Global Information Assurance Certification Paper

High Availability Firewall – WatchGuard Firebox Vclass V60

SANS Security Essentials, GSEC Practical Assignment
Name: Wee Leng, CHIA
Date Submitted: July 3, 2003
Version Number: 1.4b, Option 1

Abstract

Availability is but one of the three cornerstones in information security: Confidentiality, Integrity, and Availability. Nevertheless, its importance cannot be undermined. With the advent of simplified computing technologies, it is not impossible to achieve a high availability firewall setup within a reasonably short span of time. The focus of this paper is on the subject of high availability (HA).

It kicks off by acquainting oneself with the term HA, analyzing the need for HA, categorizing the modes of HA, understanding the technicalities of HA, and finally setting up an HA model based on the WatchGuard Firebox Vclass V60, including troubleshooting procedures. It wraps up by emphasizing the fact that high availability is not the sole factor for total system reliability. Interdependency between other factors plays a key role in ensuring the availability aspect of information security.

Preface

In the era of Information Technology:

> High Availability (HA) refers to a system or component that is continuously operational for a desirably long length of time. Availability can be measured relative to "100% operational" or "never failing." A widely-held but difficult-to-achieve standard of availability for a system or product is known as "five 9s" (99.999 percent) availability[1], equivalent to five (5) minutes of downtime per year.

With reference to Layer 3 (Network Layer) of the OSI[2] (Open Systems Interconnection) Model, High Availability can be implemented on all three (3) levels namely WAN (Wide Area Network), Security and LAN (Local Area Network), as depicted in *Figure 1*. This research paper focuses on the implementation of High Availability on the Security level (c.f. Security Block).

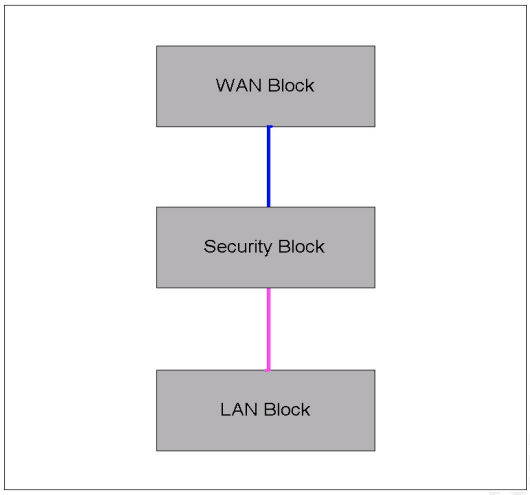

Figure 1: Generic Network Block Diagram

Understanding High Availability (HA)

Figure 1 will be used as the basis in grasping the concept of High Availability. Based on this foundation, it is easy to visualize the concept of a single-firewall network. This is done by replacing the Security Block with a Firewall, as displayed in *Figure 2*.

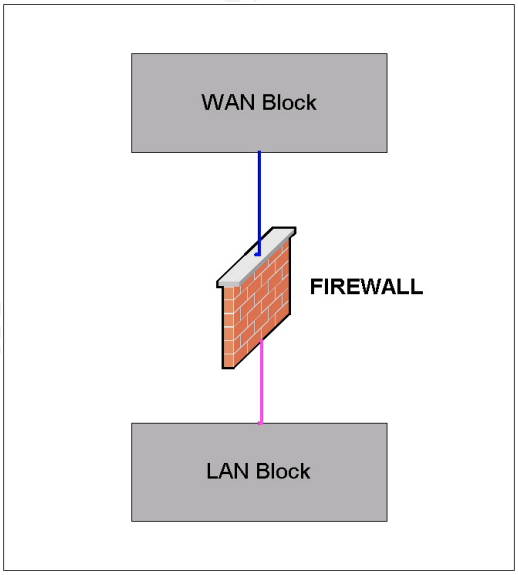

Figure 2: Single Firewall Model

Similarly, in a high availability secured network environment, more than one firewall is placed in the Security Block. The first to the N-th firewalls are what make up the firewall cluster for High Availability, as visualized in *Figure 3*.

