# TABLE OF CONTENTS

- Problem Statement ................................................................. 3
- Problem Resolution—FortiGate Next Gen Firewall Auto Scaling in AWS ... 3
- Design Criteria ........................................................................... 4
- Functionality ............................................................................... 4
- Solution Components ................................................................... 5
- How It All Works Together in the FortiGate Auto Scaling Solution ...... 5
- Fortigate Autoscaling Set-Up Procedure ....................................... 6
  - Step 1: Launch The Cloudformation Template ............................... 6
  - Step 2: Connect To The Ec2 Worker Node ................................... 8
  - Step 3: Configure The Primary Fortigate Firewall ....................... 8
  - Step 4: Review The Log File & Additional Details ....................... 8
- New VPC BYOL and OnDemand Licenses ...................................... 9
- Deploy Auto Scale Stacks Into an Existing VPC ............................ 12
- Deploy Base Setup VPC Template (if needed) ............................... 12
- Scenario 3: Existing VPC OnDemand Licenses .............................. 15
- Scenario 4: Existing VPC BYOL and OnDemand Licenses ............. 18
- Configuring FortiGate Instances .................................................. 22
- Appendix A. Troubleshooting ....................................................... 27
- Accessing the Worker Node ......................................................... 27
- Potential Issues ............................................................................. 29
- BYOL License Issues ................................................................... 29
- Script Waiting for Configuration .................................................. 30
- Traffic Issues/Impaired Route Tables/Failover Issue/Impaired Instances 31
- Failover Algorithm ....................................................................... 31
- Auto Scale Events/Scale-out ......................................................... 31
- Auto Scale/Scale-in ....................................................................... 32
- Appendix B. Automated Template Deployment and Code Updates .... 32
- Worker Node Status ..................................................................... 32
- Worker Code Updates ................................................................... 33
- Automated Template Deployment .................................................. 33
- Appendix C .................................................................................... 35
- Appendix D .................................................................................... 35
DEPLOYMENT GUIDE: FORTIGATE AUTO SCALING

PROBLEM STATEMENT
How do you protect dynamically scaling AWS compute resources with advanced, consistent security postures without purchasing multiple firewalls that sit idle in non-peak sessions? How can you leverage AWS elasticity for true pay-as-you-go cloud networking protected with advanced network security? This document provides the information you need to solve the challenge of scaling security and capacity management in conjunction with workloads in AWS.

PROBLEM RESOLUTION - FORTINET'S FORTIGATE UTM AUTO SCALING IN AWS
Security has become an essential enabler of application and service delivery in cloud environments. For organizations contemplating the migration of essential activities to the cloud, the ability to match security to workloads is a key business consideration.

The effective application of AWS cloud security requires an ability to scale up and down in concert with the workload. The most operationally advantageous way to support this dynamic need is with automation.

As organizations migrate their production infrastructure to the cloud, many leverage AWS’s AutoScaling capabilities to automatically scale their cloud compute resources according to conditions they define. This provides an excellent means of optimizing cloud costs, detecting faulty instances, identifying unhealthy applications and automating replacement. As cloud workloads are scaled-out, the concerns of secured data protection persist and require a security-scaling automation capability aligned to flexing compute resources.

To address this situation, Fortinet has developed an AutoScaling CloudFormation template which adds enterprise firewall instances automatically based on user-defined criteria while using AWS integrated scripts and templates to maintain a familiar UI and initiate security elasticity for optimal network utilization. Fortinet maps your security postures to scale up and down with your EC2 in an AWS CloudFormation template
- You define the min/max quantity and type of FortiGate appliances allowed to scale
- Flex criteria includes: CPU Utilization, Memory Utilization, Concurrent Sessions, or Session Setup Rate
- Four Templates available for new/existing VPC with On Demand / BYOL

SOLUTION HIGHLIGHTS
- Provides timely protection as workloads scale horizontally
- Delivers automatic scaling for best-in-class advanced security in AWS
- Pre-tunes “minimum” and “maximum” security parameters to provide refined security policies
- Minimizes cloud instance over-subscription and OPEX spending
- Eliminates error-prone manual intervention in security configurations
**DEPLOYMENT GUIDE: FORTIGATE AUTO SCALING**

**DESIGN CRITERIA**

Using native AWS service functions and FortiGate automation features, you can scale your FortiGate firewall on AWS as your workload demands fluctuate dynamically. The design criteria attributes include:

- 3 Elastic IPs are needed for the initial launch
- FortiGate AutoScaling works with instances with two interfaces. One instance interface is used for Public / One instance interface is used for routing internal subnets
- The initial instance can be different from the AutoScaling instances. For example, start with a c3.2xlarge (Hourly or Annual) and set the AutoScaling Group to incorporate c3.large (all Scaling instances being On Demand Hourly).
- The CloudFormation template and Worker Node are integrated closely; therefore, it is important to maintain the set parameters/resources naming conventions
- Inbound traffic is distributed equally by the Internet inbound facing Elastic Load Balancer / Outbound traffic goes through 2-instances (the number of instances in the AutoScaling group does not change this)
- Source NAT must be used at the FortiOS level for inbound traffic

