IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 21, 2019, accepted June 20, 2019, date of publication July 4, 2019, date of current version December 27, 2019.

*Digital Object Identifier 10.1109/ACCESS.2019.2926925*

# A Theoretical Model for Analysis of Firewalls Under Bursty Traffic Flows

**YAHYA SHAHSAVARI[1,2], HADISHAHRIAR SHAHHOSEINI[2], KAIWEN ZHANG[1], AND HALIMA ELBIAZE[3]**

[1]Department of Software and IT Engineering, École de Technologie Supérieure, Montréal, QC H3C 1K3, Canada
[2]School of Electrical Engineering, Iran University of Science and Technology, Tehran 16846-13114, Iran
[3]Department of Computer Science, Université du Québec á Montréal, Montréal, QC H2L 2C4, Canada

Corresponding author: Hadishahriar Shahhoseini (hshsh@iust.ac.ir)

**ABSTRACT** Firewalls are located at the front line of the network against outside threats. Performance modeling and analysis of network firewalls help to better understand their behavior and characteristics. Moreover, having an analytical model in hand helps firewall designers avoid developing multiple design alternatives and thus considerably reduce the design costs. Moreover, the network administrators can proactively identify the performance bottlenecks of the network and fix them before any malicious attack which targets the network or the firewall itself. In this paper, we propose a novel analytical approach for performance modeling and analysis of network firewalls based on a discrete-time queuing system in which the bursty nature of the incoming traffic is taken into account, where traditional queuing models such as $M/M/1$ model fails to capture peculiar characteristics of the Internet traffic. Throughput, packet loss, delay, and firewalls CPU utilization are employed as performance evaluation indicators in our proposed model. In addition, we introduce a potential DoS attack with a very low rate which can be launched against firewalls with different burstiness factors.

**INDEX TERMS** Network firewalls, performance modeling, discrete-time queuing system, three dimensional Markov chain, burstiness factor, DoS attack.

## I. INTRODUCTION

Firewalls typically are deployed at the entry point of the network and defend against malicious threats and hostile attacks. Firewalls operate by inspecting incoming and outgoing traffic flows using a rule-based engine. This engine matches the packets sequentially with a predefined set of rules, namely access rules, and decide whether to block the packet or not.

Most of the commercial firewalls, especially those deployed in industrial networks have a huge rule-base or Access Control List (ACL) (e.g. Cisco ASA [1]).

Thus, firewalls consume large amounts of resources in the network and spend a significant time for higher-level packet assessment. Since firewalls are usually deployed at the entry point of the network, they can naturally become a performance bottleneck spot in the network. Firewall performance is an important factor in enforcing network security, especially when the network is under attack. These attacks are generally distributed denial of service attacks (DDoS) launched from botnets. From holding up a simple website, to blocking access to an application to make a political statement, DDoS is a growing concern for enterprises, and these threats are expanded in scope and impact. If the firewalls are not well designed to withstand against the mentioned attacks, they may jeopardize the overall security of the network in which they are deployed. Performance modelling and analysis of network firewalls is beneficial for a deeper understanding of their dynamics and behavior. For instance, by having a theoretical model in hand, firewall designers can conduct simulations in order to avoid developing multiple design alternatives before settling on the implementation of the system, thus reducing the design cost. As well, network designers and network administrators can identify the optimal parameters and resource allocation in order to improve the overall performance of the network. Furthermore, we can employ a mathematical model to help the administrators chose the best reaction against the attack. In this paper, we propose the following contributions:

- We develop a mathematical performance model for a network firewall using a discrete-time queuing system.

The associate editor coordinating the review of this manuscript and approving it for publication was Jiafeng Xie.

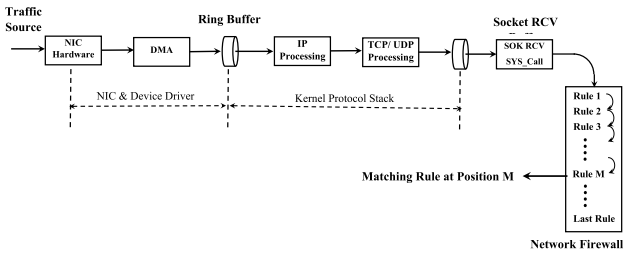



**FIGURE 1.** Path of incoming packets from the NIC to the firewall.

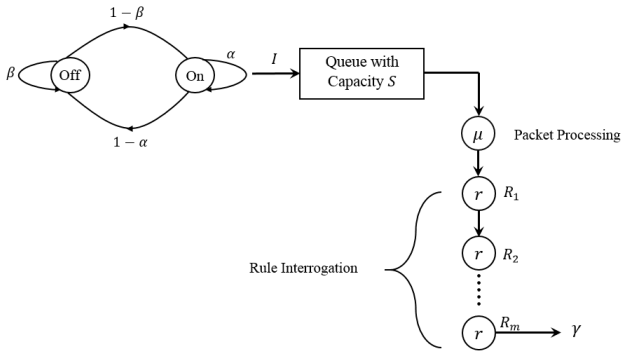



**FIGURE 2.** Single server, finite capacity queuing system with multiple stages of service.

To accomplish this, we derive closed-form equations for throughput, delay, packet loss, and firewall CPU utilization. This model also considers the correlation between arriving packets.

- We extend our model to consider multiple flows. This approach is closer to a realistic scenario where attacks are conducted using botnets and hence come to the network with different throughput and trigger different rules in the firewall's rule-base.
- We use our proposed model to introduce a potential DoS attack against the firewalls. This kind of attacks can be launched with a manipulated burstiness factor in order to attack the firewall with a low rate of DoS flow. We show that it can significantly increase the attack efficiency.
- We evaluate the accuracy of our model by simulating using Matlab and comparing with the results already reported in [2].

The rest of this paper is organized as follows. Section II discusses related works. Our analytical model of a queuing system with correlated packet arrivals which represents the behavior and dynamics of the network firewall under batch arrivals is described in Section III. In Section IV, we propose an analytical solution for performance modelling and analysis of the mentioned firewall. Section V is dedicated to numerical results and comparison of DoS attacks that target a certain rule with different burstiness factors. Finally, Section VI concludes the paper and discusses our future works.

## II. RELATED WORKS

To the best of our knowledge, there are few works on theoretical modeling and analysis of network firewalls. Particularly there are few models for analyzing network firewalls under DoS attacks. The most related work to our contributions is [2] in which a two-dimensional Markov model for network firewall is proposed. In this work, it is assumed that packets arrive in the firewall with a Poisson distribution. But the nature of traffic in today's Internet is bursty with correlated packet flows. Moreover, in [3] it is shown that packet inter-arrival times are not exponentially distributed and are not suitable for modeling as a Poisson process. In a most recent work [4], a performance model for firewalls within a mobile network using queuing theory is proposed. But it suffers from a defect similar to [2]. Without giving an explicit model for the entire firewall system, works such as [5]–[9] discuss optimizing the firewall performance using different approaches. In [10], bottlenecks for system resources such as CPU and memory usage that affect the firewall's performance were studied. An analytical model based on queuing theory and Markov chains for cloud-based firewalls is presented in [11]. As a virtual network function (VNF), the performance of the network firewall for selection of cloud instance is evaluated in [12]. However, to the best of our knowledge, there is no research work which considers the batch arrival of packets.

In this paper, we propose a discrete-time queuing model with constant service time and correlated arrivals. The mathematical logic of such a queuing model is fully described in [13]. Discrete-time queue with Bernoulli bursty source arrival and generally distributed service times is presented in [14].