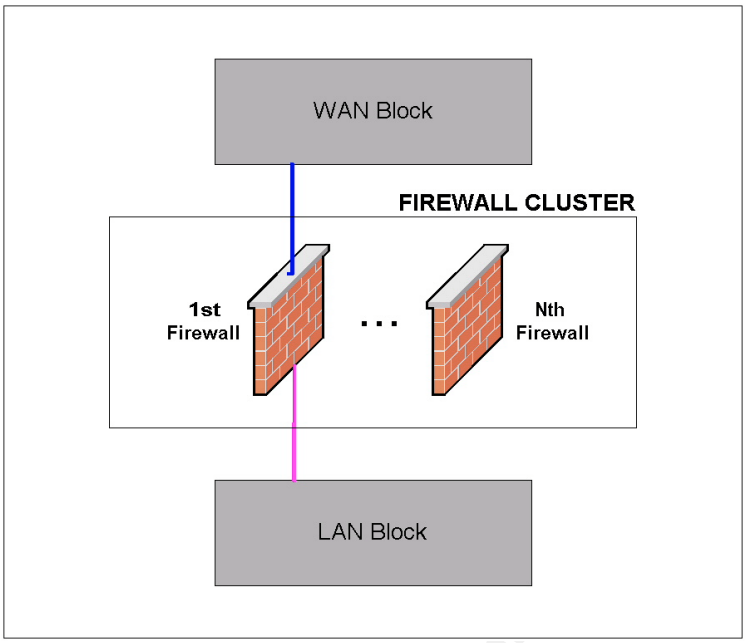

Figure 3: High Availability Firewall Model

The Need for High Availability System

How essential is a high availability system to an organization? Many questions arise as to whether a redundant firewall configuration is really necessary, doubting the Returns On Investment (ROI) of an HA implementation. The feasibility, however, very much depends and boils down to the business criticality, security policies and other requirements of an organization.

There are some main justifications to a high availability firewall implementation. The firewall, as the first line of defense, provides perimeter security for all data, voice and video networks and beyond. Hence, the firewall needs to be up and running at all times in order to avoid any disruptions to any organization's network. Should the firewall fails, essential productivity tools such as e-mail and the Internet (World Wide Web), among others, will not be accessible to users of an organization, thus plummeting productivity and causing the organization to suffer from other negative consequences.

It is also obvious that the Internet has become a necessary part of everyday life for many people. It is because of this heavy reliance on Internet services on the businesses' and consumers' part that further propels the need to implement high availability firewalls, the world over. The expected availability, therefore, is 24 x 7 (24 hours a day, 365 days a year). Downtime is no longer tolerable, whether it is due to failures (unplanned) or scheduled maintenance (planned), and is deemed unacceptable.

This point is strongly reiterated by a Gartner Group report which "predicts that the proportion of enterprise applications requiring 24x7 availability will nearly triple in the next five years, growing from less than 10% in 2001 to more than 25% by 2005"[3].

High Availability Classification

Basically, high availability can be categorized into three modes, namely Cold Standby, Hot Standby, and Load Balancing Cluster[3].

In a *Cold Standby* mode, the backup (standby) is offline[3]. Thus, in the event that the primary device fails, human intervention is required to manually substitute the standby device. Although this is the lowest-cost option, it comes with other setbacks such as the inability to automatically update the configuration of spare (standby) device, should there be any changes to the primary device.

In a *Hot Standby* solution, similar to the Cold Standby solution, only one link actually handles network traffic. The difference is that in a hot standby mode, also known as Active/Standby HA, the backup (standby) is online, but not handling any live data[3]. This standby device monitors the primary device, whereby it automatically takes over the functions of the primary device should the primary be taken offline in the event of any failures. The advantage of having this solution, as compared to the cold standby solution, is that human intervention is not required in this case. Besides, it is also possible to achieve 99.9% or higher availability via this mode[3].

On the other hand, the better of the three modes, the *Load Balancing Cluster*, also known as Active/Active HA, provides manifold layers of redundancy. This is possible as all nodes (devices) are active and each carries portion of the load. Therefore, if one fails, others automatically take over. With this mode, it is possible to achieve 99.999% or higher availability[3]. Having said that however, this method is usually more costly than the simple standby solutions[3], i.e. cold standby or hot standby. From a different point of view, this investment is fully leveraged, as the backup systems do not sit idle; instead it is being fully utilized.

Based on the above three classifications, the Hot Standby mode (Active/Standby HA) on WatchGuard Firebox Vclass V60 will be used as the focus of this research.