**FUNCTIONALITY**

To ensure availability and optimization of Fortinet’s advanced threat protection over the entire Auto Scaling Group, Fortinet maps your AWS security postures to scale up and down with your EC2 in an AWS CloudFormation template. This template can be held in a repository, making it reproducible and easily deployable as new instances require secure elasticity.
SOLUTION COMPONENTS

The FortiGate Auto Scaling solution utilizes AWS native tools, templates, and infrastructure including:

- **CloudFormation**: Enables you to use a template file to create and provision a collection of resources together as a single unit (a stack) predictably and repeatedly.

- **CloudWatch**: Monitors your AWS resources and the applications you run on AWS in real time.

- **IAM**: Identity and Access Management (IAM) is an AWS service that enables managing users and user permissions in AWS.

- **ELB**: Elastic Load Balancer automatically distributes traffic across multiple AWS EC2 instances.

- **S3**: Simple Storage Service, object-based storage service in AWS.

- **SQS**: A messaging queue service that handles message or workflows between components in a system.

- **EC2**: Elastic Compute Cloud (EC2) is a service that provides resizable computing capacity—literally, servers in Amazon’s data centers—that you use to build and host your software systems.

- **AZs**: The AWS infrastructure is built around Regions and Availability Zones (“AZs”). A Region is a physical location with multiple Availability Zones. Availability Zones consist of one or more discrete data centers, each with redundant power, networking, and connectivity, housed in separate facilities. These Availability Zones offer you the ability to operate production applications and databases that are more highly available, fault tolerant, and scalable than would be possible from a single data center.

- **VPC**: Virtual Private Cloud (VPC) enables you to launch AWS resources into a virtual network that you have defined. This virtual network closely resembles a traditional network that you would operate in your own data center, with the benefits of using the scalable infrastructure of AWS.

HOW IT ALL WORKS TOGETHER IN THE FORTIGATE AUTO SCALING SOLUTION


- The FortiGate AutoScaling CloudFormation template performs the following actions in the set-up:
  - Two FortiGates are launched in two Availability Zones (AZs) in a High Availability (HA) architecture
  - The two FortiGates are added to an Internet facing ELB
    - An Internal ELB is created for the application traffic
  - An EC2 worker node instance is launched (t2micro) into its own AutoScaling Group
  - The Launch Configuration is created for the AutoScaling Group
  - A LifeCycle Hook is created to post to the SQS Queue when a scaling event occurs
  - AutoScaling Group Scaling policies are created
  - CloudWatch Alarms are created for ScaleUp and ScaleDown triggering
  - An SQS queue is created with AutoScaling instance launch permission events
  - Dynamic IAM roles are created for both EC2 launch and also to write to the SQS queue for AutoScaling LifeCycle Hook

- Your FortiGate AutoScaling policy set up will take approximately ~20 minutes.
  - The AWS CloudWatch set threshold for ScaleOut will trigger the ScaleOut policy.
  - The AWS CloudWatch set threshold for ScaleIn will trigger the ScaleIn policy.

Summary: AutoScaling components are created via the CloudFormation Template with specific automation steps completed via autoscale service in the t2micro ‘Worker Node’ (see diagram)
FORTIGATE AUTOSCALING SET-UP PROCEDURE

SCENARIO 1: NEW VPC ON DEMAND LICENSES

STEP 1: LAUNCH THE CLOUDFORMATION TEMPLATE.
- From the AWS Management Console, navigate to the CloudFormation service, create a new stack and launch the CloudFormation template.
- Select Template
### Specify Details

![CloudFormation stack creation interface](image)

#### Parameters

**Auto Scaling Configuration**
- **AutoScalingGroup**
- **InstanceType**
- **InstanceSize**
- **ListenerPort**
- **VPCConfiguration**
- **FortigateInstanceConfiguration**
- **ExternalELBConfiguration**
- **ClientCert**
- **VPCConfiguration**
- **AutoScalingGroupInformation**
- **NetworkConfiguration**
- **KeypairUsedForScaling**
- **SQSQueueUsedInAutoScalingProcess**

**Key Input Parameters**

- **AZforFirewall1** and **AZforFirewall2** must be in different Availability Zones
- **Scaling Parameter**
- **ScaleUp Threshold**
- **ScaleDown Threshold**
- **FortiGate Instance Type and Size**
  - **Initial FortiGate**
  - **Scale Up and Scale Down FortiGate**
- **Listener port configuration for the ELBs**
- **AutoScaling Group information**
- **Network information including the VPC CIDR, public and private subnets**
- **Keypair used for scaling purposes**
- **SQS Queue used in AutoScaling process**
The CloudFormation template creates the stack and outputs seen in the screen shot:

![CloudFormation template screenshot]

**STEP 2: CONNECT TO THE EC2 WORKER NODE**


**STEP 3: CONFIGURE THE PRIMARY FORTIGATE FIREWALL**

- Log-in to the primary FortiGate Firewall listed in the Output Section of the CloudFormation Template
- Once the Primary FortiGate is configured, add a Tag to the Primary FortiGate instance with the following “FortiGate-State” as the Key with the Value “Configured” as seen in the screen shot below

![Tagging screenshot]

**STEP 4: REVIEW THE LOG FILE & ADDITIONAL DETAILS**

- To review the Log File from the Worker Node, do so with the following command:
  - `tail -f -n 100 ~/aws_autoscale/subsys/autoscale.log`
- To restart the autoscale service, do so with the following two commands:
  - `sudo service autoscale stop`
  - `sudo service autoscale start`
- To check the status of the autoscale service, do so with the following command:
  - `sudo service autoscale status`
SCENARIO 2: NEW VPC BYOL AND ON DEMAND LICENSES

NEW VPC BYOL AND ONDEMAND LICENSES

(NewVPC_Fortigate542_Autoscale_ELB_WorkerNode_BYOL.template)

1. Pull the template from the proper repository to local storage.
2. From the AWS Management Console, navigate to the CloudFormation Service and “Create New Stack.”
3. “Upload a template to Amazon S3” and browse to its location on local storage.
4. Click Next.

5. Auto Scaling Configuration:
   a. Stack Name (Region Unique)
   b. ASQueue (SQS Queue Name that is unique within the scope of your queues)
   c. AZ For Firewall1 (Availability Zone within region to place FortiGate 1)
   d. AZ For Firewall2 (Availability Zone within region to place FortiGate 2)

Note: Availability Zone may not support the instance size of the FortiGate Instance. If you get a warning that a specific instance size is not supported, choose a different size or choose a different zone.

For example, the m3.medium instance type is not supported in the us-east-1b AZ.
6. VPC Configuration Section:
   a. Choose a CIDR block (if different from the defaults provided) that will hold the subnets specified for Public1, Private1, Public2, Private2 subnets.
   b. Provide unique subnet range for each of the Public and Private Subnets.

7. FortiGate Instance Configuration:
   a. Choose Instance Type for initial FortiGates and Auto Scaled FortiGates.
   b. Provide CIDR Range to define Security Group for FortiGate Access.
   c. Provide a key pair to allow ssh to FortiGate instances.
   d. Provide the name of the S3 bucket used to store licenses. If the S3 bucket does not currently exist, provide a valid S3 bucket name and create the bucket and upload the licenses after this template is deployed. The auto scale script will “sleep” waiting for the creation of the S3 bucket and the licenses to be uploaded.

Note: This is the only difference between this template and the previous template.

8. Defaults for ELB Configuration (Please see AWS documentation if you need to change defaults.)
9. Worker Node Instance Configuration:
   b. Provide a key pair to allow ssh to FortiGate instances.
   c. Click Next to deploy template.

10. If you wish to add additional Tags, Permissions, or Add Advanced Notification Options, you can do so on this screen. Please see AWS documentation for more information. Click Next to continue.

11. Review your parameters and acknowledge the IAM resources notification and click “Create.”
12. Verify “Stack Creation In Progress” message.

![Stack Creation In Progress](image)

13. Monitor Stack Creation Events if desired.

![Events](image)

**DEPLOY AUTO SCALE STACKS INTO AN EXISTING VPC**

The next two templates will deploy FortiGate Auto Scale Groups into an existing customer VPC. The first template will deploy into an existing VPC using OnDemand licenses only. The second template will deploy into an existing VPC and use BYOL licenses for perpetual FortiGates and OnDemand licenses for auto scaled instances. Once again, before deploying a template that uses BYOL licenses, you will need to create an S3 bucket to store the licenses and upload the BYOL licenses to the S3 bucket. Please see the steps above to create the S3 bucket and upload the licenses.

**DEPLOY BASE SETUP VPC TEMPLATE (IF NEEDED)**

The sections after this section will deploy auto scale groups into an existing VPC. If you want to demonstrate deploying into an existing VPC, but you do not currently have a deployed VPC, use the NewVPC_BaseTemplate.template to create a minimal base VPC that can be used before proceeding to the next step.