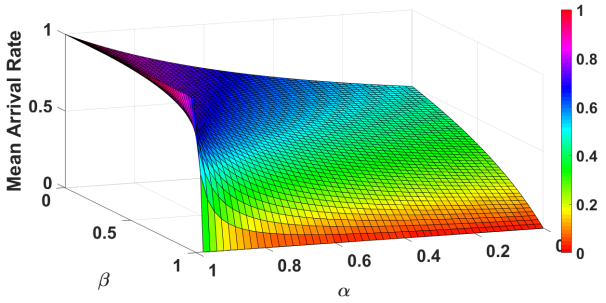
Our analytical model is suitable for analyzing the performance of network firewalls when those are subjected to normal traffic flows as well as DoS traffic with different values of correlation between arriving packets. It is useful to analyze the resiliency of firewalls when those encounter worst-case DoS attacks. In [15], it has been shown how outside attackers can remotely discover firewall rules located at the bottom of a rule-base. Thus, they will be able to launch complex algorithmic DoS attacks [16] targeting the bottom rules in order to efficiently decrease the performance of the firewall rapidly. This methodology allows attackers to do their attack with a very low rate of DoS traffic flow.

## III. SYSTEM MODEL

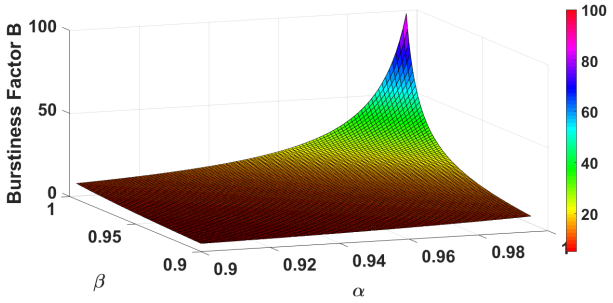
In this section, we present an analytical model for performance modeling and analysis of network firewalls which considers the bursty nature of traffic flows, especially under DoS attacks.

### A. RULE-BASED NETWORK FIREWALL

Incoming packets that arrive in the firewall's Rx NIC (Receiving Network Interface Card) will be queued in the system buffer to be processed in multiple stages. Specifically, in Linux systems, packet processing is done in three stages. As shown in Figure 1, arriving packets are first copied from the NIC to a ring buffer via DMA (Direct Memory access).




(a) Mean arrival rate versus $\alpha$ and $\beta$




(b) Burstiness factor versus $\alpha$ and $\beta$

**FIGURE 3.** Mean arrival rate and Burstiness factor versus $\alpha$ and $\beta$.

Secondly, after IP and TCP/UDP processing, packets are copied from the ring buffer to a socket receive buffer. Finally packets are copied from the socket receive buffer to the firewall's rule-base application.

### B. QUEUING MODEL

This section describes our proposed discrete-time queuing system with correlated packet arrivals, referred to as bursty packet arrivals, and constant service times of arbitrary length. The system is illustrated in Figure 2.

$$M \times L = \frac{m}{r} + \frac{1}{\mu} \tag{1}$$

In this system, incoming packets arrive in a stochastic manner with a mean arrival rate of $I$ packets/time-slot from a Bernoulli bursty source according to a first-order Markovian process. As depicted on the left side of Figure 2, the source alternates between On periods in which there is a packet arrival and Off periods in which there is no packet arrival. This source is called Bernoulli bursty source. Table 1 presents some of the notations needed for understanding this paper. The system has a buffer with a size of $S+1$ packets and queue size of $S$. An incoming packet is first queued in the buffer and then served by the first stage consisting of the kernel's packet processing. In this stage, the mean service time is $1/\mu$. In the next stage, the packet will encounter the firewall rule-base. In this stage, rules are interrogated one by one until there is a matching rule with number $m$. Then a certain action, either allow or block, is performed. The mean time of interrogation for each rule is $1/r$. Packets are served FCFS (First Come, First Served). We assume each packet requires a constant service time of $M$ time-slots in such a way that:
where $L$ is the time length of each time-slot. During an arbitrary time-slot, there can be one packet arrival or no packet arrival. The number of packet arrivals during a time-slot is a random variable dependent on the number of arrived packets during the immediately preceding time-slot. We define two independent parameters $\alpha$ and $\beta$ as follows:

$$\alpha(t) = Prob[N(t) = 1 | N(t-L) = 1] \tag{2}$$

$$\beta(t) = Prob[N(t) = 0 | N(t-L) = 0] \tag{3}$$

where $N(t)$ is the number of packet arrivals during each time-slot. Both of the On period and Off period are geometrically distributed. Therefore, they have mean values of $1/(1-\alpha)$ and $1/(1-\beta)$. We also define the mean arrival rate in steady state $I$ (packet/slot) and burstiness factor $B$ as

**TABLE 1.** Notations used in this paper.

| Metrics | Unit | Description |
|---|---|---|
| $\lambda$ | $Packets/Second$ | Mean packet arrival rate |
| $\alpha(t)$ | – | P [one packet arrival during a slot \| one packet arrival during previous slot] |
| $\beta(t)$ | – | P [no packet arrival during a slot \| no packet arrival during previous slot] |
| $N(t)$ | – | Number of packet arrival during each time-slot |
| $I$ | $Packets/Slot$ | Mean packet arrival rate |
| $B$ | – | Burstiness factor |
| $\gamma$ | $Packet/Second$ | Mean system throughput |
| $1/\mu$ | $Seconds$ | Mean kernel packet processing time |
| $1/r$ | $Seconds$ | Mean rule interrogation time |
| $m$ | – | Targeted rule position number |
| $S$ | $Packets$ | System queue size |
| $M$ | $Slots$ | Mean service time |
| $L$ | $Seconds$ | Time length of each time slot |
| $k$ | – | Slot Number |
| $s_k$ | $Packets$ | Buffer Occupancy |
| $a_k$ | – | Number of Packets entering the system buffer during slot $k$ |
| $r_k$ | – | Number of slots received for service (delivered service time) |
| $p(i,n,j)$ | – | Equilibrium probabilities of the Markovian chain $(r_k, a_{k-1}, s_k)$ |
| $S(z)$ | $Packets$ | Buffer Content |
| $E[s]$ | $Packets$ | Mean buffer occupancy at the start of arbitrary slot |
| $P_loss$ | – | Mean ratio of dropped packets |
| $E[d]$ | $Slot$ | Average packet delay |
| $W$ | $Seconds$ | Average packet delay |
| $\overline{X}$ | $Seconds$ | Mean service time |
| $CPU_{util}$ | % | CPU Utilization |

$$I = \frac{E[T_{on}]}{E[T_{on}] + E[T_{off}]} = \frac{1-\beta}{2-\alpha-\beta} \tag{4}$$

$$B = \frac{E[T_{on}]E[T_{off}]}{E[T_{on}] + E[T_{off}]} = \frac{1}{2-\alpha-\beta} \tag{5}$$

For uncorrelated arriving packet flows, $B$ equals 1. Given the $I$, $B$ varies between $Max(I, 1-I)$ and $\infty$. Figure 3 illustrates the diagram of $I$ and $B$ as functions of $\alpha$ and $\beta$.

When the burstiness factor $B$ approaches to infinite, the denominator in Equation (5) approaches 0. Therefore: $2-\alpha-\beta = 0$ and $2 = \alpha+\beta$. Since $\alpha \leq 1$ always and $\beta \leq 1$, we can conclude $\beta = \alpha = 1$. Also, $1-\alpha = 0$ and $1-\beta = 0$. This implies that when the state of the system

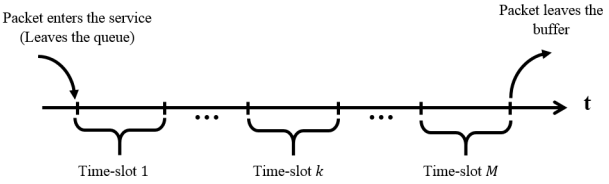



**FIGURE 4.** The time axis is divided into several time-slots with equal lengths.

is On, it remains On and when the state of system is Off, it remains Off. In such conditions, the result of Equation (4) approaches the indeterminate value of 0/0.