WatchGuard Overview

WatchGuard Technologies (www.watchguard.com), a leading provider of dynamic, comprehensive Internet security solutions[4], was founded in February 1996. To date, its new product line includes WatchGuard Firebox Vclass, in addition to the established WatchGuard Firebox System.

Mechanics of High Availability – How does it work?

This section delves into the workings of high availability in a WatchGuard firebox appliance, with particular emphasis on Firebox Vclass V60, based on the Hot Standby mode (Active/Standby HA).

> When HA is active, the Primary (Active) appliance sends a "heartbeat" to a Secondary (Standby) appliance. This heartbeat tells the Secondary appliance that the Primary appliance is still "alive," or up. If the primary appliance fails, the heartbeat ceases. When the Secondary appliance detects three consecutive missed heartbeats, it assumes all processing tasks[5].

"In a WatchGuard Firebox V60 HA system, two firebox Vclass appliance are connected so that one serves as a ready backup to the other if the Primary appliance fails while managing network traffic"[6]. When HA is first activated at the Primary device, the MAC addresses of all interfaces are changed. The Secondary is contacted via the HA1 port (c.f. *Figure 4*). When the Primary synchronizes with the Secondary device, a complete set of configurations and policies are sent to the Secondary device.

The Secondary device restarts and begins to listen to heartbeats sent out by the Primary device. The respective interfaces on the Primary and Secondary are set to the same MAC addresses. The Primary device sends heartbeats while the Secondary device monitors the heartbeats. The Secondary device does not send or receive packets. This, however, has no impact on the system throughput. High availability provides a fail over process, which occurs after five (5) seconds of inactivity, in the event of a Firebox Vclass failure.

When the Primary device fails, upon detecting a link down or firmware failure, it stops sending heartbeats. If three heartbeats are lost, the Secondary device takes over[6]. All Media Access Control (MAC) addresses of the interfaces (Primary and Secondary) are the same. Clients need not do an Address Resolution Protocol (ARP) for the Secondary device.

When a Primary device comes back, the Primary (now receiving heartbeats) does not preempt the Secondary. The Primary device takes over if no heartbeat is received from the Secondary device. However, if both the Primary and Secondary devices are sending heartbeats at the same time then the Secondary device yields to the Primary device.

Virtual Router Redundant Protocol (VRRP) provides an Internet Engineering Task Force (IETF) standards based approach to having mirrored configurations on two routers so that when one ceases to function, a backup takes over. "VRRP allows both HA security appliance to share the same MAC and Internet Protocol (IP) address"[6]. For instance, in a Firebox V60, the hostname and IP address is the same for both Primary and Secondary device. In a similar vein, the MAC address for both Primary and Secondary is identical; using the VRRP defined MAC addresses. "The advantage gained from using VRRP is a higher availability default path without requiring configuration of dynamic routing or router discovery protocols on every end-host"[7].

Configuring HA on WatchGuard Firebox Vclass V60

In the subsequent sections, the proposed high availability design diagram will be used as the model example for the complete setup of HA on Firebox Vclass V60. The Firebox V60 is recommended for Large or Mid-Size Enterprises[8].

1. HA Model Overview

Figure 4, adopted from *Figure 3*, represents the model design diagram for the high availability implementation. For the purpose of this example, three (3) interfaces on the WatchGuard Firebox V60 will be used, namely the External (EXT) interface, Internal (INT) interface, and the De-Militarized Zone (DMZ) interface where the web server is normally located.

The Firebox V60 comes with four (4) accelerated, Ethernet interfaces (RJ-45 connectors) labeled 0 *(Private*), 1 *(Public*), 2 *(DMZ*1), and 3 *(DMZ2)* which act as the primary conduits through which passes all of the network data traffic[9]. Another two Ethernet interfaces, labeled *HA2* and *HA1* are connected with a crossover Ethernet cable to the other HA-ready Firebox V60 for fail over (redundancy) protection. This corresponds to the Firebox V60 hardware interfaces as follows:

- E0: Internal VLAN
- E1: External VLAN
- E2: DMZ VLAN
- E3: Not used
- HA1: HA Link
- HA2: Not used

