

Nokia IP Clustering

January 2007

Table of Contents

INTRODUCTION	3
CLUSTERING SOLUTION	3
CLUSTERING VS. VRRP	4
CLUSTERING VS. BLADED SYSTEMS	4
TYPICAL CLUSTER TOPOLOGY	4
HOW IT WORKS	6
THE CLUSTER PROTOCOL	6
FORMING AND MAINTAINING THE CLUSTER	6
WORK LOAD ASSIGNMENT (BALANCING)	7
DYNAMIC LOAD BALANCING VS. STATIC LOAD SHARING	8
HEALTH CHECKS / ACTIVE SESSION FAILOVER	9
CLUSTER CONFIGURATION	10
IP CLUSTERING SPECIFICATIONS: (SUBJECT TO CHANGE)	14
PERFORMANCE SCALING	15
CONCLUSION	15

Introduction

Modern networks are carrying both the financial transactions and the information that is the lifeblood of today's business, which makes network outages (or even degraded performance) unacceptable. Just as the networks are evolving so have the frequency and sophistication of malicious attacks against those networks. Providing secure communications to the customers and employees has become a high priority for CIOs and security officers. And to provide secure communications various security appliances are deployed at key points in a network.

Failure of these security appliances is a painful reality that leaves critical systems and data exposed or unavailable. Down time in the network leads to a loss of productivity for employees, customer frustration, negative impact on business reputation and significant loss in revenue. Various high availability solutions have been developed by different vendors to ensure the availability of security appliances. Nokia's patented Clustering technology provides an unmatched, robust and scalable high availability solution for security appliances.

Clustering Solution

Nokia's Clustering has been developed to address the aforementioned business challenges. Clustering allows several independent appliances to join together for a common security goal as one virtual machine. In addition to processing network traffic in parallel, clustered appliances share information about the context of that traffic to enable the cluster to survive the failure or degradation of any of its individual appliances. By dividing and conquering, clustering can allow several appliances to work in concert to take on a task that would tax any single member. And all the appliances can be centrally managed from one location.

Nokia's clustering solution provides the following to its security appliances:

Scalability, Strength in numbers allows for easy scalability. As increasing traffic is placed on modern networks, network administrators can add cluster members (N+1) to divide the increased load among more devices, ensuring that every device can handle the load assigned to it.

Availability, On the rare occasions when problems develop with an appliance, its workload is transparently redistributed to the surviving cluster appliances without disrupting communication throughout the cluster.

Resiliency and fault tolerance in clusters is based on the statistical improbability of multiple simultaneous failures. Transparent workload redistribution also makes active appliance maintenance possible. Administrators can perform transparent "rolling upgrades," in which nodes are gracefully removed from the cluster, upgraded, and reinserted, all without any disruption to end-user operations.

Clustering vs. VRRP

Virtual Router Redundancy Protocol (VRRP) is another solution for availability. In VRRP a hot-standby appliance is deployed in which one active appliance would handle the entire incoming traffic load with a backup appliance ready to assume its functions in case of problems. However, single-appliance solutions do not provide the scalability to handle the load of modern networks.

To provide maximum possible security, network packets have to be carefully inspected by security applications. Deep packet inspection causes severe overhead on the security appliances. In contrast to VRRP additional appliances can be added to the cluster to maintain the same levels of performance while delivering the required inspection. By its nature, clustering adds scalability. When the cluster is reaching its capacity limitations, additional cluster members (N+1) can be added to increase performance. Also, Clustering provides sub-second fail-over while with VRRP the fail-over time to the stand by appliance is usually a few seconds.

Clustering vs. Bladed Systems

A bladed system is another solution for availability. Bladed system consists of a chassis with multiple blades serving different functions. A chassis typically would have a network blade (consisting of ports and traffic forwarding capabilities), an application blade (running security applications) and a management blade (used for management). Application blades provide availability and in some cases, load balancing.

Bladed systems are typically expensive solutions to manufacture, which directly translates to the level of customer investment. Price to performance ratios for bladed systems is very high. The price of each blade can be as high as a single appliance. Bladed systems are usually at least 5U with the performance being in par or less than a comparable security appliance.

Security Appliances are now running multiple security applications, which provide multiple layers of security (unified threat management). These security appliances are now capable of delivering multiple gigabit performance and come in form factor as low as 1 U for a nominal price compared to bladed systems. A security appliance cluster is also known to scale especially when the traffic is CPU intensive. So, it is becoming less popular to use a bladed system solution when security appliances can be used as a centrally managed cluster solution to provide not only availability but also sustained scalability.