This situation will be very sensitive to a slight difference between $\alpha$ and $\beta$. Consider a very small real positive number of $\epsilon$. Let us assume $\{\alpha, \beta | \alpha > \beta, \alpha = \beta + \epsilon\}$, then the result of the fraction of Equation (4) and the steady state probability of having a packet arrival during an arbitrary slot approaches 1. On the other hand, if $\{\alpha, \beta | \beta > \alpha, \beta = \alpha + \epsilon\}$, then the result of the fraction of Equation (4) and the steady state probability of having a packet arrival during an arbitrary slot approaches 0. This fact is readily visible in Figure 3a. Another interesting case is when the burstiness factor $B = 1$. This case occurs when $\alpha + \beta = 1$. This situation refers to uncorrelated packet arrivals with exponential distribution.

## IV. MODEL ANALYSIS AND SOLUTION

In this section, we propose an analytical solution based on a Markov model for performance modeling and analysis of the network firewall.

### A. MARKOV MODEL FOR A RULE-BASED NETWORK FIREWALL

Suppose the time axis is divided into time-slots with equal length ad depicted in Figure 4. We define the random variable $s_k$ as the buffer occupancy (i.e. the total number of packets stored in the buffer including the packet which is under service) at the beginning of time-slot $k$. We also define the random variable $a_k$ as the number of packets entering the system buffer during time-slot $k$. Moreover, let the random variable $r_k$ denote the number of time-slots taken for servicing the packet (i.e. the packet which is currently under service in the buffer) so far. We call $r_k$ the *delivered service time* at the beginning of time-slot $k$. Note that if $s_k = 0$, then $r_k = 0$. Since the service time is assumed constant, $r_k$ can be at most $M - 1$ time-slots. After receiving $M - 1$ time-slots of service, the packet will leave the buffer at the end of next time-slot. Suppose at the beginning of the time-slot $k$, the system is not empty and the delivered service time is less than $M - 1$. Therefore, there is no departure at the end of this time-slot. Also, buffer occupancy is increased by the number of arriving packets during this slot (zero or one) and the delivered service time is increased by one.

As well, suppose the packet which is under service has already received service for $M - 1$ time-slots at the beginning of time-slot $k$. Hence, there will be one departure at the end of time-slot $k$. If the leaving packet is the last packet in the system and no packet arrival occurres during slot $k$, the system becomes empty.

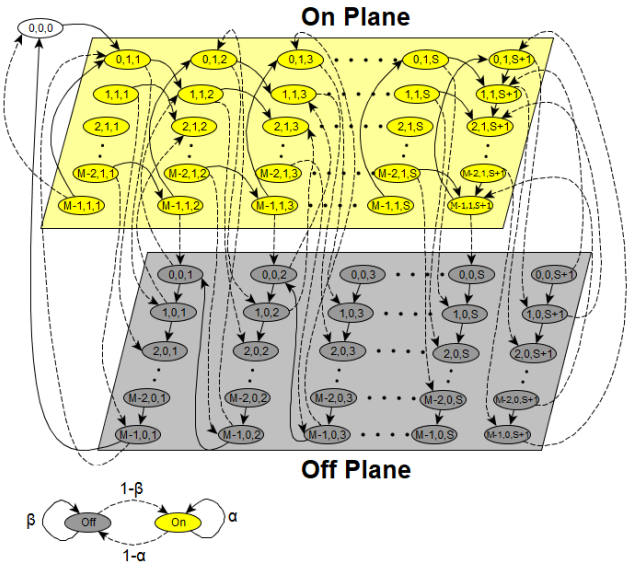



**FIGURE 5.** State transition for a rule-based network firewall with finite buffer capacity $S + 1$.

If the buffer is empty at the start of time-slot $k$, the buffer occupancy at the beginning of the next slot equals to the packets arrived during this slot. Also, $r_{k+1} = 0$.

Whatever we mentioned above, can be summarized in Equations (6) and (7) as follows.

$$s_{k+1} = \begin{cases} \min(s_k + a_k, S + 1) & s_k > 0 \ \& \ r_k \neq M - 1 \\ \min(s_k + a_k, S + 1) - 1 & s_k > 0 \ \& \ r_k = M - 1 \\ a_k & s_k = 0 \ \& \ r_k = 0 \end{cases} \tag{6}$$

and also,

$$r_{k+1} = \begin{cases} r_k + 1 & s_k > 0 \ \ \& \ \ r_k \neq M - 1 \\ 0 & s_k > 0 \ \ \& \ \ r_k = M - 1 \\ 0 & s_k = 0 \ \ \& \ \ r_k = 0 \end{cases} \tag{7}$$

From the terms and equations we described above, we can claim that the vector $(r_k, a_k, s_k)$ is a 3D (three dimensional) Markovian state description of the system at the start of time-slot $k$. Now, it is helpful to define $p(i, n, j)$ as the equilibrium probabilities of the Markovian chain $\{(r_k, a_k, s_k)\}$. The state transition diagram is depicted in Figure 5. The diagram consists of two planes, each referring to *On* and *Off* periods. The state (0, 0, 0) represents the special case when the system is empty. Dashed arrows denote state transitions, at which the state changes from On to Off (with probability $1 - \alpha$) or from Off to On (with probability $1 - \beta$). In such conditions, the system state moves from one plane to another. For the other state transitions, when the system remains On (with probability $\alpha$) or when Off (with probability $\beta$), the state

transitions are shown by solid lines. We define:

$$p(i,n,j) \triangleq \lim_{k\to\infty} Prob[r_k = i, a_{k-1} = n, s_k = j] \quad 0 \le i \le M-1 \quad 0 \le j \le S+1 \tag{8}$$

Now let us define the partial probability generation function of

$$Q_{i,n}(z) \triangleq \sum_{j=0}^{S+1} p(i,n,j) z^j \quad 0 \le i \le M-1 \quad n = 0,1 \tag{9}$$

Also, we define the function $R_n(y,z)$ as

$$R_n(y,z) \triangleq \sum_{i=0}^{M-1} Q_{i,n}(z) y^i \tag{10}$$

Appendix describes an analytical approach to calculate $Q_{i,n}(z)$ and $R_n(y,z)$.