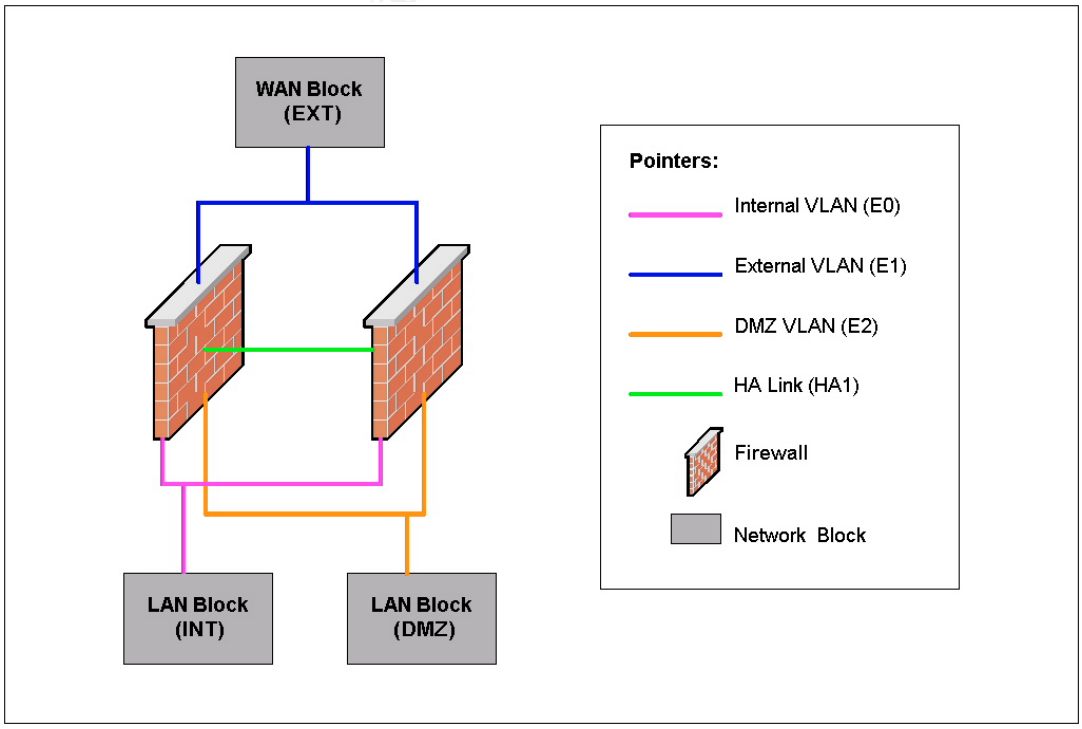

Figure 4: Model HA Design Diagram

2. Physical Setup: Preparing the Equipment, Connecting the Devices

In order to connect the devices physically, the logical diagram as depicted in *Figure 4* has to be interpreted correctly.

Method 1

Based on a straightforward interpretation, using the application of three separate physical LANs, *Figure 4* implies that both the firewalls are connected to an Internet-facing router on the External interfaces of the firewalls (E1) via a hub. This router is then linked to an Internet Service Provider (ISP) via a leased line. On the other end, the firewalls are also connected to the internal Local Area Network (LAN) on the Internal interfaces (E0) via another hub, and the optional LAN on the DMZ interface (E2) via a different hub. A connection between the two firewalls on the HA1 interface is required for HA functionality.

In summary, the equipment and materials needed for this configuration includes two (2) units of WatchGuard Firebox V60, three (3) units of hubs, one (1) unit of router, two (2) units of switches (for INT LAN and DMZ LAN respectively), nine (9) units of RJ-45 straight-over color-coded cables: 3– blue (EXT) 3 – pink (INT), 3 – orange (DMZ), one (1) unit of RJ-45 cross-over cable (for HA link).

Method 2

Alternatively, using the application of Virtual LAN (VLAN), *Figure 4* implies that both the firewalls are connected to an Internet-facing router on the External interfaces (E1) via a VLAN-enabled switch. An example of this switch is the Cisco Catalyst 3550 switch. This router is then linked to an Internet Service Provider (ISP) via a leased line, similar to *Method 1*. On the other end, the firewalls are connected to the internal Local Area Network (LAN) on the Internal interfaces (E0) via the same switch. The same applies to the optional LAN on the DMZ interface (E2). A connection between the two firewalls on the HA1 interface is required for HA functionality.

In summary, the equipment and materials needed for this configuration is similar to *Method 1* with the exception of replacing the three (3) units of hubs with one (1) unit of VLAN-enabled switch. The switch is configured for three (3) VLANs – INT VLAN, EXT VLAN and DMZ VLAN.