Typical Cluster Topology

From the outside, the cluster looks just like a single gateway – it has a single-system image. The cluster has two or more network interfaces; each with its own unique IP address – “cluster” or “virtual” addresses that don’t change regardless of the internal makeup of the cluster. Strictly internal to the cluster are one or two dedicated cluster protocol networks. Security application’s synchronization traffic can also travel over these

networks, except in the most connection-intensive traffic environments, where a separate traffic network is preferable.

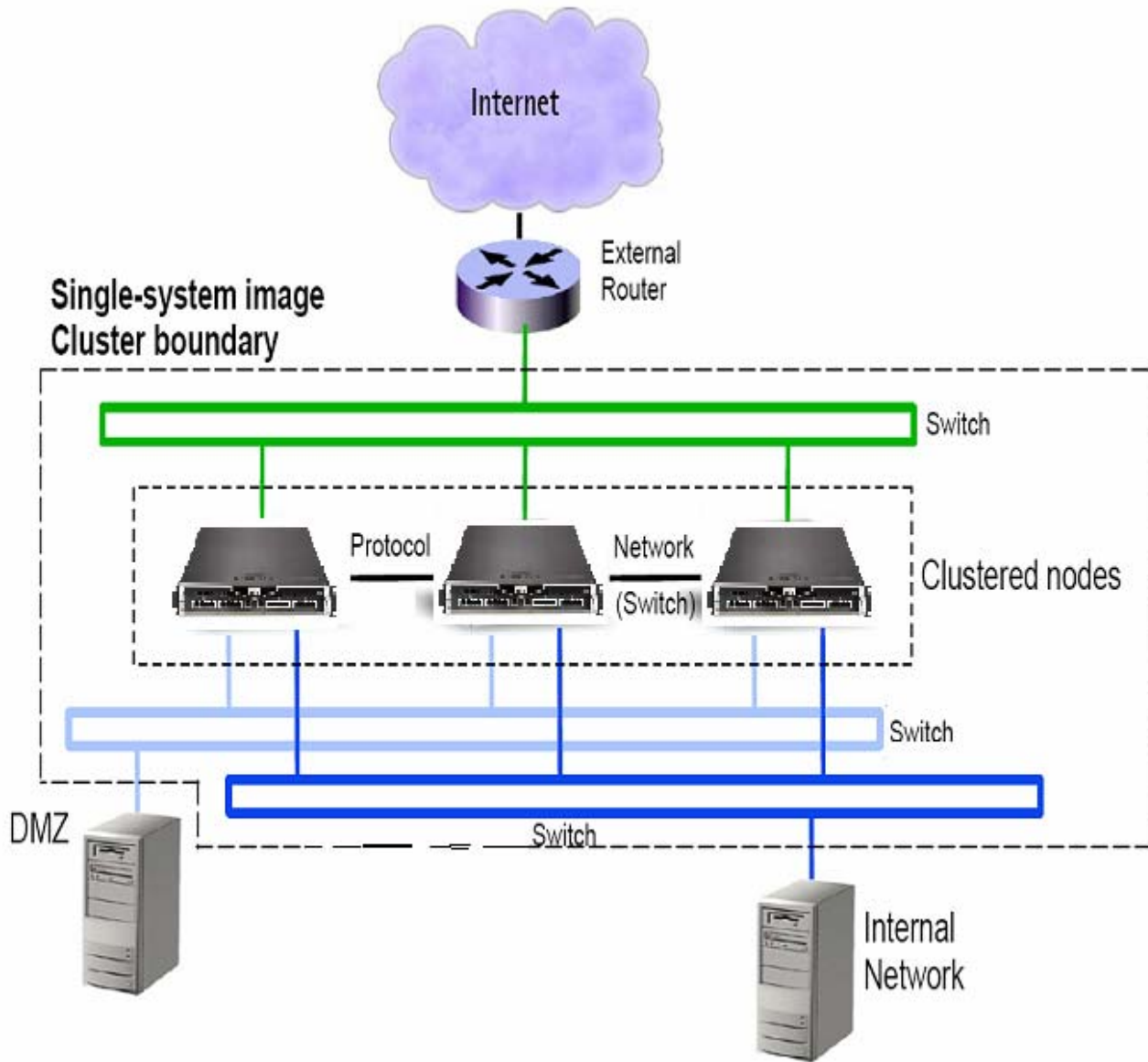


Figure 1: A typical three-node, three-interface cluster

There are three clustered network interfaces: Internet, Internal Network, and DMZ. Each clustered network interface is assigned a unique cluster IP address – a “virtual” IP address that is shared by the respective “real” interfaces on each of the three cluster nodes. Switches aggregate each set of real interfaces into a single virtual cluster interface.