1. Pull the template from the proper repository to local storage.
2. From the AWS Management Console, navigate to the CloudFormation Service and “Create New Stack.”
3. “Upload a template to Amazon S3” and browse to its location on local storage.
4. Click Next.
5. Stack Name (Region Unique).
6. Choose CIDR Block to contain the Public1, Private1, Public2, Private2 subnets.
7. Choose AZ for Public1 and Private1 subnets.
8. Choose AZ for Public2 and Private2 subnets.
9. Accept defaults for subnets or provide values that fit within the chosen CIDR block.
10. Click Next.

11. If you wish to add additional Tags, Permissions, or Add Advanced Notification Options, you can do so on this screen. Please see AWS documentation for more information. Click Next to continue.
12. Review your parameters and acknowledge the IAM resources notification and click “Create.”


14. Monitor Stack Creation Events if desired.
SCENARIO 3: EXISTING VPC ONDEMAND LICENSES

(ExistingVPC_Fortigate542_Autoscale_ELB_WorkerNode.template)

1. Pull the template from the proper repository to local storage.
2. From the AWS Management Console, navigate to the CloudFormation Service and “Create New Stack.”
3. “Upload a template to Amazon S3” and browse to its location on local storage.
4. Click Next.

5. Auto Scaling Configuration:
   a. Stack Name (Region Unique)
   b. ASQueue (SQS Queue Name that is unique within the scope of your queues)
   c. AZ For Firewall1 (Availability Zone within region to place FortiGate 1)
   d. AZ For Firewall 2 (Availability Zone within region to place FortiGate 2)

Note: Availability Zone may not support the instance size of the FortiGate Instance. If you get a warning that a specific instance size is not supported, choose a different size or choose a different zone.

For example, the m3.medium instance type is not supported in the us-east-1b AZ.
6. VPC Configuration Section:
   a. Use dropdowns to fill in the values for VPC.
   b. Use dropdowns to fill in the values for Public and Private1 Subnet.
   c. Use dropdowns to fill in the values for Public and Private2 Subnet.

   Note: These values are pulled from the existing VPCs.

7. FortiGate Instance Configuration:
   a. Choose Instance Type for initial FortiGates and Auto Scaled FortiGates.
   b. Provide CIDR Range to define Security Group for FortiGate Access.
   c. Provide a key pair to allow ssh to FortiGate instances.

8. Defaults for ELB Configuration (Please see AWS documentation if you need to change defaults.)
9. Worker Node Instance Configuration:
   
   d. Provide CIDR Range to define Security Group for Worker Node Access.
   
   e. Provide a key pair to allow ssh to FortiGate instances.
   
   f. Click Next to deploy template.

10. If you wish to add additional Tags, Permissions, or Add Advanced Notification Options, you can do so on this screen. Please see AWS documentation for more information. Click Next to continue.

11. Review your parameters and acknowledge the IAM resources notification and click “Create.”
12. Verify “Stack Creation In Progress” message.

13. Monitor Stack Creation Events if desired.

SCENARIO 4: EXISTING VPC BYOL AND ONDEMAND LICENSES

(ExistingVPC_Fortigate542_Autoscale_ELB_WorkerNode_BYOL.template)

1. Pull the template from the proper repository to local storage.

2. From the AWS Management Console, navigate to the CloudFormation Service and “Create New Stack.”

3. “Upload a template to Amazon S3” and browse to its location on local storage.

4. Click Next.

5. Auto Scaling Configuration:
   a. Stack Name (Region Unique)
   b. ASQueue (SQS Queue Name that is unique within the scope of your queues)
   c. AZ For Firewall1 (Availability Zone within region to place FortiGate 1)
   d. AZ For Firewall 2 (Availability Zone within region to place FortiGate 2)

Note: Availability Zone may not support the instance size of the FortiGate Instance. If you get a warning that a specific instance size is not supported, choose a different size or choose a different zone.

For example, the m3.medium instance type is not supported in the us-east-1b AZ.
6. VPC Configuration Section:
   a. Use dropdowns to fill in the values for VPC.
   b. Use dropdowns to fill in the values for Public and Private1 Subnet.
   c. Use dropdowns to fill in the values for Public and Private2 Subnet.

Note: These values are pulled from the existing VPCs.

7. FortiGate Instance Configuration:
   a. Choose Instance Type for initial FortiGates and Auto Scaled FortiGates.
   b. Provide CIDR Range to define Security Group for FortiGate Access.
   c. Provide a key pair to allow ssh to FortiGate instances.
   d. Provide the name of the S3 bucket used to store licenses. If the S3 bucket does not currently exist, provide a valid S3 bucket name and create the bucket and upload the licenses after this template is deployed. The auto scale script will "sleep" waiting for the creation of the S3 bucket and the licenses to be uploaded.

Note: This is the only difference between this template and the previous template.
8. Defaults for ELB Configuration (Please see AWS documentation if you need to change defaults.)

9. Worker Node Instance Configuration:
   b. Provide a key pair to allow ssh to FortiGate instances.
   c. Click Next to deploy template.