3. Logical Setup: Configuration on Vcontroller v4.0

The Logical Setup for Firebox V60 will be based on the Vcontroller application version 4.0, to be installed on a Windows workstation. The version that precedes v4.0 is v3.2. The Vcontroller v4.0 is installed on a machine with the following prerequisites[5]:

Operating System	: Windows 98/ME/NT 4.0/2000/XP
CPU	: Pentium II or later
Processor speed	: 500 MHz or faster

Memory	: 64 MB minimum (128 MB is recommended)
Input device	: CD-ROM or DVD
Hard disk space	: 10 MB minimum Additional space as required for log files, backup and archive configuration files
Network interface	: NICs or embedded network connections
Other requirements	: Java Run-time Environment (JRE) and Java Development Kit (JDK)

This machine, which serves as the firewall Management Station, is connected to the management station to a hub or switch that is connected to interface E0 (Internal LAN) on both firewalls. The management station can also be connected to an HA2 port. This management station is the primary administrative access to the firewalls.

It is assumed that on the Primary Firebox, the basic system configuration has been completed under the system administration column of the Vcontroller prior to configuring the HA options for the Firebox V60. This includes the configuration of interface IP addresses, routing table, as well as the security policies, among others.

With this assumption in place, and with all devices physically connected for HA, the Primary Firebox is powered on. At this point, the Secondary Firebox is not powered on and is still in factory default configuration. The Primary Firebox is configured using the executed Vcontroller application. From the System Configuration menu, the High Availability tab is selected. The “Enable High Availability” checkbox is checked to turn on the HA functionality on both the firewalls.

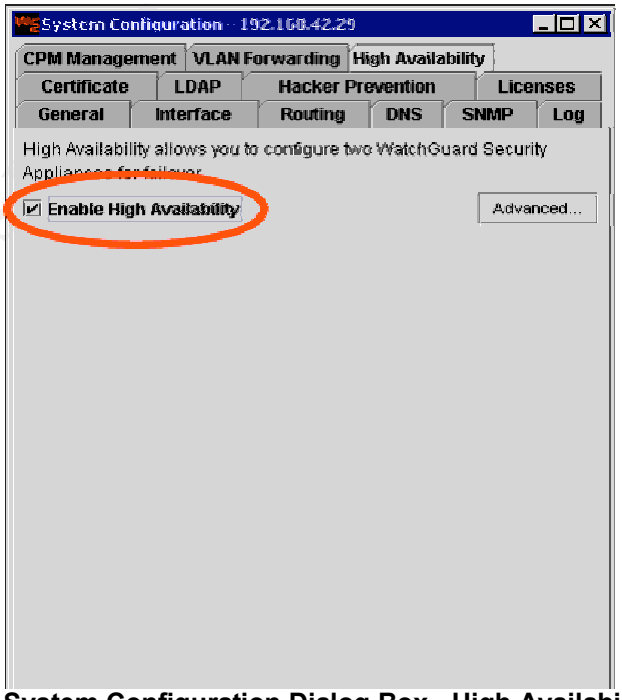

Figure 5: System Configuration Dialog Box - High Availability Tab[6]

Under the Advanced HA Parameters dialog box (obtained by clicking on "Advanced…" button under High Availability Tab), administrators are able to control which interfaces are monitored for fail over, advertisement interval, HA group ID, and interface IP assignments. If the HA2 port was configured for management via the system configuration tab then it cannot be used for fail over. The advertisement interval is also known as the heartbeat; the period of time a signal is sent to the other device to make sure it is functioning properly. The HA group ID specifies a group of devices that back up one another. This identification applies if more than one HA group exists on the same network.

Although it is possible to turn off monitoring for the DMZ and public ports, the private interface is monitored by default and cannot be turned off. Any interface that needs to be HA ready must be checked. In addition, any interface that needs to be HA ready must have that interface (public, private or DMZ) connected with a hub, and from the hub to the network device (hub, switch, server...). Otherwise, a failure of that interface does not prompt HA for fail over.