In addition to clustering security applications, the cluster serves as a single-system image router supporting both static routes and dynamic routing protocols including OSPF, BGP, and PIM-SM/DM. Integral routing helps to create a complete high available gateway, manageable as a single-system image.

How it Works

Nokia's IP Clustering technology is described in great detail in United States Patents 6,006,259 and 6,078,957. These patents can be downloaded at no charge from the U.S. Patent Office web site at <http://patft.uspto.gov>.

The Cluster Protocol

Each security appliance (node) in the cluster participates in the cluster protocol. The protocol is the language that cluster nodes use to form the cluster, to dynamically join and leave the cluster, detect node failures, and assign work to the cluster nodes in order to balance the load.

This cluster protocol is separate and distinct, but used in conjunction with a security application's state synchronization protocol. The state sync protocol replicates connections state and tunnels state across the cluster so that upon fail-over resulting from a defective or manually removed appliance that appliance's work can be instantly re-assigned to another appliance.

Forming and Maintaining the Cluster

When the cluster is first formed (e.g. on boot up of one or more appliances configured for a cluster), an election process selects one of the appliances as the *master*. If only one appliance is present initially, it elects itself. If two or more nodes are present *simultaneously* initially, the higher performance hardware appliance wins the election. If two or more highest performance appliances are of the same appliance model, then their IP address is used as a tie-breaker.

The master is responsible for assigning workload to cluster nodes. Additional nodes can join the existing cluster at any time by sending a join request; the master accepts the node into the cluster and begins assigning it work.

On joining the cluster, a node installs the pre-configured cluster IP addresses, multicast cluster MAC addresses or unicast cluster MAC addresses, and MAC filters on its interfaces. The MAC filters enable the node to receive all Ethernet frames sent to the cluster IP address / multicast cluster MAC.

Cluster protocol messages are sent as IP packets identified by next layer protocol 144. They are sent to multicast group 224.0.1.144. After the cluster forms, protocol messages continue to be exchanged to assign and change work assignments and to detect cluster node failure or removal.

Work Load Assignment (Balancing)

Nokia IP Clustering occurs at the IP layer. A particular node handles packet flows through the cluster. This provides a clean division of work for the cluster nodes. Nokia IP Clustering is performed by 'session balancing' data flows in such a way that flows are assigned to be handled by a particular node in the cluster. All the communication and handling of flows is achieved on a per packet level. IP packets are received directly and flows are processed by a particular node within the cluster. Each node has flows dynamically assigned to it and that node is responsible for handling the processing of those flows through the cluster. Also all intra-cluster communication is performed using the industry standard IP protocol. This has two major benefits: communication can be done over standard and existing interfaces and the loss of a packet, for whatever reason, is easily dealt with through standard IP-based mechanisms. This is a huge improvement over traditional clustering. Also, with the high overhead of traditional clustering, additional nodes don't scale linearly.

Each node updates the other members as to the state of all flows it is responsible for processing, so that at any time the flow could be moved to another node for processing. This state information includes data such as TCP sequence numbers, IPSec security associations and all of their related information. There is very little overhead because only changed state needs to be distributed across the cluster. Complete and reliable state replication results in Active Session Failover.

The Nokia IP Cluster handles IP flows on a cluster-wide basis. Whenever a new session or flow is established, information about that session is distributed among all nodes in the cluster. Because all session and tunnel state information is distributed to all nodes in the cluster and the availability of nodes is constantly monitored, the failure of one device will not cause noticeable disruption of traffic.