10. If you wish to add additional Tags, Permissions, or Add Advanced Notification Options, you can do so on this screen. Please see AWS documentation for more information. Click Next to continue.
11. Review your parameters and acknowledge the IAM resources notification and click “Create.”

12. Verify “Stack Creation In Progress” message.

13. Monitor Stack Creation Events if desired.
CONFIGURING FORTIGATE INSTANCES

After the desired template is deployed, there will be two FortiGate Instances running in two different Availability Zones. By default, no configuration exists on the instances and the primary FortiGate will need to be configured via the FortiOS GUI or a pre-populated configuration file will need to be restored via the GUI. Properly configuring a FortiGate policy is beyond the scope of this document, but this section will provide a walk-through of using a backup configuration to restore config to a FortiGate Instance.

Proper configuration of a FortiGate requires proper configuration of the following:

1. Network Interfaces
2. Routing Information
3. Administrator and User Accounts
4. Address and Service Objects
5. Firewall Policies
6. VPN Configuration
7. Etc.

Please see the FortiGate Cookbook and the FortiGate System Administration Guide for more information on properly configuring the FortiGate Firewall.

http://cookbook.fortinet.com/creating-security-policies/
http://docs.fortinet.com/fortigate/admin-guides

To upload/restore a configuration previously backed up from a FortiGate device:

1. From the AWS Management Console, navigate to the EC2 Service and
2. Click on Running Instances.
3. Click on the OnDemandA instance in the stack to view details.
4. Copy the assigned Public Elastic IP address of the OnDemandA instance.
5. Open a new browser page and browse to “https://<ip address of OnDemandA>.”

6. Log in to FortiGate with User Name “admin” password “instance-id.”

Note: Retrieve Instance-Id from EC2 screen.

7. If the login prompts you to install a FortiGate License before continuing, follow the instructions above for creating an S3 bucket and uploading the licenses.
8. Restore configuration using Dashboard->System Configuration->Restore.

9. Browse to local backup file and restore.

10. Update EC2 instance tags and set “Fortigate-State” to “Configured.”
   a. Go back to EC2 Dashboard.
   b. Highlight OnDemandA.
   c. Click “Tags” tab and “Add/Edit Tags.”
d. Change value of “Fortigate-State” to “Configured.”

11. The Auto Scale script will now begin to run and configure all the instances in the auto scale group. After each instance is configured, the script will add instances that are ready to accept traffic to the auto scale group and to the elastic load balancer. This can be monitored by checking the auto scale group and the members of the elastic load balancer.
   
a. Go to EC2 Dashboard -> Auto Scaling Groups.
   
b. Here you will see three auto scale groups that are part of the stack:
      i. ASG0—Instances in the perpetual (BYOL) auto scale group
      ii. ASG1—Instances in the dynamic (OnDemand) auto scale group
      iii. ASG2—Worker Node Instance
The perpetual auto scale group is limited to the number of defined perpetual firewalls. If a perpetual instance becomes “unhealthy,” CloudWatch will terminate the instance and spawn a new instance and the auto scale script will push configuration and reload with an available BYOL license. Instances in the auto scale group will not be part of the scale-out/scale-in policy.

Instances in the ASG1 auto scale group can dynamically grow and shrink based on scaling metrics that are monitored by CloudWatch. The worker node instance in ASG2 will be replaced by CloudWatch if it should become unhealthy, but will not auto scale based on the scale-out/scale-in policy.

12. Auto scale instances can be monitored by highlighting the auto scale group and clicking on the “instances” tab. This example shows the two BYOL instances that are members of ASG0.

13. Instances that are accepting traffic from elastic load balancers can be monitored from the load balancer screen:
   a. Go to EC2 Dashboard -> Load Balancers.
   b. Highlight the “EELB” (external elastic load balancer) to see which FortiGates are accepting traffic.
   c. Click “Instances” tab.
   d. This example shows the OnDemandA and OnDemandB instances are accepting traffic.
ACCESSING THE WORKER NODE

Troubleshooting issues with the stack is done via the worker node, AWS CLI commands, AWS Console, and the auto scale script logs. If you have an issue and want to monitor the status of the worker node, log in to the worker node via ssh using the following steps:

1. Pull the IP address of the worker node from the EC2 Dashboard on the AWS console. Navigate to the EC2 Dashboard and click on “Running Instances.”

2. Copy the IPv4 Public IP from the EC2 Dashboard for “ASInstance.”

3. You may also use the AWS CLI command to retrieve the IP address of the worker node:

   ```bash
   aws cloudformation describe-stack-resources --stack-name demostack --output text |grep ASEIP
   ```
4. Proceed to a terminal and ssh into the worker node using the IP address and "ec2-user" as the username and the key pair specified for the worker node when the template was deployed.

