On Generalized Low-Rate Denial-of-Quality Attack Against Internet Services

Yajuan Tang[†], Xiapu Luo[‡], Qing Hui[§], Rocky K. C. Chang[‡]
[†]Department of Electronic Engineering, Shantou University, China
[‡]Department of Computing, The Hong Kong Polytechnic University, Hong Kong SAR, China
[§]Department of Mechanical Engineering, Texas Tech University, USA
yjtang@stu.edu.cn, {csxluo,csrchang}@comp.polyu.edu.hk, qing.hui@ttu.edu

***Abstract*—Low-rate Denial of Quality (DoQ) attacks, by sending intermittent bursts of requests, can severely degrade the quality of Internet services and evade detection. In this paper, we generalize the previous results by considering arbitrary attack intervals. We obtain two sets of new results for a web server with feedback-based admission control. First, we model the web server under the attack as a switched system. By proving the Lyapunov and Lagrange stability of the system, we show that the admission rate can always be throttled to a bounded low value. Second, we investigate the worst impacts of a DoQ attack by optimizing a utility function for the attacks. As a result, we obtain for the first time optimal attack patterns for both periodic and aperiodic attacks. Extensive simulation results also agree with the analytical results.**

I. Introduction

The Denial-of-Quality (DoQ) attack has been shown effective in degrading the quality of Internet service and evading detection [1]. The DoQ attack exploits the adaptation mechanism employed in a victim service by sending intermittent bursts of requests to cause a transient overloading on that service. Although the request arrival rate within a burst is high, the average arrival rate of such DoQ attacks is low. Therefore, the DoQ attack is also considered as a low-rate attack.

However, the analysis (and the understanding) of the DoQ attacks is so far limited to constant attack intervals (i.e., a constant time between two successive attack bursts). Such periodic attacks can also be easily detected [2], [3]. Therefore, a more sophisticated attacker could randomize the attack intervals to evade the detection while achieving a certain level of damage to a victim service. Motivated by this, we therefore investigate in this paper the impact of DoQ attacks with arbitrary attack intervals (i.e., aperiodic attacks), with the periodic attack as a special case.

Following the Reduction of Quality (RoQ) attack [1], we use a web server's admission control mechanism as an example to address two fundamental issues for the generalized DoQ attack:

1. Given a target victim, is it always possible to launch a DoQ attack against it?
2. If it is possible, what are the worst impacts of the periodic and aperiodic attacks given a utility function for the attacks?

Modeling the DoQ attacks with arbitrary attack intervals is very challenging. The main difficulty is that the victim's system states exhibit discontinuities at the arrivals of attack bursts. As a result, the classical control theory could not be applied to model the victim system. In our approach, we model it as a switched system, which is a hybrid system involving a series of subsystems and changes at discrete times. Modeling the victim as a switched system enables us to tackle the two aforementioned fundamental issues. We summarize our main contributions below:

1) We have modeled a web server with a PI controller for admission control under a DoQ attack using the switched system theory. By proving the Lyapunov and Lagrange stability of the switched system, we have shown that the victim's admission rate can always be throttled to a bounded low value.
2) We have formulated attack optimization problems for both periodic and aperiodic attacks. By solving them using a gradient-based algorithm and a particle swarm optimization method, we have obtained the worst attack impacts. The impact of the aperiodic attack is found to be more severe than that of the periodic attack.

II. Modeling the DoQ Attack

A. A DoQ attack model

To launch a DoQ attack, an attacker sends a sequence of request bursts to a victim web server. We denote the attack as $\mathbb{A} = \{\Lambda, \tau, N\}$, where $\Lambda = (\lambda_{a_1}, \lambda_{a_2}, \dots, \lambda_{a_N})^T$ indicates the arrival rate of each burst, $\tau = (\tau_1, \tau_2, \dots, \tau_N)^T$ denotes the arrival time of each burst, and N is the number of bursts in the sequence. Here, we assume $\lambda_{a_i} = \lambda_a$ (a constant) to simplify the analysis. In the following we refer $\tau = (\tau_1, \tau_2, \dots, \tau_N)^T$ to as an attack sequence.