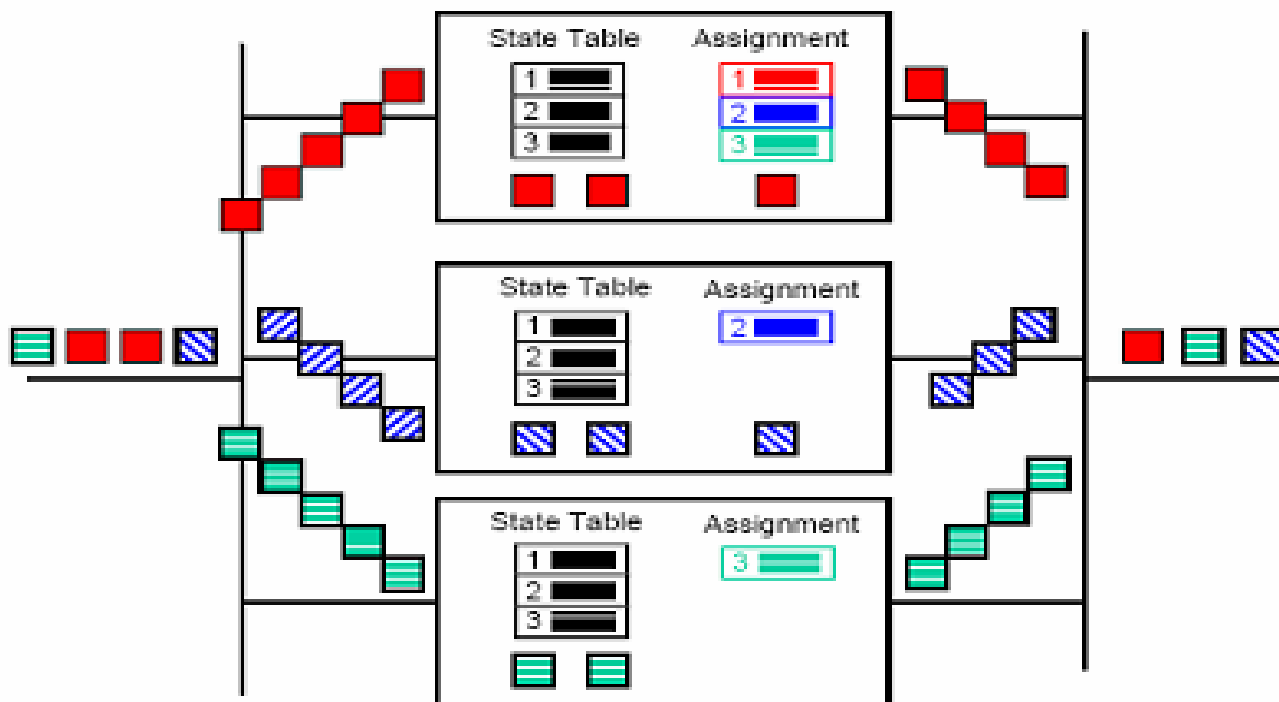


Figure 2: Work Load Assignment

Dynamic Load Balancing vs. Static Load Sharing

Dynamic – After initial assignment, if the workload of other nodes increases or decreases significantly, the master will reassign work among the cluster nodes to restore relative balance.

Static – After initial assignment to a particular node, it retain its workload for the life of the session.

Node workload is distributed dynamically, using an algorithm that assigns new work to the least-loaded node. As workload assigned to a node gets completed (sessions expire), the node becomes available to take on new work and work from other nodes that are more heavily loaded.

The effect of dynamic load balancing and rebalancing is illustrated below and compared with static load sharing (which is implemented more commonly by other vendors).

