comments of the three referees on an earlier version of this paper are greatly appreciated. The work of Wang and Lin was supported in part by the National Natural Science Foundation of China under Grant Nos. 60932003 and 60803123 and the National Grand Fundamental Research 973 Program of China under Grant Nos. 2010CB328105 and 2009CB320504. The work of Li was supported in part by the National Natural Science Foundation of China under Grant Nos. 10671107, 10871114 and 60736028, and the National Grand Fundamental Research 973 Program of China under Grant No. 2006CB805901.

References

Yang Wang is an Assistant Professor in the Department of Computer Science and Technology, Peking University, Beijing, PR China. She received her Ph.D. degree from the Department of Computer Science and Technology, Tsinghua University in 2009. Her research interests include network security, system performance evaluation, wireless networks and communication networks etc.

Chuang Lin is a professor in the Department of Computer Science and Technology, Tsinghua University, Beijing, China. He is a Honorary Visiting Professor, University of Bradford, UK. He received his Ph.D. degree in Computer Science from Tsinghua University in 1994. His current research interests include computer networks, performance evaluation, network security analysis, and Petri net theory and its applications. He has published more than 300 papers in research journals and IEEE conference proceedings in these areas and has published four books.

Quan-Lin Li is a Full Professor in the School of Economics and Management Sciences, Yanshan University, Qinhuangdao, PR China. He received his Ph.D. degree from the Institute of Applied Mathematics, Chinese Academy of Sciences, Beijing, PR China, in 1998. His research interests include stochastic models, queueing systems, block-structured Markov processes, stochastic process algebra, automotive supply chains, computer networks, network security and RFID technique and applications. He has published a book “Constructive Computation in Stochastic Models with Applications: The RG-Factorizations” and over 30 papers in international journals, such as Advances in Applied Probability, Queueing Systems, Stochastic Models, Computer Networks, Performance Evaluation, International Journal of Production Economics and Computer and Mathematics with Applications.