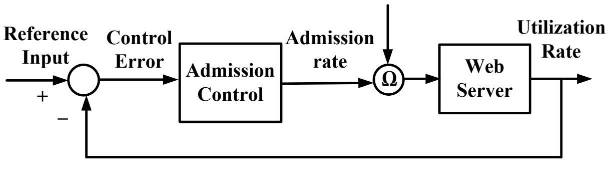


Fig. 1: The admission control of a web server.

B. A web server model

Following [1], the model of a web server's admission control mechanism is illustrated in Figure 1. The admission system is based on the feedback control. The server's utilization rate is fed back and compared with a reference value. The control error indicates the difference between the measured utilization rate and the reference value. The admission controller bases on the control error to determine the admission rate. Table I summarizes the main notations used in this paper.

TABLE I: The main notations used in this paper.

	Description
$\alpha(\cdot)$	admission rate
$\rho(\cdot)$	utilization rate
ρ^*	desired utilization rate
α^*	admission rate when $\rho = \rho^*$
$n_m(\cdot)$	number of waiting requests
$n_p(\cdot)$	number of backlogged requests
μ	service rate
A, B, C, D, ℓ	constants for determining $\rho(\cdot)$
$\lambda(\cdot)$	the total arrival rate
λ_n	arrival rate of normal requests
λ_a	attack intensity
K	controller's parameter
τ_i	arrival time of the ith attack pulse
N	total number of attack pulses

The system uses a proportional-integral (PI) controller to adjust the admission rate α:

$$\dot{\alpha}(t) = K(\rho^* - \rho(t)), \quad \alpha \in [0, 1]. \tag{1}$$

The number of requests (i.e., $n_m(t)$) waiting for service consists of two parts: newly admitted requests (i.e. $\lambda(t)\alpha(t)$) and backlogged requests (i.e., $n_p(t)$):

$$n_m(t) = \lambda(t)\alpha(t) + n_p(t), \quad n_m \in [0, +\infty). \tag{2}$$

The backlogged requests are those admitted requests that cannot be processed in $[t-1, t)$. The number of new requests arriving in period $[t-1, t)$ is $\alpha(t-1) \times \lambda(t-1) \times 1$ and the number of backlogged requests is

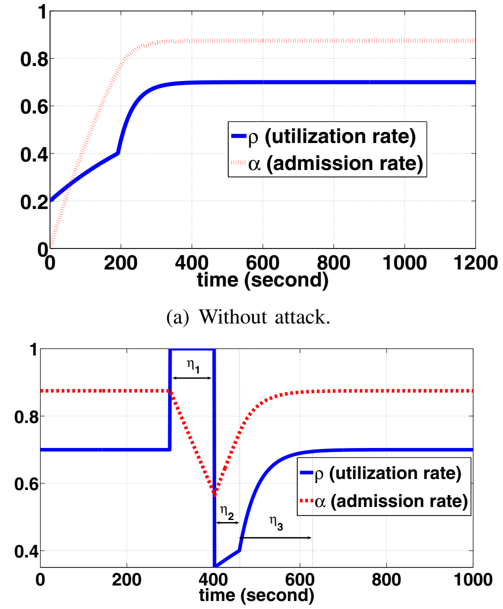
$$n_p(t) = \lambda(t-1)\alpha(t-1) - \mu(t-1), \quad n_p \in [0, +\infty). \tag{3}$$

Moreover, ρ is estimated from the queue length by

$$\rho(t) = \begin{cases} An_m(t) + B & \text{if } n_m(t) < \ell \\ Cn_m(t) + D & \text{if } n_m(t) \geq \ell \end{cases}, \quad \rho \in [0, 1]. \tag{4}$$

The trajectories of ρ and α in the absence of attack are illustrated in Figure 2(a). The system starts from an initial condition where $\alpha = n_m = n_p = 0$ and $\rho = 0.2$. After elapsing for 300 seconds α and ρ have reached constant values, and we refer this stage to as steady state.