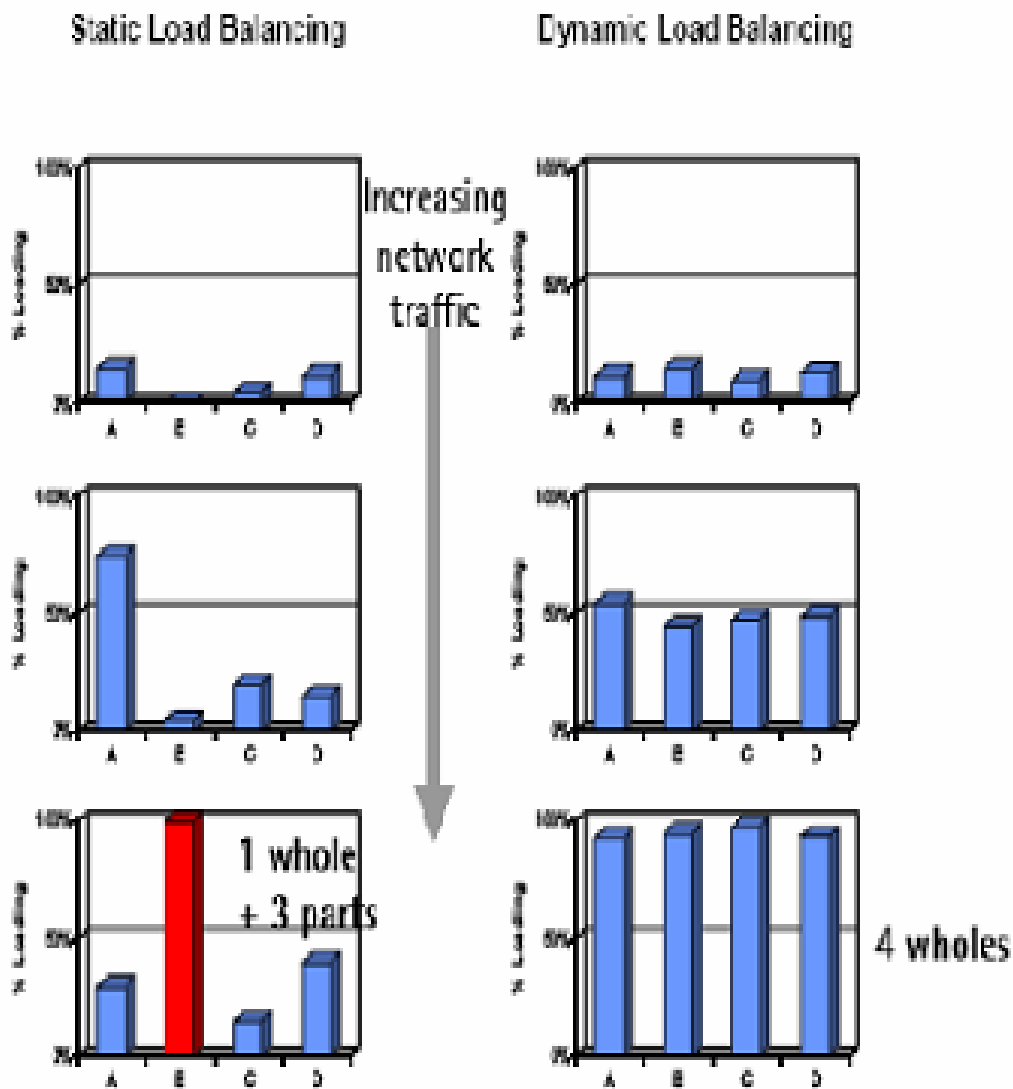


Figure 3: Static vs. Dynamic Load Balancing

Static load sharing algorithms can't anticipate how traffic will end up being distributed because the static algorithm can't adapt to the nature of the traffic. This will quickly result in workload imbalances among cluster nodes as shown above. The net effect (in the 4 node example below) is that the cluster load will scale with increasing network traffic only to "one whole plus three parts." In contrast, dynamic load balancing and rebalancing results in an equal workload distributed across the cluster nodes.

Dynamic Load Balancing Algorithm

Nokia's load balancing algorithm involves a level of indirection, in which elements of workload are assigned to "buckets," then buckets are assigned to cluster nodes. The "least-buckets" algorithm directly assigns new work to the cluster node currently assigned the least number of buckets, continuously evening out the workload among nodes

The "buckets" are actually discrete 10-bit hash values. A hash value is computed on each incoming packet using its source and destination IP address, port numbers, (in the case of VPN) security associations, etc. The hash value determines which of 1024 buckets that packet will fall in. When a bucket receives its first packet, that bucket is assigned to a cluster node based on which node currently has the least buckets already assigned. A timer attached to each bucket expires when no packets have been received after several minutes, emptying the bucket and freeing the assigned node to pick up work (a bucket) from a more heavily loaded node.

Health Checks / Active Session Failover

The health of each appliance in a cluster is constantly monitored so that a fail-over can occur in case of a problem. Each Nokia IPSO system monitors itself for signs of degradation or partial failure, and separately, a cluster-wide heartbeat mechanism detects failure of any node if it fails to send its periodic heartbeat.

The following built-in self-health-checks are continuously performed inside each cluster node, in addition to the cluster wide heartbeat mechanism:

- Examine the state of all interfaces to see that they're up, that valid addresses are assigned to them, and that interface configuration is internally consistent.
- Listen to the security and high availability daemons to ensure that they exist, are running, and are giving proper indications that they are running correctly. The daemons periodically notify the master that all is well.
- Ensure that the node is state-synchronized with the other nodes in the cluster.
- Check that the security policy is loaded.