5. From here, the worker node script is kept in the aws_autoscale directory and the log files are kept in the aws_autoscale/subsys directory. You can monitor the logfile using a "tail -f" command or use the alias provided (worker_node_log) to monitor the status of the auto scale script.
POTENTIAL ISSUES

BYOL LICENSE ISSUES

1. S3 Bucket for licenses not created:
   b. If you deployed a BYOL template and the auto scale script is waiting for licenses, you will see the script sleeping waiting for licenses.
   c. Make sure the following is true:
      i. S3 Licenses Bucket is created using the name provided in the template deploy or via CLI.
ii. Make sure licenses have been uploaded to the bucket

or via CLI

ii. Once the licenses are in the bucket, the auto scale will apply them to the proper EC2 instance and the following will appear in the auto scale script log.

SCRIPT WAITING FOR CONFIGURATION

After the FortiGates and Worker Node are deployed into the VPC, the auto scale script will wait for the Primary Firewall to be configured. The auto scale script detects this when the FortiGate Tag for OnDemandA with Name = Fortigate-State and Value is changed to “Configured.” This state can be detected by monitoring the Auto Scale Script Log. If the instances are detected, but the scripts are not proceeding past “detection,” then you need to configure the FortiGate via the FortiOS GUI or upload a configuration to the firewall and then change the tag. Follow the section above that provides instruction for “Configuring FortiGate Instances.”

Once the tag value is changed, the script should proceed with configuring the instances in the auto scale group.
TRAFFIC ISSUES/IMPAIRED ROUTE TABLES/FAILOVER ISSUE/IMPAIRED INSTANCES

As the FortiGates are initializing in the VPC, the auto scale script will determine which route tables are associated with each subnet. The auto scale script will then modify the route table for each subnet to direct traffic through the internal interface of the FortiGate located in the AZ, rather than send traffic to the “intrinsic router.” By redirecting the traffic, this gives the FortiGate the opportunity to inspect the traffic and apply security policy to the flows. However, the auto scale script will not redirect the traffic until the FortiGate instances are finished initializing and ready to accept traffic. This state can be detected when the logs indicate that FortiGates are “impaired.”

Once the FortiGates reach a healthy state, the route tables will be modified and the “impaired” messages will cease.

This “impaired” state can happen at initialization or after a failover event. If the logs continue to indicate an “impaired” state, you need to determine why the route tables are misconfigured or why an instance continues to be marked as “unhealthy” or “impaired” by AWS.

FAILOVER ALGORITHM

If a primary firewall in an AZ is terminated or becomes unhealthy and is terminated by CloudWatch, the auto scale script will use the following algorithm to modify the route tables and maintain HA within the AZ:

If the primary firewall for an AZ disappears or is “impaired”:
1. Look for a firewall within the same AZ and mark it as primary and change AZ route table to point to secondary interface.
2. Look for secondary firewall in another AZ in auto scale group. If found, change AZ route table to point to secondary interface.
3. Look for primary firewall in another AZ in auto scale group. If found, change AZ route table to point to secondary interface.
4. If none of the above are found, exit and wait for an auto scale event to happen to replace failed firewalls. After a healthy FortiGate is configured and added to the auto scale group is found, rerun failover algorithm.

AUTO SCALE EVENTS/SCALE-OUT

If CloudWatch detects that a scale-out metric has been exceeded for the specified threshold time, a new instance will be spawned and the auto scale script will be notified via the SQS named queue. This will appear in the auto scale log.
The auto scale script will then configure the new instance and add the instance to the auto scale group and the external elastic load balancer. If an event prevents the auto scale script from configuring the FortiGate for any reason, the script can ABORT the life cycle hook for the instance and the instance will be terminated. CloudWatch can then use a number of decision points to decide if a new instance should be created to replace the failed instance. The logs should return to a steady state after the new instance is configured and ready to receive traffic, and all active instances should appear in the list with the proper route table associated with the instance after a couple of iterations.

AUTO SCALE/SCALE-IN

If a scale-in event occurs, the auto scale script will detect and log that a FortiGate instance was terminated, but CloudWatch will not notify the auto scale script via the SQS named queue. CloudWatch will update the auto scale group and the external load balancer and no action is required by the auto scale script.

APPENDIX B. AUTOMATED TEMPLATE DEPLOYMENT AND CODE UPDATES

The worker node has access to the repositories containing updated templates and auto scale script code. The code can be updated from the worker node or scripts deployed from the worker node using the following commands.

WORKER NODE STATUS

1. Worker Node Status
   The current run state of the worker code can be determined using the “worker_code_status” command. If the code is not running, it can be started or stopped using “worker_code_start” or “worker_code_stop.”