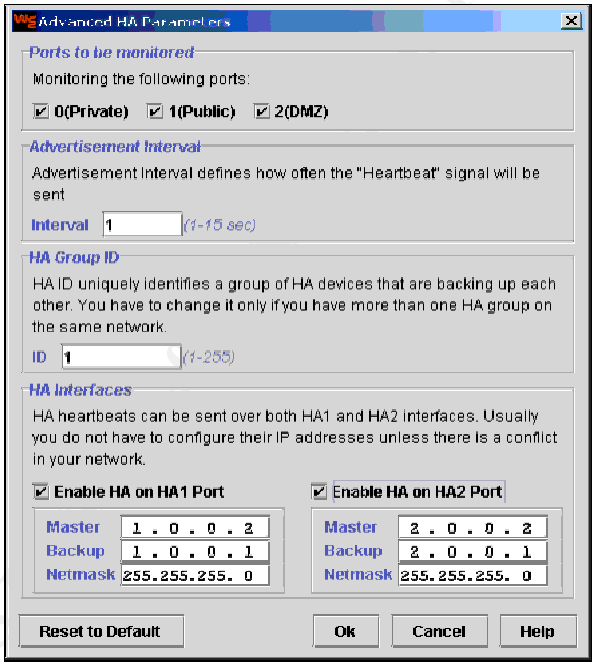

Figure 6: Advanced HA Parameters[6]

Finally, the database on the Primary Firebox needs to be synchronized with the Secondary Firebox. At this point, turn on the Secondary Firebox and click on the Hotsync button. The Primary Firebox then attempts to communicate with the Secondary Firebox. If successful, the window displays the "Ready for fail over" message. The rule of thumb is to execute HotSync each time a change is made to the Primary firebox.

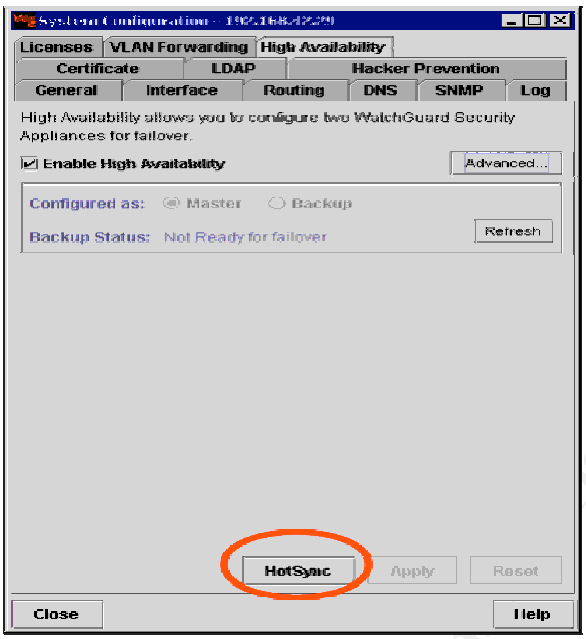

Figure 7: Synchronize Both Fireboxes[6]

4. Testing the HA Setup

It is the normal practice in any network/system implementation that testing and commissioning be conducted to verify complete functionality of the network/system setup.

In substantiating the WatchGuard Firebox V60 high availability setup, the objective would be to simulate a failure on the Primary device, thus the expected end result would be for the Secondary device to take over the operations of the Primary device seamlessly.

The simplest way to achieve this is to power off the Primary device. This power failure signifies a device failure. The Secondary device should resume all firewall functionality within five (5) seconds of inactivity.

Another way to test the HA setup is to deliberately reset the Primary device to Factory Default configuration. By restoring the device to factory default, all configuration files and database information will be wiped out completely from the device. Without these files and information, the device will cease to function. This will in return prompt the Secondary device to take over from the Primary device. Again, consistent to the first test method, the secondary device should resume all firewall functionality within five (5) seconds of inactivity.

