Pay-As-You-Go (PAYG)/Hourly Auto Scaling

https://campus.barracuda.com/doc/73007159/

To deploy the Pay-As-You-Go/Hourly Barracuda Web Application Firewall for AWS in the auto scaling model, follow the instructions in this article.

The Pay-As-You-Go/Hourly Auto Scaling CloudFormation Template:

- Includes the number of Barracuda Web Application Firewall instances to be deployed and provisioned.
- Creates an IAM role that can be used to access the S3 storage and create the S3 bucket for the stack. Typically, an S3 bucket stores the instance data such as the serial number and primary IP address (i.e., WAN IP address) of the deployed Barracuda Web Application Firewall VM(s).
- Includes the security group created and attached to the deployed Barracuda Web Application Firewall instances.
- Includes alarms created for CPU and network usage to determine the scaling up/down of instances.

> Before proceeding with the deployment, ensure that the AWS services required for the auto scaling setup are created/configured. See to the section "AWS Services Required for the Auto Scaling Setup" in the article Auto Scaling of Barracuda Web Application Firewall using CloudFormation Template on Amazon Web Services.

The PAYG auto scaling CFT is available on GitHub.

Prerequisites

- Latest Barracuda Web Application Firewall CFT Template.
- Availability zone(s), VPC ID, and subnet ID where you want to deploy the Barracuda Web Application Firewall and protect your servers.
- Elastic Load Balancer to load balance the traffic between the deployed Barracuda Web Application Firewalls. For more information, see Elastic Load Balancing in the AWS documentation.
- Ability to create an IAM role with access to S3. The CFT will create an IAM role that has permissions to create and modify an S3 bucket. The S3 bucket stores the IP address and serial number details of the deployed Barracuda Web Application Firewall instances. The IAM role uses "AssumeRole" and "STS keys" for maximum security while accessing the S3 bucket.

Default Values of the Barracuda Web Application Firewall PAYG CloudFormation

Template

The following are the default values of the Barracuda Web Application Firewall PAYG CloudFormation Template (CFT). You can modify the values as needed.

- **ScalingMinSize** - The minimum number of Barracuda Web Application Firewall instances to be deployed initially to serve the web traffic. Default: 1
- **Scaling MaxSize** - The maximum number of instances to be scaled up to handle the traffic whenever required. Default: 4
- **Instance Type** - Instance type to be used in Amazon Web Services (AWS). Default: m3.medium
- **Health Check Grace Period** for Auto Scaling is set to 1200 seconds.
- **Pause Time** for Update Policy is set to 600 seconds.
- **Security Group** with the following ports opened:

Port	Protocol	Description
8000	TCP	Provides HTTP access to the Barracuda Web Application Firewall web interface.
8443	TCP	Provides HTTPS access to the Barracuda Web Application Firewall web interface.
8002	TCP	Required for clustering the instances and to auto scale the instances up/down.
32575	TCP	Required for clustering the instances and to auto scale the instances up/down.
32576	UDP	Required for clustering the instances and to auto scale the instances up/down.
Server port specified in the CFT when creating the stack	TCP	Required for the service(s) configured on the Barracuda Web Application Firewall.

- **Default Cool Down time** for scaling the instances up/down is set to 300 seconds.
- **Alarms** for CPU and bandwidth. Note: These alarms are designed to ensure that auto scaling does not lead to instability. The alarms will scale up quickly and scale down slowly to ensure traffic to the site is not disrupted.

Alarm Type	Threshold Value (Average)	Action	Evaluation Periods
Network-In High Alarm	70% of max throughput for 5 minutes	Bring up one instance	5 minutes
Network-In Low Alarm	< 50% of max throughput for 2 hours 30 minutes	Bring down one instance	2 hours 30 minutes
Network-Out High Alarm	70% of max throughput for 5 minutes	Bring up one instance	5 minutes

Network-Out Low Alarm	< 50% of max throughput for 2 hours 30 minutes	Bring down one instance	2 hours 30 minutes
CPU High Alarm	> 85% for 5 minutes	Bring up one instance	5 minutes
CPU Normal Alarm	< 60% for 2 hours 30 minutes	Bring up one instance	2 hours 30 minutes

Next Step

Continue with the article How the Barracuda CloudFormation Template Works in Pay-As-You-Go (PAYG)/Hourly Instance.