2. Worker Code Status
   New revisions of code may be available in the repository. Current status can be retrieved using the “worker_code_status” command.
WORKER CODE UPDATES

1. Worker Code Update
   New revisions of code may be available in the repository. Current status can be retrieved using the “worker_code_status” command.

```
2 package(s) needed for security, out of 5 available
Run "sudo yum update" to apply all updates.
[ec2-user@worker-demostack-ASInstance-ip-10-0-0-92 ~]
# Fetching Repo Status: /aws autoscale_worker/

<table>
<thead>
<tr>
<th>REPO_NAME</th>
<th>STATUS</th>
<th>COMMIT TIMESTAMP</th>
</tr>
</thead>
<tbody>
<tr>
<td>aws autoscale_worker</td>
<td>ok</td>
<td>2017-04-14 14:38:22 +0000</td>
</tr>
</tbody>
</table>
```

2. Worker Code Update
   Code updates can be pulled from the repository using the “worker_code_update” command.

```
[ec2-user@worker-demostack-ASInstance-ip-10-0-0-92 ~]
worker_code_update
```

AUTOMATED TEMPLATE DEPLOYMENT

Template deployment can be automatically deployed directly from the worker node and avoid error-prone deployment by manually clicking through the AWS CloudFormation Console.

1. To get a list of available templates in the Fortinet AWS Repository, use the “aws_env_template list” command:

```
[ec2-user@worker-demostack-ASInstance-ip-10-0-0-92 ~]
aws_env_template list
```
2. To list resources required by a specified template:

3. To deploy a specified template:
   a. aws configure and add AWS account credentials
   b. aws_env_template <template name> create

User will be prompted for values that need to be specified.
APPENDIX C

SCALING GUIDELINES (FORTIGATE ON DEMAND C3 / C4 / M3 INSTANCES)

- Suggested Scale Up / Scale Down Criteria Parameters

<table>
<thead>
<tr>
<th>Template: CPU Utilization</th>
<th>Medium Instance FG-VM01-AWS</th>
<th>Large Instance FG-VM02-AWS</th>
<th>Xlarge Instance FG-VM04-AWS</th>
<th>2Xlarge Instance FG-VM08-AWS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale Up Threshold</td>
<td>80</td>
<td>80</td>
<td>80</td>
<td>80</td>
</tr>
<tr>
<td>Scale Down Threshold</td>
<td>70</td>
<td>70</td>
<td>70</td>
<td>70</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Template: Memory Utilization</th>
<th>Medium Instance FG-VM01-AWS</th>
<th>Large Instance FG-VM02-AWS</th>
<th>Xlarge Instance FG-VM04-AWS</th>
<th>2Xlarge Instance FG-VM08-AWS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale Up Threshold</td>
<td>80</td>
<td>80</td>
<td>80</td>
<td>80</td>
</tr>
<tr>
<td>Scale Down Threshold</td>
<td>70</td>
<td>70</td>
<td>70</td>
<td>70</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Template: Concurrent Sessions</th>
<th>Medium Instance FG-VM01-AWS</th>
<th>Large Instance FG-VM02-AWS</th>
<th>Xlarge Instance FG-VM04-AWS</th>
<th>2Xlarge Instance FG-VM08-AWS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale Up Threshold</td>
<td>320,000</td>
<td>450,000</td>
<td>1,000,000</td>
<td>3,000,000</td>
</tr>
<tr>
<td>Scale Down Threshold</td>
<td>270,000</td>
<td>400,000</td>
<td>800,000</td>
<td>2,400,000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Template: Session Set-Up Rate</th>
<th>Medium Instance FG-VM01-AWS</th>
<th>Large Instance FG-VM02-AWS</th>
<th>Xlarge Instance FG-VM04-AWS</th>
<th>2Xlarge Instance FG-VM08-AWS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale Up Threshold</td>
<td>1,500</td>
<td>8,000</td>
<td>30,000</td>
<td>120,000</td>
</tr>
<tr>
<td>Scale Down Threshold</td>
<td>1,200</td>
<td>6,000</td>
<td>24,000</td>
<td>100,000</td>
</tr>
</tbody>
</table>

APPENDIX D

REMINDERS: AUTOSCALING PRODUCTS / SUPPORTED REGIONS / CONTACT RESOURCES

FortiGate UTM Auto Scaling is based on AWS On Demand

- On Demand is a subscription-based transaction via the AWS Marketplace (Annual / Hourly).
- The “Scaling” instances, which scale in and out per selected criteria, are also On Demand.
- AutoScaling is supported with v5.4.2 only.
- Bring Your Own License (BYOL), the Fortinet-AWS-VM perpetual license, is a supported option.