When an attack burst arrives at time t, the total arrival rate of requests at t is $\lambda(t) = \lambda_n + \lambda_a$. To ease the analysis, we assume λ_n, λ_a and μ (the service rate) are constant. The server therefore admits $(\lambda_n + \lambda_a)\alpha(t)$ requests. If λ_a is large enough, it will increase $n_m(t)$ and consequently cause $\rho(t) = 1$ (i.e., fully utilized). In this case, the server enters the saturated state and remains there until most of backlogged requests are processed. After that, ρ will first decrease (as n_m decreases) and then increase (as α increases). Accordingly, the server enters the recovery stage. As ρ has two increasing rates governed by constants A and C, the recovery stage can be further divided into two distinct stages. Figure 2 illustrates the four stages for $\rho(t)$ and $\alpha(t)$, where η_1, η_2 and η_3 are respective durations for the saturated stage, and the recovery stages one and two.



(a) Without attack.

(b) An attack burst arrival at 300th second.

Fig. 2: Trajectories of the admission rate and utilization rate.

C. A model of a web server under DoQ attacks

Since the web server's behavior in the presence of a DoQ attack consists of a set of state changes, we model these state changes using the switched system theory [4].

We first consider a family of subsystems:

$$\dot{x} = f_p(x), \tag{5}$$

where $x \in \mathbb{R}^n$, $p \in \mathcal{P}$ ($\mathcal{P}$ is finite), and for each $p \in \mathcal{P}$, f_p is Lipschitz continuous. Next, we consider the switched system

$$\dot{x} = f_\sigma(x), \tag{6}$$

where $\sigma : [0, \infty) \to \mathcal{P}$ is a piecewise constant switching signal, continuous from the right. We denote by t_i, $i = 1, 2, \ldots$, the consecutive discontinuities of σ called the switching times. Here, we assume that if there are infinitely many switching times, there exists a $\tau > 0$, such that for every $T \geq 0$, one can find a positive integer i for which $t_{i+1} - \tau \geq t_i \geq T$. For $t \in [t_k, t_{k+1})$ and $\sigma(t) = i_k$, the i_kth subsystem is active. Hence, the trajectory $x(t)$ of the switched system (6) is defined as the trajectory $x_{i_k}(t)$ of the i_kth subsystem for $t \in [t_k, t_{k+1})$. An equilibrium point of (6) is a point $x_{\mathrm{e}} \in \mathbb{R}^n$ satisfying $f_p(x_{\mathrm{e}}) = 0$ for all $p \in \mathcal{P}$.

We rewrite the web server model as a set of functions of $\alpha(t)$:

$$\dot{\alpha}(t) = \begin{cases} K(\rho^* - A\lambda\alpha(t) - B) \\ K(\rho^* - C\lambda\alpha(t) - D) \\ K(\rho^* - 1) \\ K(\rho^* - 0) \\ 0 \end{cases} \tag{7}$$

whose switching time is controlled by attack sequence, switching law is determined by current status of the system. When

$\dot{\alpha}(t) = 0$ for $t \in [t_k, t_{k+1})$, $\alpha(t)$ is a constant. Since this case only inflates the convergence time without affecting the stability of the switched system, we can ignore $\dot{\alpha}(t) = 0$ when calculating equilibrium point for stability analysis.