If any of these self-health-checks fail, the node voluntarily removes itself from the cluster until and unless "health" is restored. If a node or the master within the cluster should become unavailable for any reason, the cluster automatically reassigns that node's active sessions among the remaining nodes. The time required to detect an unavailable node and rebalance the cluster load is as low 500 milliseconds, so there is little or no disruption in service after the failure of a device.

Cluster Configuration

IP clusters can be configured to work in one of three modes: multicast, multicast with IGMP or forwarding. Each mode has certain advantages and is best used in particular situations. Limitations introduced by certain Ethernet switches might require that a certain mode be configured. Two modes allow all cluster nodes to receive all traffic addressed to the cluster's IP address: multicast and multicast with IGMP.

- Multicast mode uses multicast MAC addresses at the cluster node interfaces. The Master node dynamically assigns each cluster node a subset of the overall workload. Each node then receives all traffic to the cluster (multicast to all nodes) and independently determines which packets to process and which to ignore.
- Multicast with IGMP adds IGMP snooping support to the multicast mode. Cluster members dynamically teach the switches about the network topology and to what switch interfaces to send the multicast cluster traffic.
- The third mode, forwarding mode, allows the cluster's master to receive packets at its hardware MAC address and then forward allocated traffic to the other nodes via their individual node unicast addresses.

Detailed cluster configuration information can be found on Nokia's support site.

<https://support.nokia.com/knowledge/kbdetail.jsp?ch=documents&id=06101731a6a1a7010a93dbf676007934> The following snapshots show how easy it is to configure Nokia IP Clusters and its central management capabilities.

The cluster can be managed as a single entity using Cluster Voyager. This single-node view through Voyager illustrates how each node is individually configured.

Cluster ID 10		
ID:	IP390-1	IP390-2
Model:	IP390 (Flash Based)	IP390 (Flash Based)
Software Release:	4.1-BUILD014	4.1-BUILD014
Software Version:	releng 1515 04.13.2006-194903	releng 1515 04.13.2006-194903
Serial Number:	0123456789	950609NP053
Current Time:	Mon May 22 01:55:12 2006 GMT	Mon May 22 02:53:26 2006 GMT
Uptime:	3 days 3 hours 2 minutes	3 days 3 hours
Physical Memory (Bytes):	1073741824	1073741824

Figure 4: Cluster Voyager Home Page

The “Cluster Configuration” page contains the controls for configuring and viewing the status of a node that is part of the cluster.

Clustering Configuration

Cluster Status

Cluster ID: 10

Cluster Mode: Multicast Multicast with IGMP Forwarding

Work Assignment: static dynamic

Failure Interval (Milliseconds):

Cluster Members

Member ID	Performance Rating
172.32.33.2	<input type="text" value="1100"/>
172.32.33.3	<input type="text" value="1100"/>
172.32.33.4	<input type="text" value="1100"/>

[Clustering Monitor](#)

Network Configuration

Network	State	Select	Cluster IP Address	Primary	Secondary
10.0.0.0/16	●	<input type="radio"/> Yes <input checked="" type="radio"/> No	<input type="text"/>	<input type="radio"/>	<input type="radio"/>
172.32.33.0/24	●	<input checked="" type="radio"/> Yes <input type="radio"/> No	<input type="text" value="172.32.33.1"/>	<input checked="" type="radio"/>	<input type="radio"/>
172.32.55.0/24	●	<input type="radio"/> Yes <input checked="" type="radio"/> No	<input type="text"/>	<input type="radio"/>	<input type="radio"/>
192.168.1.0/24	●	<input checked="" type="radio"/> Yes <input type="radio"/> No	<input type="text" value="192.168.1.1"/>	<input type="radio"/>	<input type="radio"/>
192.168.2.0/24	●	<input checked="" type="radio"/> Yes <input type="radio"/> No	<input type="text" value="192.168.2.1"/>	<input type="radio"/>	<input type="radio"/>

Figure 5: Cluster Configuration Page

The “Cluster Monitor” page can be used to monitor different parameters for all the cluster members at one place.