### B. KEY PERFORMANCE MEASURES

The above equations and definitions enable us to compute the buffer content probability generating function (PGF) $S(z)$ as:

$$\begin{aligned} S(z) &= \sum_{j=0}^{S+1} Prob[s=j] z^j = \sum_{j=0}^{S+1} \sum_{i=0}^{M-1} \sum_{n=0}^{1} p(i,n,j) z^j \\ &= \sum_{i=0}^{M-1} \sum_{n=0}^{1} (\sum_{j=0}^{S+1} p(i,n,j) z^j) = \sum_{i=0}^{M-1} \sum_{n=0}^{1} Q_{i,n}(z) \\ &= R_0(1,z) + R_1(1,z) \end{aligned} \tag{11}$$

where $\sum_{i=0}^{M-1} \sum_{n=0}^{1} p(i,n,j)$ yields the marginal probability of $p(i,n,j)$. Regarding the Equations (A.12)-(A.18), the $S(z)$ can be calculated as follows:

$$\begin{aligned} S(z) = {} & \frac{p_0(z-1)}{N(z)} \{ z(\frac{\xi_1^{M+1} - \xi_2^{M+1}}{\xi_1 - \xi_2}) \\ & - z\xi_1\xi_2(\frac{\xi_1^{M} - \xi_2^{M}}{\xi_1 - \xi_2}) - (\xi_1\xi_2)^M \} \\ & + \frac{(z-1)z^{S+1}}{IN(z)} \sum_{1}^{M-1} p_i \{ \xi_1\xi_2[I + (1-I)z] \\ & \times (\frac{\xi_1^{M-i-1} - \xi_2^{M-i-1}}{\xi_1 - \xi_2}) \\ & + \frac{1}{z}[I + (1-I)z](\xi_1\xi_2)^{M-i} (\frac{\xi_1^{i+1} - \xi_2^{i+1}}{\xi_1 - \xi_2}) \\ & - z(\frac{\xi_1^{M-i} - \xi_2^{M-i}}{\xi_1 - \xi_2}) \\ & - (\xi_1\xi_2)^{M-i} (\frac{\xi_1^{i} - \xi_2^{i}}{\xi_1 - \xi_2}) \} + \frac{1}{I} \sum_{i=1}^{M-1} p_i z^{S+1} \end{aligned} \tag{12}$$

where $\xi_1$ and $\xi_2$ are eigenvalues of the matrix $F$

$$F \triangleq \begin{bmatrix} \beta & 1-\alpha \\ (1-\beta)z & \alpha z \end{bmatrix} \tag{13}$$

And $M$ unknown parameters of $p_i$ $(0 \le i \le M-1)$ are defined as follows:

$$p_i = \lim_{k\to\infty} Prob[r_k = i, a_k = 1, s_k = S+1] \tag{14}$$

Note that the parameter $p_0$ is defined as the steady state probability of having an empty buffer at the start of any arbitrary time-slot. $N(z)$ in Equation (12) is given by Equation (A.13). We find unknown parameters of $p_i$ $(0 \le i \le M-1)$ as follows. It can be seen that the denominator and the numerator of $Q_{0,n}(z)$ are polynomial in $z$. Also, both of them disappear for $z = 0$ and $z = 1$. The denominator of $Q_{0,n}(z)$, i.e. $N(z)$, is a polynomial degree $M+1$, which guarantees that it has exactly $M+1$ zeros inside the complex plane. Also we know that $Q_{0,n}(z)$ is an analytic function of $z$ in the whole complex plane. Therefore each zero of the denominator of $Q_{0,n}(z)$ is also a zero of the numerator. Since both of them disappear for $z = 0$ and $z = 1$, we obtain only $M-1$ linear equations for $M$ unknowns. From the Equation (11) and using the normalization condition for $S(1) = 1$, we obtain the $M$-th linear equation:

$$p_0 - M \sum_{i=1}^{M-1} p_i = 1 - MI \tag{15}$$

Consequently, we obtain a linear system of equations with $M$ independent linear equations and $M$ unknowns. Now the system can be solved to obtain the parameters $p_i$ $(0 \le i \le M-1)$. Substituting these parameters in Equation (12) yield $S(z)$ in terms of known values. This enables us to compute the mean buffer occupancy $E[s]$ at the beginning of any arbitrary slot:

$$\begin{aligned} E[s] = S'(1) = {} & \frac{1}{1-MI} \{ \frac{I}{2}((M+1)p_0 - M + 1) \\ & + \frac{1-I-p_0}{2-\alpha-\beta} - \sum_{i=1}^{M-1} p_i [M(S+1) - i - 1] \} \end{aligned} \tag{16}$$

When an incoming packet encounters a full buffer, it will be dropped. Let $P_{loss}$ denote the ratio of dropped packets. A packet will be lost if it arrives at the beginning of a time-slot at which the system is full. Therefore we obtain a strict expression for the packet loss ratio:

$$p_{loss} = \frac{1}{I} \sum_{i=1}^{M-1} p_i \tag{17}$$

$P_{loss}$ is independent of the buffer size $S$. According to Little's result, we can claim:

$$E[d] = \frac{E[s]}{I(1-p_{loss})} \quad (time - Slots) \tag{18}$$

where $E[d]$ is the average packet delay. Note that, to obtain average packet delay in seconds $W$, we calculate:

$$W = E[d] \times L \quad (Seconds) \tag{19}$$

where $L$ is the length of each time-slot in seconds. One of the key performance measures is the system throughput and can be estimated as:

$$\gamma = \frac{I}{L}(1 - p_{loss}) \tag{20}$$

Since the length of each time-slot equals to $L$ seconds, the mean service time can be represented as:

$$\overline{X} = ML = \frac{m}{r} + \frac{1}{\mu} \quad (Seconds) \tag{21}$$

Another performance measure which we are interested in to estimate is the firewall's CPU utilization:

$$CPU_{util} = \gamma \overline{X} = IM(1 - P_{loss}) \tag{22}$$

As well, the offered load can be expressed as:

$$\rho = IM \tag{23}$$

### C. MULTIPLE FLOWS

In practice, packet flows are not originated from a singular source but rather come from multiple sources. Moreover, modern DoS attacks may be launched from multiple networks and nodes (e.g., botnets). Therefore, we need to adapt the presented analytical model with realistic conditions. Suppose there are multiple flows arriving in the firewall and trigger some of the rules. Without loss of generality, we can assume that each flow triggers only one rule. In the situation where a flow triggers multiple rules, we can decompose it to multiple flows that each of them triggers one rule. Incoming flows are indicated by $I_\phi$ $\{I_\phi : 1 \leq \phi \leq Q\}$ such that any of the flows triggers a specific rule $\{R_\psi : 1 \leq \psi \leq V\}$ in the rule-base, where $Q$ denotes the total number of arriving flows and $V$ denotes the total number of rules in the firewall rule-base. This process is depicted in Figure 6 . To find a solution, we follow the presented approach below. Let us define the aggregated flow $\overline{I}$ as follows:

$$\overline{I} = \sum_{\phi=1}^{Q} I_\phi \tag{24}$$

The average number of time-slots taken to service the packets belonging to each individual flow (position of triggered rules) can be calculated as:

$$\overline{M} = 1 + \lceil \sum_{\phi=1}^{Q} \frac{I_\phi}{\overline{I}} \times M_\phi \rceil \tag{25}$$

In the equation above, the two added expressions are consist of the kernel processing time $(1/\mu)$ considered equal to one time-slot and the time taken for rule inspection. Consequently,

$$\gamma_\phi = \frac{I_\phi}{I} \overline{\gamma} \tag{26}$$

where,

$$\overline{\gamma} = \sum_{\phi=1}^{Q} \frac{I_\phi}{L}(1 - p_{loss}) \tag{27}$$

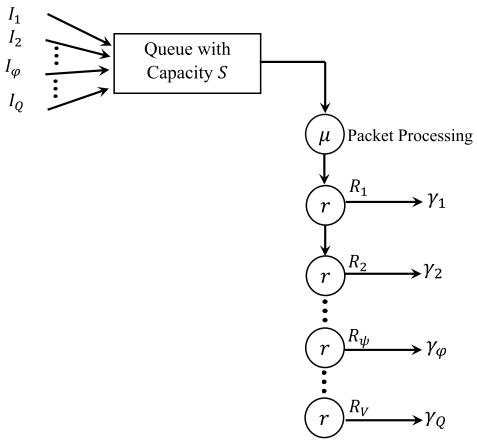



**FIGURE 6.** Generalized model for multiple flows when each flow triggers a different rule.

Now the CPU utilization of each flow can be calculated as:

$$CPU_{util,\phi} = I_\phi M_\phi (1 - p_{loss}) \tag{28}$$

Note that the average packet delay for each of the flows is equal to the total average packet delay for overall flows. Hence,

$$E[d] = \frac{E[s]}{\overline{I}(1 - p_{loss})} \quad (Slots) \tag{29}$$

or,

$$W_\phi = \overline{W} = \frac{E[s]}{\overline{I}(1 - p_{loss})} \times L \quad (Seconds) \tag{30}$$

### D. INFINITE BUFFER CAPACITY

Given the special case of a queuing system with infinite buffer capacity, when $S$ approaches infinity (i.e. $S = \infty$) the following expression can be obtained for $Q_{0,n}(z)$:

$$\lim_{S \to \infty} = \frac{p_0 Ln(z)}{N(z)} \quad n = 0, 1 \tag{31}$$

The expression contains only one unknown parameter of $p_0$. This case is a good approximation for a queuing system with a large buffer size (e.g. $S \geq 100000$).