It is worth noting from Figure 2 that there are two discontinuities for $\rho(t)$: the point separating η_1 and η_2, and the point separating η_2 and η_3. Let

$$0 < s_1 < s_2 < \cdots < s_i < \cdots \tag{8}$$

denote the time instants of switching events $n_m(s_i) = \ell$ and η_k^1 indicate the duration of the saturated stage (η_1 in Figure 2) in $t \in [\tau_k, \tau_{k+1})$. Then, for $t \in [\tau_k, \tau_{k+1})$,

$$\tilde{\rho}(t) = \begin{cases} 1, & \tau_k \le t < \tau_k + \eta_k^1, \\ A\lambda\alpha(t), & \tau_k + \eta_k^1 \le t < s_i, \\ C\lambda\alpha(t), & s_i \le t < \tau_{k+1}, \end{cases} \tag{9}$$

if $n_m(s_i) = \ell$ and $\tau_k < s_i < \tau_{k+1}$, or

$$\tilde{\rho}(t) = \begin{cases} 1, & \tau_k \le t < \tau_k + \eta_k^1, \\ A\lambda\alpha(t), & \tau_k + \eta_k^1 \le t < \tau_{k+1}, \end{cases} \tag{10}$$

if $n_m(t) < \ell$ for all $t \in [\tau_k, \tau_{k+1})$.

If the attack sequence satisfies $\tau_{k+1} - \tau_k \le \eta_k^1$ for all $k = 1, 2, \ldots$, then

$$\widetilde{\rho(t)} \equiv 1 \tag{11}$$

for all $t \ge 0$, implying that the admission rate keeps decreasing by $\dot{\alpha} = K(\rho^* - 1) < 0$. Since $\alpha \in [0, 1]$, it follows that $\alpha(t)$ converges to 0 asymptotically. In this case, the admission rate cannot converge to a targeted value unless this value is 0. We further assume that an attacker can let a subsequence $\{\tau_{k_i}\}$ of the attack sequence $\{\tau_k\}$ fulfill that $\tau_{k_i+1} - \tau_{k_i} > \eta_{k_i}^1$ for all $k = 1, 2, \ldots$. Moreover, if the attack sequence is an infinite sequence, then this subsequence is infinite as well.

Hence, the hybrid model for (7) is given by

$$\dot{\alpha}(t) = K(\rho^* - \mathrm{sat}(\tilde{\rho}(t))), \tag{12}$$

where

$$\mathrm{sat}(x) = \begin{cases} 1, & x > 1, \\ x, & 0 \le x \le 1, \\ 0, & x < 0. \end{cases} \tag{13}$$

This hybrid model is a switched system with state-dependent switching s_i and time-dependent switching τ_k. The switching time t_k of (6) can be expressed as $t_k = \tau_k$ and $t_{k+1} = \tau_{k+1}$ if there is no s_i between τ_k and τ_{k+1}, or $t_k = \tau_k$, $t_{k+1} = s_i$, and $t_{k+2} = \tau_{k+1}$ if there exists a s_i between τ_k and τ_{k+1}.

D. Stability Analysis

Due to the limited space, we report only the main stability results. Other details omitted here will be reported in a forthcoming paper.

Proposition 1: Assume that there exists a subsequence $\{\tau_{k_i}\}$ of the attack sequence $\{\tau_k\}$, such that

$$\alpha(\tau_{k_i}) < \frac{K}{2}(1 - \rho^*)(\tau_{k_i+1} - \tau_{k_i}) + \frac{\mu}{\lambda} \tag{14}$$

for all $i = 1, 2, \ldots$ and

$$\alpha(\tau_k) \ge \frac{K\gamma}{2}(1 - \rho^*) + \frac{\mu}{\lambda} \tag{15}$$

for all $k = 1, 2, \ldots$, where $\gamma > 0$ is a constant. Moreover, if the attack sequence is an infinite sequence, then this subsequence is infinite as well. Next, assume that if there are infinitely many switching times for (12), there exists a $\kappa > \varrho + \eta_2$, such that for every $T \ge 0$ one can find a positive integer i for which $\tau_{i+1} - \kappa \ge \tau_i \ge T$, where

$$\varrho = \frac{2\lambda - 2\mu}{\lambda K(1 - \rho^*)}, \quad \eta_2 = -\frac{1}{KA\lambda} \ln \frac{AN + B - \rho^*}{A\mu + B - \rho^*}. \tag{16}$$

Finally, assume that for every $k = 1, 2, \ldots$, holds. Then the switched system (i.e. (12)) is Lyapunov stable on $[0, 1]$.