FortiGate On Demand Instances in AWS Marketplace

- Instance Size: m3. medium, large, xlarge, 2xlarge | c3/c4.large, xlarge, 2xlarge, 4xlarge, 8xlarge
- Instances can be purchased through AWS Marketplace as either Hourly or Annual subscriptions. In AWS Marketplace, https://aws.amazon.com/marketplace/, search ‘Fortinet’ for listings.
- Select your region(s) and subscription preference. Fortinet can provide guidance on instance sizing based on subscriber count, bandwidth requirements and concurrent sessions.
- You may terminate the Hourly On Demand instance at any time to stop incurring charges.
- Annual subscription cancellations or downgrades are not supported. If you need help with or want to upgrade your subscription, please contact AWS Customer Support.

FortiGate BYOL-VM Licenses in AWS

- FortiGate SKUs: FG-VM01-AWS, FG-VM02-AWS, FG-VM04-AWS, FG-VM08-AWS, FG-VM16-AWS, FG-VM32-AWS
- See www.Fortinet.com/AWS for more detail
- AWS BYOL-VMs can be purchased through certified Fortinet resale partners, contact Fortinet Sales at awssales@fortinet.com or call our worldwide offices for more information
Fortinet Support for your FortiGate

- The FortiGate-VM for AWS hourly and annual subscription includes full FortiGuard UTM services that provide fully automated updates to ensure protection against the latest application layer threats in addition to online and phone support.

- When contacting Fortinet Support, please be ready to provide:
  1. The serial number of your FortiGate-VM instance
  2. The email ID of your Fortinet account

  If you do not have an account yet, please sign up at https://support.fortinet.com/login/CreateAccount.aspx
  http://www.fortinet.com/support/contact_support.html

Fortinet Global Sales Offices

<table>
<thead>
<tr>
<th>Americas</th>
<th>EMEA</th>
</tr>
</thead>
<tbody>
<tr>
<td>US: +1-866-868-3678</td>
<td>Austria: +43 1 22787 120</td>
</tr>
<tr>
<td>US Federal Government: +1-703-915-3817</td>
<td>Belgium/Luxembourg: +32 (0)2 716 49 27</td>
</tr>
<tr>
<td>Canada: +1-866-868-3678 press 1, then press 5</td>
<td>Czech Republic: +420 773 788 788</td>
</tr>
<tr>
<td>Latin America</td>
<td>Finland: +358 40 744 9531</td>
</tr>
<tr>
<td>Mexico: +52-(55) 5524-8428</td>
<td>Germany: +49 69 310 192 0</td>
</tr>
<tr>
<td></td>
<td>Ireland: +353 1 6087703</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>APAC</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia: +61 2 8007 6000</td>
<td>Austria: +43 1 22787 120</td>
</tr>
<tr>
<td>China: 8610-6296 0376</td>
<td>Belgium/Luxembourg: +32 (0)2 716 49 27</td>
</tr>
<tr>
<td>Hong Kong: +852 3708 3500</td>
<td>Czech Republic: +420 773 788 788</td>
</tr>
<tr>
<td>India: 080 41321689/ 99</td>
<td>Finland: +358 40 744 9531</td>
</tr>
<tr>
<td>Indonesia: (62) 21 2358 4548</td>
<td>France: +33-1-8003-1655</td>
</tr>
<tr>
<td>Japan: +81-3-6434-8531</td>
<td>Germany: +49 69 310 192 0</td>
</tr>
<tr>
<td>Korea: +82-2-559-9500</td>
<td>Ireland: +353 1 6087703</td>
</tr>
<tr>
<td>Malaysia: +603-2723 0300</td>
<td>Israel: +972 77-6935670</td>
</tr>
<tr>
<td>Philippines: +632-808-8798</td>
<td>Italy (Rome): +39 06-51573-330</td>
</tr>
<tr>
<td>Singapore: +65-6395-2788</td>
<td>Italy (Milano): +39 039 687211</td>
</tr>
<tr>
<td>Taiwan: +886-2-27961666</td>
<td>Poland: +48 22 449 00 29</td>
</tr>
<tr>
<td>Thailand: +66 2 658 658 12</td>
<td>Russia: +7 499 9552499</td>
</tr>
<tr>
<td></td>
<td>Saudi Arabia: +966 1 261 1402</td>
</tr>
<tr>
<td></td>
<td>Spain: +34 915 024 874</td>
</tr>
<tr>
<td></td>
<td>Sweden: +46 70 237 9090</td>
</tr>
<tr>
<td></td>
<td>Switzerland: +41 44 833 68 48</td>
</tr>
<tr>
<td></td>
<td>The Netherlands: +31 (0)33 454 67 50</td>
</tr>
<tr>
<td></td>
<td>Turkey: +90 (216) 250 3259 / 60</td>
</tr>
<tr>
<td></td>
<td>United Arab Emirates: +971 4 423 9601</td>
</tr>
<tr>
<td></td>
<td>United Kingdom: +44 (0) 203 752 6880</td>
</tr>
</tbody>
</table>