Clustering Monitor

Mon May 22 01:58:46 2006 GMT

Cluster Status

Cluster Id:	10
Cluster Mode:	mcast
Work Assignment:	dynamic
Cluster Uptime:	0 days 00 hrs 37 mins.
Number of Members:	3
Number of Interfaces:	3

Cluster Members Table

Member Id	Hostname	Platform	OS Release	Rating	Protocol Network	Time since join	Work Assigned(%)
172.32.33.2(master)	IP390-1	IP390	4.1-BUILD014	1100	172.32.33.0/24	0 days 00 hrs 37 mins.	0
172.32.33.3(member)	IP390-3	IP390	4.1-BUILD014	1100	172.32.33.0/24	0 days 00 hrs 01 mins.	0
172.32.33.4(member)	IP390-2	IP390	4.1-BUILD014	1100	172.32.33.0/24	0 days 00 hrs 11 mins.	0

[Clustering Setup Configuration](#)

Figure 6: Cluster Monitor Page

The IPSO IP Cluster is represented as a “Gateway Cluster” object in UTM-1 / VPN-1 (below, “IP390-Cluster”). The Cluster Members list the nodes that make up the IP cluster (below, “IP390-1, IP390-2 and IP390-3”). The Gateway Cluster member’s IP address is (any) one of the member’s IP addresses.

Gateway Cluster Properties - IP360-Cluster

General Properties
Cluster Members
 3rd Party Configuration
 (+) Topology
 NAT
 (+) VPN
 (+) Remote Access
 Authentication
 (+) Logs and Masters
 Capacity Optimization
 (+) Advanced

Cluster Members

Gateway Cluster members List:

Name	IP Address
<input checked="" type="checkbox"/> IP390-1	10.0.60.14
<input checked="" type="checkbox"/> IP390-2	10.0.60.15
<input checked="" type="checkbox"/> IP390-3	10.0.60.16

Add... Edit... Remove...

Figure 7: Check Point Cluster Object Properties

In the edit topology window, all the cluster members can be automatically filled by using the 'get topology' button.