Proposition 1 proves the Lyapunov stability of the switched system. It also points out that an attacker can determine the upper and lower bounds of admission rate according to (14) and (15) respectively.

Proposition 2: Assume that there exists a subsequence $\{\tau_{k_i}\}$ of the attack sequence $\{\tau_k\}$, such that $\tau_{k_i+1} - \tau_{k_i} > \eta_{k_i}^1$ for all $i = 1, 2, \ldots$. Moreover, if the attack sequence is an infinite sequence, then this subsequence is infinite as well. Next, assume that if there are infinitely many switching times for (12), there exists a $\tau > 0$, such that for every $T \ge 0$ one can find a positive integer i for which $t_{i+1} - \tau \ge t_i \ge T$. Then the switched system (i.e. (12)) is Lagrange stable. Furthermore,

$$\alpha(t) \le \varepsilon, \tag{17}$$

where

$$\varepsilon = \max_{p \in \mathcal{P}} (\alpha_{\mathrm{ep}} + |\alpha(0) - \alpha_{\mathrm{ep}}|) \le \max_{p \in \mathcal{P}} (2\alpha_{ep} + 1), \tag{18}$$

and α_{ep} is an equilibrium point.

Proposition 2 proves that the switched system is Lagrange stable. It also shows that there exists an attack sequence that can force α to be less than ε, which is determined by the equilibrium point and the initial condition.

III. Optimizing the DoQ Attack

The cost of launching a DoQ attack could be measured by the number of attack requests involved in the attack. Sending more attack requests is more vulnerable to detection and also requires more resources. A sophisticated attacker would like to maximize the attack damage and reduce the attack cost simultaneously by manipulating the attack sequence.

To obtain optimal attack sequences, we first define the attack damage Γ as the percentage of normal requests dropped by the victim server due to a DoQ attack:

$$\Gamma = \frac{\text{Number of rejected normal requests}}{\text{Number of normal requests}} \tag{19}$$

On the other hand, the cost of a DoQ attack, denoted by γ, is a function of the arrival rate of the attack requests.

Besides the damage and cost, an attacker's willingness of taking risk also affects his utility function [5]. A conservative attacker may send out bursts of requests with large intervals to avoid detection. However, an aggressive attacker may prefer to inflict more damage by decreasing the attack intervals. Therefore, we use $(1 - \gamma)^\kappa$ to measure an attacker's risk preference, where κ is a risk index of an attacker. If an attacker is willing to take more risk, κ is small; otherwise, κ is large.

We next formulate an optimization problem to maximize an attacker's utility function:

$$\max_{\bar{\tau}^*} J = \Gamma \times (1-\gamma)^{\kappa}, \tag{20}$$

where $\bar{\tau}^* = (\tau_1^*, \tau_2^*, \ldots, \tau_k^*, \ldots)^T$ is the optimal attack sequence. We first investigate optimal periodic DoQ attack and propose a nonlinear optimization algorithm to solve it. Then we extend our discussions to aperiodic DoQ attacks and propose a particle swarm optimization algorithm to obtain the optimal result. Generally an attacker can gain a higher utility by launching an aperiodic attack than a periodic attack.