### E. LIMITATIONS

In this paper, it is assumed that arriving packets have a fixed size. The impact of this assumption is that bigger packets may cause the queue buffer to overflow faster than smaller packet sizes and yield less throughput. In a practical setting, network packets do not have a fixed size. Despite this limitation, results obtained from our analytical model closely match to experimental results reported in [2] for $B = 1$ (Poisson arrivals).

## V. NUMERICAL RESULTS AND COMPARISON

To validate our model and accuracy checking, we simulated it using MATLAB [17] and compare the results with those reported in [2]. We set all of the numerical values and inputs

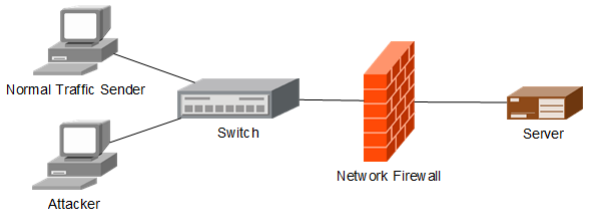



**FIGURE 7.** The hypothetical testbed used for model validation.

identical to the values already used in the aforementioned paper. To accomplish this, we also consider a hypothetical testbed (similar to what used in [2]) as depicted in the Figure 7. In this paper, we assume a firewall rule-base contains 10000 rules.

According to the values of parameters considered in [2], average processing time per rule (i.e. $1/r$) is 0.05 $\mu s$. The estimated time for kernel processing in addition to the IP processing time (i.e. $1/\mu$) is considered to be 2.65 $\mu s$. As already discussed, the time taken to service a packet is equal to $M$ and is always greater than 1. Now suppose the length of each time-slot equals $L$ seconds. The most accurate value to be assigned to $L$ is the shortest time interval assigned to the parameters of the experimental work. In practice, $1/r < 1/\mu$ [2]. Hence in this paper, we should take $L = 0.05\ \mu s$ but for convenience and to avoid excessive computational overhead, we employ $L = 1/\mu = 2.65\ \mu s$. Thus, approximately 53 rules will be interrogated during a specific time-slot. In this case, when the rule number 1000 is triggered, $M$ can be calculated as follows:

$$M = 1 + \lceil \frac{1000}{53} \rceil = 20$$

where $\lceil\ \rceil$ refers to the ceiling of the quantity inside it. The calculated $M$ is equivalent to the rule number 1007. In such a case, the average delay is calculated 0.35 $\mu s$ more than reality. However, it is a good approximation for this work.

### A. MEAN ARRIVAL RATE I

As already discussed, The mean arrival rate in steady state, $I$, is the steady state probability of having a packet arrival during an arbitrary time-slot. This parameter can be used to describe the packet arrival rate to the system. But in [2], packet arrivals are described using $\lambda$ which is the mean packet arrivals per second. One can easily convert $\lambda$ and $I$ together using the following equation:

$$\lambda = \frac{I}{L} \tag{32}$$

### B. BURSTINESS FACTOR

In uncorrelated arrivals, the burstiness factor equals 1. Therefore:

$$2 - \alpha - \beta = 1$$
$$\alpha + \beta = 1$$

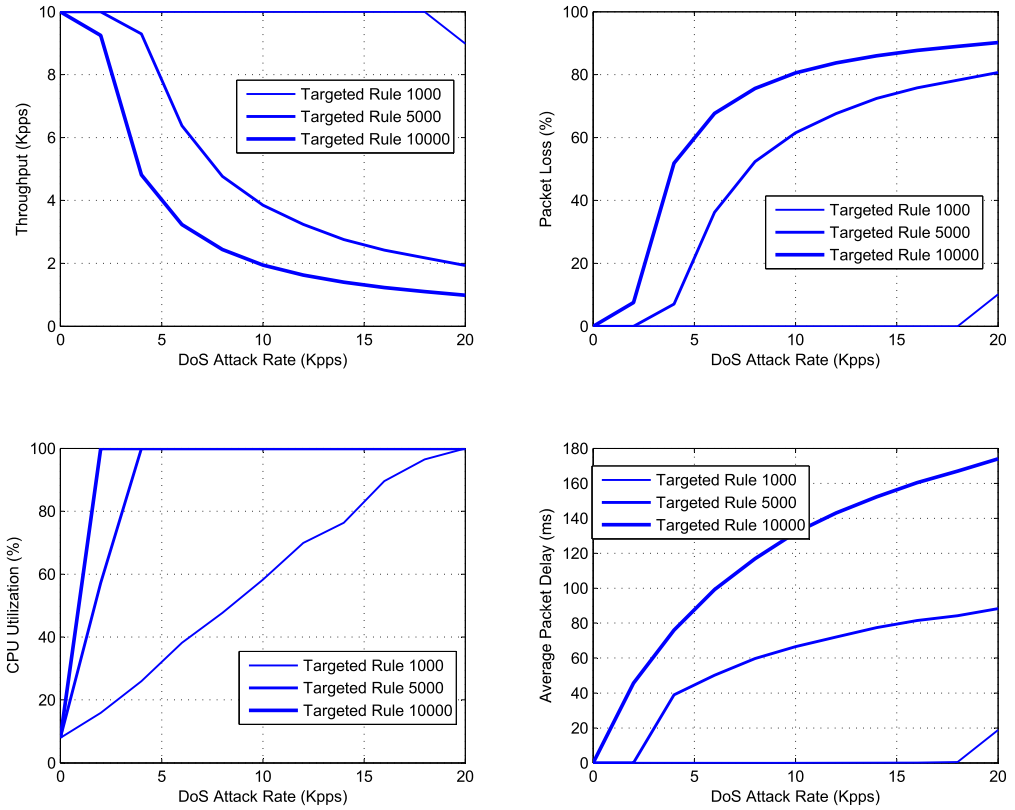
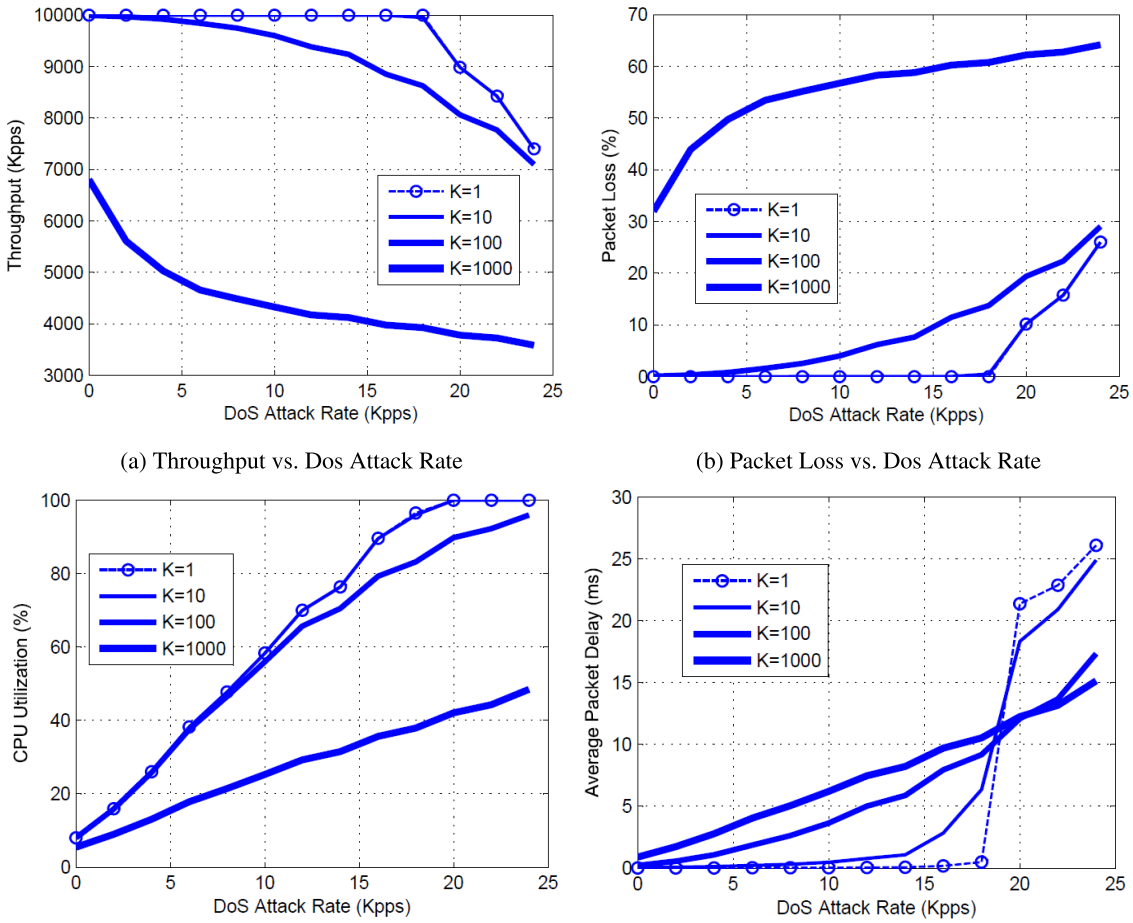