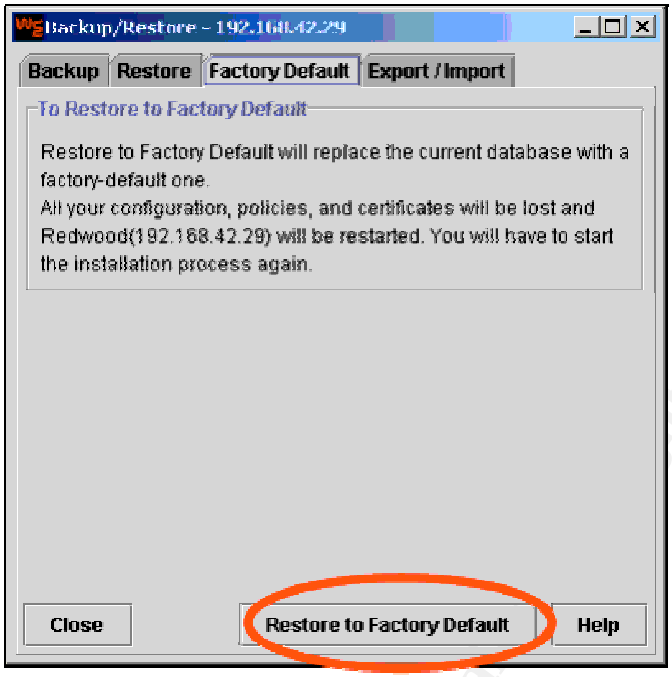

Figure 8: Restoring Firebox to Factory Default[6]

5. What Went Wrong? How Do I Rectify It?

In the process of setting up HA on WatchGuard Firebox V60, problems or hiccups may or may not occur. The possible problems and suggested solutions to tackle the underlying problems are put forth in this part.

One of the common mistakes, often overlooked, which causes the failure of detecting the secondary device by the primary device, lies in the physical connections of devices. This can be resolved by performing troubleshooting analyses on the Physical Layer (Layer 1) of the OSI model. The first step is to ensure that the connection links both HA1 ports on the primary and secondary devices. Once this is verified, check that an RJ-45 cross over cable is used to connect both HA1 ports between the primary and secondary devices. More often than not, without realization, an RJ-45 straight over cable is misused in place of a cross over cable! Once done, simply click on the Refresh button to redetect the secondary device.

One other possible problem during the high availability setup is the inability to perform a HotSync operation. Sometimes, users cannot seem to click on the HotSync button to duplicate the entire configuration and policy database from the primary device to the secondary device. This happens primarily because the secondary device has not been powered on. Without activating the secondary device, the HotSync button is inactive, and the status indicator in the High Availability tab does not display an “OK” message. Therefore, always ensure that the both HA devices are turned on before any HotSync operation is conducted.

Conclusion

To put it in a nutshell, the implementation of high availability firewalls in itself cannot be considered sufficient to ensure overall system reliability. Undeniably, the surest way to eliminate all single points of failure is to replicate the entire system with transcendent technology, including hardware, software, and connectivity. Nonetheless, the ultimatum is to be able to strike a balance between cost, reliability, performance and security.

List of References:

[1]**"Definitions - High Availability"**. 11 Apr. 2003. URL: http://searchdatabase.techtarget.com/gDefinition/0,294236,sid13_gci761219,00.html (June 30, 2003)

[2]**"RFC1042 – A Standard for the Transmission of IP Datagrams over IEEE 802 Networks"**. February 1988. URL: http://www.faqs.org/rfcs/rfc1042.html (May 1, 2003)

[3]**"White Papers: Achieving High Availability"**. Fall 2001. URL: http://www.rainfinity.com/products/wp_achieving_ha.html (May 1, 2003)

[4]**"Corporate Information - WatchGuard Technologies, Inc".** 2003. URL: http://www.watchguard.com/corporate/index.asp (May 2, 2003)

[5]**"Firebox Vclass High Availability Guide v4.0".** (v40VclassHAGuide.pdf - PDF Format). 2003. URL: http://help.watchguard.com/documentation/vclass.asp (Available from the "Product Documentation" option) (June 30, 2003)

[6]**"Firebox Vclass User Guide – Vcontroller 4.0".** (v40VclassUserGuide.pdf - PDF Format). 2003. URL: http://help.watchguard.com/documentation/vclass.asp (Available from the "Product Documentation" option) (June 30, 2003)

[7]**"RFC2338 – Virtual Router Redundancy Protocol".** April 1998. URL: http://www.faqs.org/rfcs/rfc2338.html (May 1, 2003)

[8]**"Firebox Vclass Brochure".** (v_bro_ltr.pdf - PDF Format). 2003. URL: https://www.watchguard.com/infocenter/brochures.asp (Available from the "Brochures & Datasheets" option) (June 30, 2003)

[9]**"Firebox V10, V60, V80, V100 Hardware Guide"**. (v40VclassHardwareGuide.pdf - PDF Format). 2003. URL: http://help.watchguard.com/documentation/vclass.asp (Available from the "Product Documentation" option) (June 30, 2003)