A. Optimizing Periodic Attacks

We proved in [6] that the web server's admission rate converges to periodic behavior under a periodic DoQ attack. Here, we assume the system already converges. Following (19), the damage is represented as the percentage of rejected normal requests in an attack period.

$$\Gamma = \frac{1/T \int_0^T (\alpha^* - \alpha(t))\lambda_n dt}{1/T \int_0^T \alpha^* \lambda_n dt} = 1 - \frac{\int_0^T \alpha(t) dt}{\alpha^* T}, \tag{21}$$

where T is the attack period. Similarly, the attack cost is the normalized number of attack requests in an attack period, defined as

$$\gamma = \frac{\int_0^T \lambda_a \delta(T) dt}{\int_0^T \lambda_a \delta(T) dt + \int_0^T \lambda_n dt} = \frac{\lambda_a}{\lambda_a + \lambda_n T}. \tag{22}$$

Given (21) and (22), the optimization problem for the periodic DoQ attack is:

Problem 1: Given that an attacker sends an infinite number of periodic bursts of requests to a victim web server, find a period T^* that maximizes J:

$$\max_{T^*} J = \Gamma \times (1-\gamma)^{\kappa}, \tag{23}$$

subject to the state evolution described in section II and $T > 0$.

We have solved Problem 1 using a gradient-based nonlinear optimization algorithm.

B. Optimizing Aperiodic Attacks

In this subsection we investigate the problem of optimizing aperiodic DoQ attacks. Suppose there are N bursts sent within a fixed interval $[t_0, t_f]$ with the following attack sequence;

$$t_0 = \tau_0 < \tau_1 < \tau_2 < \ldots < \tau_N < \tau_{N+1} = t_f.$$

The damage is the average number of normal requests dropped during the given time interval $[t_0, t_f]$:

$$\Gamma = \frac{1}{t_f - t_0} \int_{t_0}^{t_f} (\alpha^* - \alpha(t))\lambda_n dt = \int_{t_0}^{t_f} L(x) dt. \tag{24}$$

The attack cost is defined as:

$$\gamma = \frac{\frac{1}{N}\sum_i \int_{\tau_{i-1}}^{\tau_i} \lambda_a \delta(\tau_i) dt / (\tau_i - \tau_{i-1})}{\int_{\tau_{i-1}}^{\tau_i} \lambda_a \delta(\tau_i) dt / \min_{\delta\tau}(\tau_i - \tau_{i-1})}. \tag{25}$$

Note that the denominator in (25) is the maximum arrival rate obtained from all attack intervals. We do not use the average arrival rate, because it may not reflect the true cost of attacker due to the irregular distribution of the attack bursts. Therefore, we use the maximum arrival rate (i.e., the minimal interval between consecutive bursts) to normalize γ to the range of $[0, 1]$.

The optimization problem for the aperiodic DoQ attack is:

Problem 2: Given a fixed time interval $[t_0, t_f]$ and N bursts of requests, find a sequence $\bar{\tau}^* = (\tau_1^*, \tau_2^*, \ldots, \tau_N^*)^T$, such that

$$\max_{\bar{\tau}^*} J = \Gamma \times (1-\gamma)^{\kappa}, \tag{26}$$

subject to the state evolution described in section II, and

$$\tau_1 - \tau_0 \geq T_1, \quad \ldots, \quad \tau_{N+1} - \tau_N \geq T_{N+1}, \tag{27}$$

where T_i, $i = 1, \ldots, N+1$, are predefined non-negative values.

We have solved Problem 2 using a particle swarm optimization (PSO) algorithm which is based on the swarm intelligence approach.

IV. Simulation Results

In this section we present the Matlab simulation results for optimal DoQ attacks. The parameters are the same as those in [1]: $A = 0.00267$, $B = 0.2$, $C = 0.024$, $D = -1.4$, $N = 75$, $w = 100$, and $K = 0.01$. The service rate is $\mu = 90$ requests per second, and the desired utilization rate is $\rho^* = 0.7$. The arrival rate of normal request is 100 requests per second, and that for an attack burst is 1000 requests per second.

Figures 3(a)-3(c) demonstrate the correctness of the optimized periodic DoQ attacks. In each figure, the value of J_0 is obtained for a periodic attack using an optimal value J_{opt}. The figure shows the intersection of the two curves at the maximal value of J_0. To verify the convergence of our optimization algorithm, we set the initial period to different periods varying from 2 to 300 seconds. The result shows that our algorithm converges to the optimal value, regardless of the initial period.