**FIGURE 8.** Results for different rules targeted by different attack rates and uncorrelated flows.




(a) Throughput vs. Dos Attack Rate

(b) Packet Loss vs. Dos Attack Rate

(c) CPU Utilization vs. Dos Attack Rate

(d) Average Packet Delay vs. Dos Attack Rate

**FIGURE 9.** Results for rule no. 1000 targeted by different attack rates of correlated flows and four values of burstiness factor.

### C. MAXIMUM QUEUE CAPACITY

Since the maximum capacity of both Tx and Rx DMA rings in [2] are set to 512, we set $S = 512$. This implies that both of the works have the same buffer capacity.

### D. NUMBER OF SERVICE TIME SLOTS M

The time needed to service an individual packet is equal to $ML$ seconds. We set up three experiments. We set the normal traffic to trigger the first rule in the firewall rule-base. In each experiment, the DoS flow targets the rules positioned around the ranks 1000, 5000, and 10000, respectively. As the minimum value for $M$ is 2, we set $M = 2$ for the normal traffic (one for kernel processing plus IP processing time and one for the first approximately 53 rules at the beginning of the rule-base). We set $M$ equal to 20, 96 and 190 as the number of time-slots required to grant service to the packet which triggers the rules number 1000, 5000 and 10000 respectively. The attacks which target the rule 1000 can be considered equivalent to a traditional DoS attacks but attacks which target rule 5000 and 10000 can be considered as complex DoS attacks that target last-matching rules. All of these experiments are done with a rate of $\lambda = 10000$ packet/seconds for normal traffic. To validate our analytical model, we compare our results for $B = 1$ to the empirical results reported in [2]. Our results are depicted in Figure 8. As it can be seen, for throughput, packet loss, CPU utilization, and average packet delay (all v.s. DoS attack rate), our results closely match the empirical results already reported in the mentioned reference. But for the average packet delay, we got better results than the mentioned reference. In other words, our results still follow the empirical results closely while their model has a considerable deviation from the empirical results.

If the arriving packets are not independent, the burstiness factor $B$ will not be equal to 1 and takes on various values. To evaluate the impact of different values for the burstiness factor $B$ on firewall performance and behavior, particularly under DoS attacks, we set a test similar to the experiment already discussed above. We changed the burstiness factor $B$ while we kept the values of the other parameters constant and observed the impact of it on the firewall performance. The results for the four different amounts of burstiness factor $B = 1, 10, 100, 1000$ are depicted in Figure 9. All

of these experiments are done with a rate of $\lambda = 10000$ *packets/second* for normal traffic which triggers the rule number 1. As well, DoS traffic targets the rule number 1000.

Figure 9a shows the impact of increasing in burstiness factor $B$ on system throughput. For $B = 1$ and $B = 10$, both of the curves are almost identical. For both of the mentioned curves, a downswing starts at the DoS rate of 18 Kpps. For $B = 100$, a downswing commences considerably sooner with a lower rate and downward concavity that represents a positive increase in the rate of packet wastage. For $B = 1000$, a faster downswing commences sooner than for $B = 100$ with significantly less amount of throughput (a.e. 7000 Kpps) and upward concavity that indicates decrements in packet wastage rate. Figure 9b exhibits the packet loss percentage and follows a similar trend. Figure 9c shows the CPU utilization. By increasing $B$, the CPU utilization keeps decreasing since the throughput decreases and the service time remains constant. Figure 9d shows the average packet delay. For $B = 1$, the average packet delay is comparatively less than others. But around the convergence point of the four curves, it experiences a sudden increase. For $B = 10$, it happens in a slower manner. For $B = 100$ and $B = 1000$, the curves increase almost monotonically. The convergence point is an interesting point since the order of curves is inverted and is the point that the throughput curve for uncorrelated arrivals ($B = 1$) breaks down.

## VI. CONCLUSION

In this paper, we propose a novel analytical approach for performance modeling and analysis of rule-based firewalls based on a discrete-time queuing system. We obtain closed-form expressions for performance metrics consist of throughput, packet loss, CPU utilization and delay while considering correlated packet arrivals. This model can be used for analyzing the behaviour of firewalls in normal conditions as well as when the firewall is under DoS attacks launched from different sources or any individual attackers that aim to waste the firewall resources and consequently cause a considerable disturbance using low-rate traffic flows. We also present a method in which the bursty nature of network traffic is exploited to create a bottleneck at the firewall. Since the firewall is located at the edge of the network, this disturbance can affect the experienced performance of the internal users. In this paper, we show that when these kinds of attacks are coupled with an increase in burstiness factor, they may cause serious hazards. To defend against these attack, we recommend weight allocation for rules, where the weight of each rule is based on the number of previous matches. The rule-base should then be reordered periodically based on their weight.