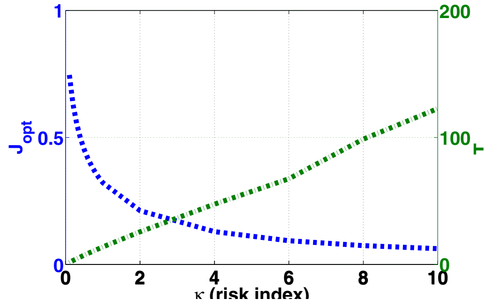
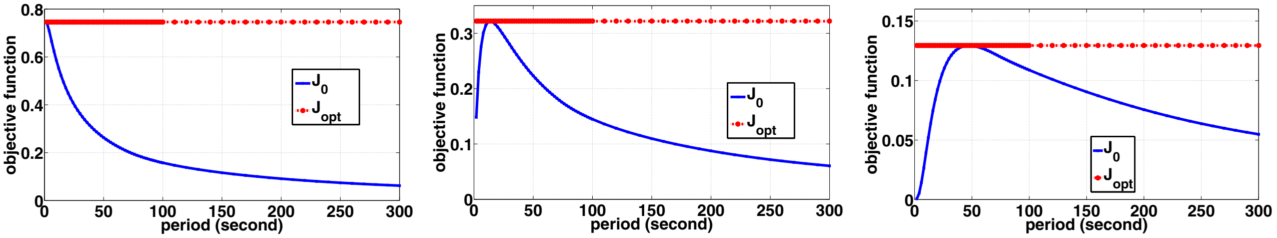


Fig. 4: Impact of the risk index on optimal period and utility function values.

We investigate the impact of κ on T_{opt} and J_{opt} and demonstrate the results in Figure 4, where the left y-axis labels J_{opt} and the right y-axis labels T_{opt}. We can see that J_{opt} decreases and T_{opt} increases with κ. Recall that a larger κ means a less willingness of taking risk. The results are therefore consistent with our analysis in section III.

Figures 5(a)-5(c) illustrate the results of optimized aperiodic DoQ attacks. In these experiments we fixed the observation time to $[0, 1200]$ seconds and generated random attacks as the initial inputs to the PSO algorithm. J_{opt} denotes the optimal result. For the purpose of comparison, we also generated a set of periodic attacks and use $J_{periodic}$ to denote their utility

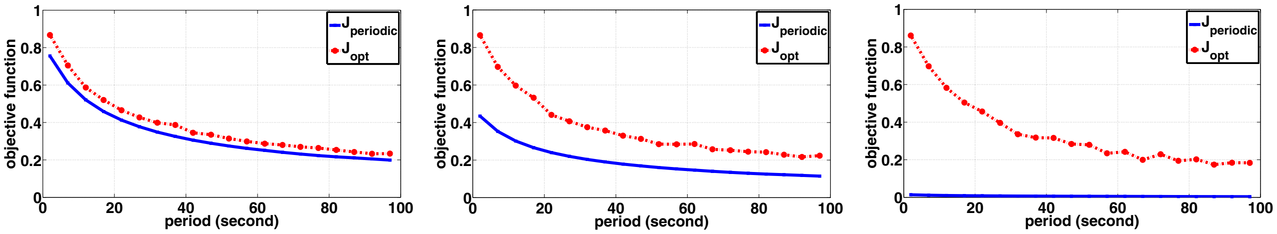


(a) $\kappa = 0.1$.

(b) $\kappa = 1$.

(c) $\kappa = 4$.

Fig. 3: Optimized objective functions of periodic DoQ attacks.



(a) $\kappa = 0.2$.

(b) $\kappa = 1$.

(c) $\kappa = 6$.