## APPENDIX

In this section, we present an approach for calculating $Q_{0,n}(z)$ and $R_n(y, z)$. A complete version of this approach is already presented in [13], but for the sake of completeness and adaption with our model, we present it in this section. Both of the functions $Q_{0,n}(z)$ and $R_n(y, z)$ that already defined in Equations (9) and (10) can be calculated as follows: According to Equations (9) and (10) and supposing that the *On* periods and *Off* periods have geometrical distribution, the following relations between $Q_{i,n}(z)$ and $Q_{i-1,n}(z)$ can be obtained:

$$\begin{pmatrix} Q_{i,0}(z) \\ Q_{i,1}(z) \end{pmatrix} = \begin{pmatrix} \beta & 1-\alpha \\ (1-\beta)z & \alpha z \end{pmatrix} \begin{pmatrix} Q_{i-1,0}(z) \\ Q_{i-1,1}(z) \end{pmatrix} + p_{i-1} z^{S+1}(1-z) \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad 2 \le i \le M-1 \tag{A.1}$$

$$\begin{pmatrix} Q_{1,0}(z) \\ Q_{1,1}(z) \end{pmatrix} = \begin{pmatrix} \beta & 1-\alpha \\ (1-\beta)z & \alpha z \end{pmatrix} \begin{pmatrix} Q_{0,0}(z) \\ Q_{0,1}(z) \end{pmatrix} - p_0 \begin{pmatrix} \beta \\ (1-\beta)z \end{pmatrix} \tag{A.2}$$

where $p_0$ is the steady-state probability of having an empty buffer at the start of any arbitrary time-slot. The parameters $p_i$ are defined as follows:

$$p_i \triangleq \lim_{k\to\infty} Prob[r_k = i, a_k = 1, s_k = S+1] \quad 1 \le i \le M-1 \tag{A.3}$$

Similarly, from the set of Equations (6) and (7), the following expressions for $Q_{0,n}(z)$ and $Q_{M-1,n}(z)$ can be obtained:

$$z\begin{pmatrix} Q_{0,0}(z) \\ Q_{0,1}(z) \end{pmatrix} = \begin{pmatrix} \beta & 1-\alpha \\ (1-\beta)z & \alpha z \end{pmatrix} \begin{pmatrix} Q_{M-1,0}(z) \\ Q_{M-1,1}(z) \end{pmatrix} + p_0 z \begin{pmatrix} \beta \\ (1-\beta)z \end{pmatrix} + P_{M-1} z^{S+1}(1-z) \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{A.4}$$

where

$$p_{M-1} \triangleq \lim_{k\to\infty} Prob[r_k = M-1, a_k = 1, s_k = S+1] \tag{A.5}$$

The Equations (A.1), (A.2) and (A.4) can be used to derive a set of linear equations with the unknown functions of $Q_{0,0}(z)$ and $Q_{0,1}(z)$. After some repeated substitutions, we reach:

$$(z\hat{I} - F^M)\begin{pmatrix} Q_{0,0}(z) \\ Q_{0,1}(z) \end{pmatrix} = (z\hat{I} - F^{M-1})\begin{pmatrix} \beta \\ (1-\beta)z \end{pmatrix} p_0 + \sum_{i=1}^{M-1} p_i z^{S+1}(1-z) F^{M-i-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{A.6}$$

where

$$\hat{I} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

and the matrix $F^M$ is defined as follows:

$$F^m \triangleq \begin{pmatrix} \beta & 1-\alpha \\ (1-\beta)z & \alpha z \end{pmatrix}^m \quad 0 \le m \le M \tag{A.7}$$

If we consider $\xi_1(z)$ and $\xi_2(z)$ as the two eigenvalues of the matrix $F$, these eigenvalues are the solutions of the characteristic equation $\xi^2 - (\alpha z + \beta)\xi + (\alpha + \beta - 1)z = 0$, In other words:

$$\xi_1(z) = u(z) + v(z) \quad and \quad \xi_2(z) = u(z) - v(z) \tag{A.8}$$

where

$$\begin{aligned} u(z) &= \frac{1}{2}(\alpha z + \beta) \quad and \ [v(z)]^2 \\ &= \frac{1}{4}[(\alpha z + \beta)^2 - 4(\alpha + \beta - 1)]z \end{aligned} \tag{A.9}$$

We obtain:

$$\xi_1(z) + \xi_2(z) = \alpha z + \beta \quad and \quad \xi_1(z)\xi_2(z) = (\alpha + \beta - 1)z \tag{A.10}$$

The matrix $F^m$, $0 \le m \le M$ can be expressed as:

$$F^m = \frac{1}{\xi_1 - \xi_2} \begin{bmatrix} \xi_1^{m+1} - \xi_2^{m+1} - \alpha z(\xi_1^m - \xi_2^m) & (1-\alpha)(\xi_1^m - \xi_2^m) \\ (1-\beta)z(\xi_1^m - \xi_2^m) & \xi_1^{m+1} - \xi_2^{m+1} - \beta(\xi_1^m - \xi_2^m) \end{bmatrix} \tag{A.11}$$

Therefore, we obtain the functions $Q_{0,0}(z)$ and $Q_{0,1}(z)$ from the Equations (A.6)

$$\begin{aligned} Q_{0,n}(z) &= \frac{1}{Z(z)}\{p_0 L_n(z) + \sum_{i=1}^{M-1} P_i z^{S+1}(1-z)T_{i,n}(z)\} \\ n &= 0, 1 \end{aligned} \tag{A.12}$$

where the functions $N(z)$, $L_n(z)$ and $T_{i,n}(z)$ are expressed in terms of $\xi_1(z)$ and $\xi_2(z)$ as follows:

$$N(z) = z[z - (\xi_1^M + \xi_2^M) + (\alpha + \beta - 1)^M z^{M-1}] \tag{A.13}$$

$$\begin{aligned} L_0(z) = \frac{1}{\xi_1 - \xi_2}\{&[\beta z^2 + (\xi_1\xi_2)^M](\xi_1 - \xi_2) \\ &+ (1+\beta)z\xi_1\xi_2(\xi_1^(M-1) - \xi_2^(M-1)) \\ &- (\beta + \xi_1\xi_2)z(\xi_1^M - \xi_2^M)\} \end{aligned} \tag{A.14}$$

$$\begin{aligned} L_1(z) = \frac{(1-\beta)Z^2}{(\xi_1 - \xi_2)}\{&\xi_1\xi_2(\xi_1^{M-1} - \xi_2^{M-1}) \\ &- (\xi_1^M - \xi_2^M) + z(\xi_1 - \xi_2)\} \end{aligned} \tag{A.15}$$

$$\begin{aligned} T_{i,0}(z) = \frac{(1-\alpha)}{(\xi_1 - \xi_2)}\{&z(\xi_1^{M-i-1} - \xi_2^{M-i-1}) \\ &+ (\xi_1\xi_2)^{M-i-1}(\xi_1^{i+1} - \xi_2^{i+1})\} \end{aligned} \tag{A.16}$$

$$\begin{aligned} T_{i,1}(z) = \frac{1}{(\xi_1 - \xi_2)}\{&-\beta z(\xi_1^{M-i-1} - \xi_2^{M-i-1}) \\ &+ (\xi_1\xi_2)^{M-i}(\xi_1^i - \xi_2^i) + z(\xi_1^{M-i} - \xi_2^{M-i}) \\ &- \beta(\xi_1\xi_2)^{M-i-1}(\xi_1^{i+1} - \xi_2^{i+1})\} \end{aligned} \tag{A.17}$$

In the Equation (A.12), the $M$ unknown probability $p_0$ and $p_i$ must be calculated. From the Equations (A.8) and (A.9),it is clear that $\xi_1^M + \xi_2^M$ is a polynomial in $z$ of degree $M$. Similarly, $(\xi_1^m - \xi_2^m)/(\xi_1 - \xi_2)$ for $m \ge 1$ is a polynomial in $z$ of degree $m - 1$. Consequently, the denominator $N(z)$ of $Q_{0,n}(z)$ is a polynomial in $z$ of degree $M + 1$. Therefore it has exactly $M+1$ zeroes inside the complex z-plane. Also, the numerator of (A.12) is a polynomial in $z$. It can be demonstrated that $N(0) = N(1) = 0$ and $L_n(0) = L_n(1) = 0$ which causes both the numerator and denominator of $Q_{0,n}(z)$ to disappear for $z = 1$. Finally, using the definition in Equation (10) and Equations (A.1) and (A.2), we can calculate linear expressions for $R_0(y, z)$ and $R_1(y, z)$ as follows:

$$\begin{aligned} &\begin{pmatrix} R_0(y,z) \\ R_1(y,z) \end{pmatrix} \\ &= \frac{\begin{pmatrix} 1-\alpha yz & (1-\alpha)z \\ (1-\beta)yz & 1-\beta y \end{pmatrix}}{[(1-\beta y)(1-\alpha yz) - (1-\beta)(1-\alpha)y^2 z]} \\ &\times\{(1 - y^M z)\begin{pmatrix} Q_{0,0}(z) \\ (Q_{0,1}(z) \end{pmatrix} - (y - y^M z)p_0 \begin{pmatrix} \beta \\ (1-\beta)z \end{pmatrix} \\ &+ \sum_{i=1}^{M-1} p_i y^{i+1} z^{S+1}(1-z)\begin{pmatrix} 0 \\ 1) \end{pmatrix}\} \end{aligned} \tag{A.18}$$

## REFERENCES

[1] J. Frahim, O. Santos, and A. Ossipov, *Cisco ASA: All-in-One Firewall, IPS, and VPN Adaptive Security Appliance*. San Jose, CA, USA: Cisco Press, 2014.

[2] K. Salah, K. Elbadawi, and R. Boutaba, "Performance modeling and analysis of network firewalls," *IEEE Trans. Netw. Service Manage.*, vol. 9, no. 1, pp. 12–21, Mar. 2012.

[3] V. Paxson and S. Floyd, "Wide area traffic: The failure of Poisson modeling," *IEEE/ACM Trans. Netw.*, vol. 3, no. 3, pp. 226–244, Jun. 1995.

[4] S. Xuan, D. Man, J. Zhang, W. Yang, and M. Yu, "Mathematical performance evaluation model for mobile network firewall based on queuing," *Wireless Commun. Mobile Comput.*, vol. 2018, Apr. 2018, Art. no. 8130152.

[5] R. Mohan, A. Yazidi, B. Feng, and J. Oommen, "On optimizing firewall performance in dynamic networks by invoking a novel swapping window-based paradigm," *Int. J. Commun. Syst.*, vol. 31, no. 15, p. e3773, 2018.

[6] A. K. Vasu, A. Ganesh, P. Ayyappan, and A. Sudarsan, "Improving firewall performance by eliminating redundancies in access control lists," *Int. J. Comput. Netw.*, vol. 6, no. 5, pp. 92–107, 2014.

[7] U. Mustafa, M. M. Masud, Z. Trabelsi, T. Wood, and Z. A. Harthi, "Firewall performance optimization using data mining techniques," in *Proc. 9th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Jul. 2013, pp. 934–940.

[8] P. J. Lee, H. Guo, and B. Veeravalli, "Enhancing CII firewall performance through hash based rule lookup," in *Proc. IEEE Region Conf.*, Nov. 2017, pp. 2285–2290.

[9] Z. Trabelsi, L. Zhang, and S. Zeidan, "Dynamic rule and rule-field optimisation for improving firewall performance and security," *IET Inf. Secur. J.*, vol. 8, no. 4, pp. 250–257, Jul. 2013.

[10] C. Wang, D. Zhang, H. Lu, J. Zhao, Z. Zhang, and Z. Zheng, "An experimental study on firewall performance: Dive into the bottleneck for firewall effectiveness," in *Proc. 10th Int. Conf. Inf. Assurance Secur.*, Nov. 2014, pp. 71–76.

[11] K. Salah, P. Calyam, and R. Boutaba, "Analytical model for elastic scaling of cloud-based firewalls," *IEEE Trans. Netw. Service Manag.*, vol. 14, no. 1, pp. 136–146, Mar. 2017.

[12] N. Ghrada, M. F. Zhani, and Y. Elkhatib, "Price and performance of cloud-hosted virtual network functions: Analysis and future challenges," 2018, *arXiv:1804.08787*. [Online]. Available: https://arxiv.org/abs/1804.08787

[13] S. Wittevrongel and H. Bruneel, "Discrete-time queues with correlated arrivals and constant service times," *Comput. Oper. Res.*, vol. 26, no. 2, pp. 93–108, 1999.

[14] W.-H. Zhou and A.-H. Wang, "Discrete-time queue with Bernoulli bursty source arrival and generally distributed service times," *Appl. Math. Model.*, vol. 32, no. 11, pp. 2233–2240, 2008.

[15] K. Salah, K. Sattar, M. Sqalli1, and E. Al-Shaer, "A potential low-rate DoS attack against network firewalls," *Secur. Commun. Netw.*, vol. 4, pp. 136–146, Jan. 2011.
[16] S. A. Crosby and D. S. Wallach, "Denial of service via algorithmic complexity attacks," in *Proc. USENIX Secur. Symp.*, 2003, pp. 29–44.
[17] *MATLAB, Version 9.4.0 (R2018a)*, The MathWorks Inc., Natick, MA, USA, 2018.

**YAHYA SHAHSAVARI** received the B.Sc. degree in electrical engineering from the University of Zanjan, and the M.Sc. degree in information and communication technology from the Iran University of Science and Technology, Iran, in 2008 and 2016, respectively. He is currently pursuing the Ph.D. degree with FUSÉE laboratory, a pioneer laboratory dedicated to blockchain systems at ÉTS, University of Quebec. His main research interests include network performance modeling (particularly blockchain networks) and traffic engineering.

**HADISHAHRIAR SHAHHOSEINI** received the B.Sc. degree in electrical engineering from the University of Tehran, in 1990, the M.Sc. degree in electrical engineering from the Azad University of Tehran, in 1994, and the Ph.D. degree in electrical engineering from the Iran University of Science and Technology, in 1999. He is currently an Associate Professor with the School of electrical engineering, IUST. He has published more than 150 papers from his research works in scientific journals and conference proceedings. His areas of research include networking, supercomputing, and reconfigurable computing.

**KAIWEN ZHANG** received the B.Sc. and M.Sc. degrees from McGill University, Montréal, and the Ph.D. degree from the University of Toronto. He was an Alexander von Humboldt Postdoctoral Fellow in computer science with TU Munich. He is currently a Professor with the Department of Software and IT Engineering, ÉTS. His research interests include blockchain technologies, publish/subscribe systems, massive multiplayer online games, performance modeling, and software-defined networking.

**HALIMA ELBIAZE** received the M.Sc. degree in telecommunication systems from the Université de Versailles, in 1998, and the Ph.D. degree in computer science from the Institut National des Télécommunications, Paris, France, in 2002. Since 2003, she has been with the Department of Computer Science, Université du Québec Montréal, QC, Canada, where she is currently an Associate Professor. Her research interests include performance evaluation, traffic engineering, cloud computing, wireless networks, and next-generation IP networks. She had been awarded many research grants from both public agencies and industry. In 2005, she received the Canada Foundation for Innovation Award to build her IP over the DWDM Network Laboratory.