Fig. 5: Optimized objective functions of aperiodic DoQ attacks.

function values. Note that in these experiments $J_{periodic}$ is calculated according to (24) and (25). The number of bursts in each experiment is set to $N = t_f/T$, where T is the period of periodic attacks. These figures clearly show that the optimized aperiodic attack is more effective than periodic attacks.

V. Related Work

The Reduction of Quality (RoQ) attacks [1] is the closest to this work in terms of the attack objectives and targets. However, the main difference is that the analysis in [1] was restricted to periodic attacks. Moreover, no stability analysis and attack optimization were performed in [1]. Besides the DoQ attacks, we have investigated the pulsing DoS attacks on Internet's TCP/AQM architecture [7], [8], [5], which could cause significant throughput degradation on TCP flows.

On another front, overload control is an important element in guaranteeing end systems' QoS. Welsh and Culler [9] proposed a feedback loop in Internet service system named SEDA to implement per-stage admission control. Yaksha system [10] adopted a PI controller for admission control, instead of the AIMD algorithm used in SEDA. Other QoS-sensitive overload controls were proposed for web services [11], [12].

VI. Conclusion and Future Work

In this paper, we investigated generalized DoQ attacks on an admission control mechanism used in web service. We have modeled the victim system under DoQ attacks as a switched system. By proving that the switched system is Lyapunov stable and Lagrange stable, for the first time we have showed that a DoQ attack can always throttle the victim system's admission rate to an arbitrary low value. We have also studied the worst impact of the attack by optimizing a utility function of the attack. As a future work, we will apply the methodology developed in this paper to analyzing other kinds of controllers, such as PID. We will also investigate the attack impacts on other overload control models discussed in the last section.

Acknowledgment

The work described in this paper was partially supported by a grant from the Research Grant Council of the Hong Kong SAR, China (Project No. PolyU 5080/02E) and a fund sponsored by Shantou University, China (Project No. Shantou University 140-074049).

References

[1] M. Guirguis, A. Bestavros, I. Matta, and Y. Zhang, "Reduction of quality RoQ attacks on Internet end-systems," in *Proc. IEEE INFOCOM*, 2005.

[2] Y. Chen and K. Hwang, "Collaborative detection and filtering of shrew DDoS attacks using spectral analysis."

[3] H. Sun, J. Lui, and D. Yau, "Defending against low-rate TCP attack: Dynamic detection and protection," in *Proc. IEEE ICNP*, 2004.

[4] W. Haddad and V. Chellaboina and S. Nersesov, *Impulsive and Hybrid Dynamical Systems: Stability, Dissipativity, and Control*. Princeton University Press, 2006.

[5] X. Luo and R. Chang, "Optimizing the pulsing denial-of-service attacks," in *Proc. IEEE/IFIP DSN*, 2005.

[6] Y. Tang, X. Luo, and R. Chang, "Degrading Internet services by intermittent false feedbacks and the countermeasures," in *Proc. IFIP CIP*, 2007.

[7] X. Luo and R. Chang, "On a new class of pulsing denial-of-service attacks and the defense," in *Proc. NDSS*, 2005.

[8] X. Luo, R. Chang, and E. Chan, "Performance analysis of TCP/AQM under denial-of-service attacks," in *Proc. IEEE MASCOTS*, 2005.

[9] M. Welsh and D. Culler, "Adaptive overload control for busy Internet servers," in *Proc. USENIX USITS*, 2003.

[10] A. Kamra, V. Misra, and E. Nahum, "Yaksha: A self-tuning controller for managing the performance of 3-tiered Web sites," in *Proc. IEEE IWQoS*, 2004.

[11] M. K. A. Robertsson, B. Wittenmark and M. Andersson, "Design and evaluation of load control in Web-server systems," in *Proc. ACC*, 2004.

[12] J. Hellerstein, Y. Diao, S. Parekh, and D. Tilbury, *Feedback Control of Computing Systems*. Wiley-IEEE Press, 2004.