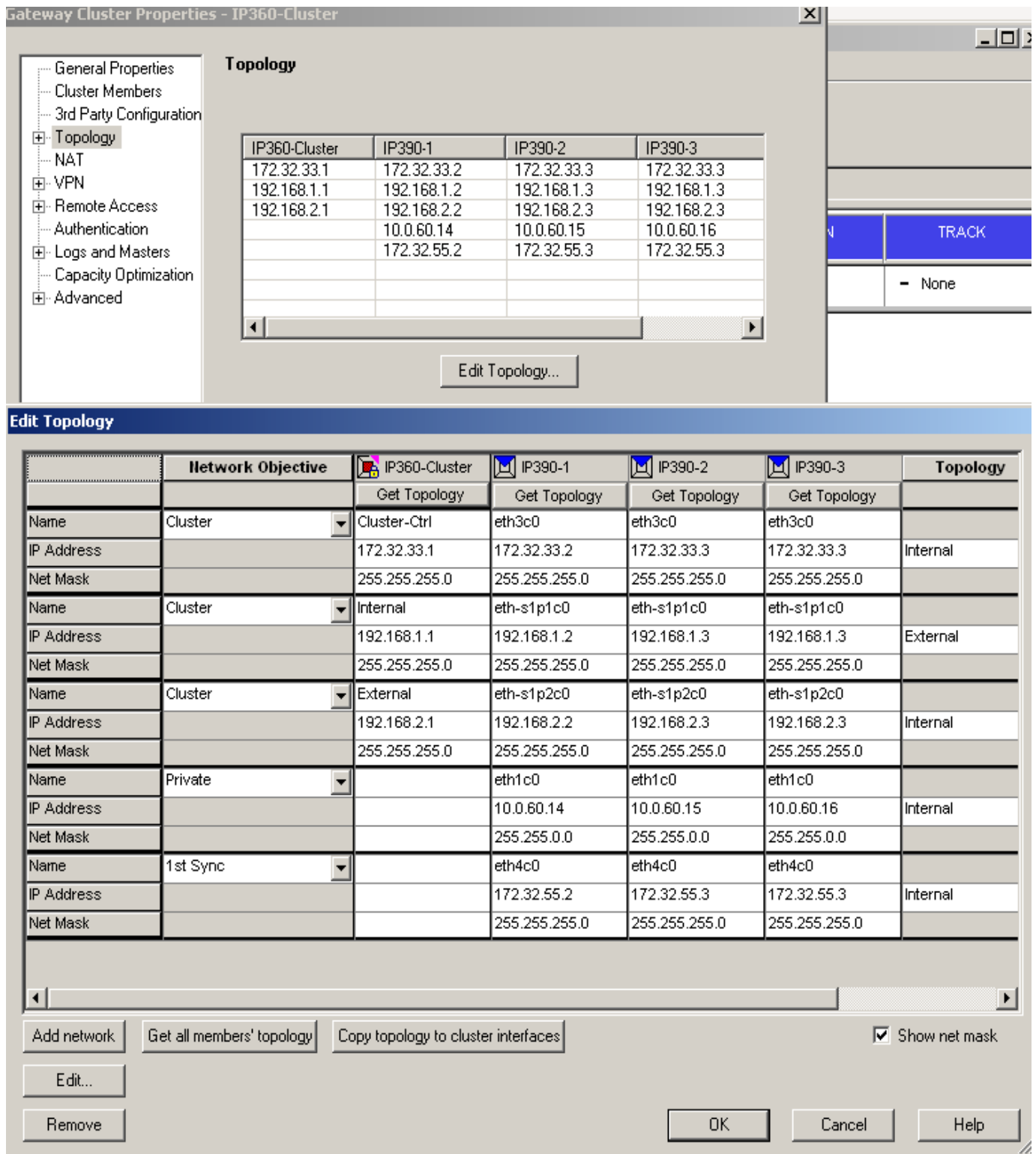


Figure 8: Check Point Cluster Object Topology Properties

Under 3rd party Configuration, cluster-operating mode needs to be set to Load Sharing (not High Availability), and Nokia IP Clustering should be selected as the 3rd party solution.

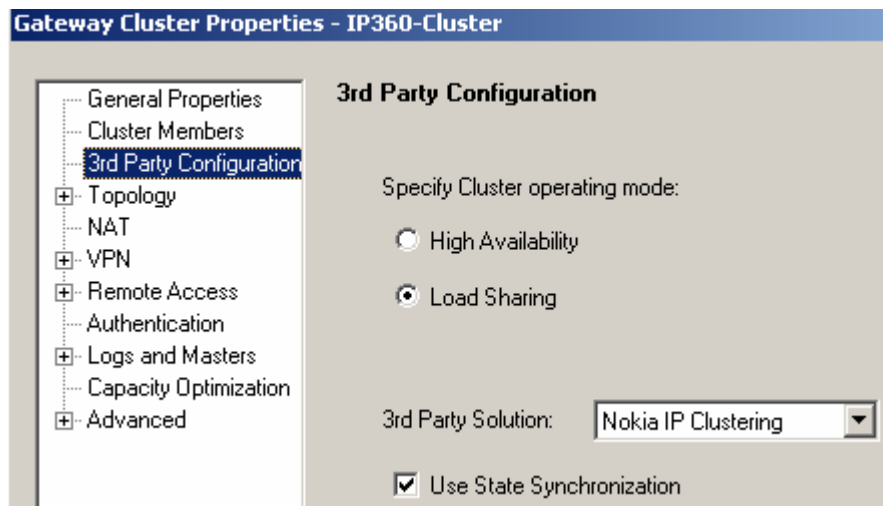


Figure 9: Selecting Nokia IP Clustering in Check Point Cluster Properties

IP Clustering Specifications: (subject to change)

Price	Free. Included as a standard feature in IPSO-3.6 and later.
Installation	Already present on the appliance as part of IPSO-3.6 and later.
Nokia IP Security Appliances Supported	Current platforms: IP2250/2255, IP1260/1220, IP740/710, IP560, IP530, IP390, IP380/350, IP265/260 & IP130.
Network Interfaces Supported	All 10/100 Mbps, Gigabit Ethernet interfaces and 10 GbE interfaces.
Check Point Applications Supported by Clustering	VPN-1 (VPN with integrated FireWall-1). SmartDefense / Web Intelligence Floodgate-1. SecureClient (without limitation).
Number of Platforms (Nodes) in a Cluster	From 2 to 4
Number of Physical Interfaces in Cluster	No limit imposed by IP Clustering (e.g. external, DMZ, multiple internal networks).
High Availability	Failover time: 0.5 seconds. Application state synchronization: Firewall connections (no dropped connections) VPN Security Associations (no dropped VPN tunnels)

Performance Scaling

	2 Nodes	3 Nodes	4 Nodes
Firewall			
UDP Throughput	2.18x	2.42x	2.51x
Transaction Rate	1.01x	1.24x	1.57x
Firewall with IPS (SmartDefense)			
UDP Throughput	1.73x	2.50x	3.27x
Transaction Rate	1.53x	2.17x	2.85x
VPN Throughput	1.9x	2.7x	3.5x

The above scalability factors are collected on an IP560 platform with a typical one-in and one-out configuration. VPN scalability is applicable for all platforms. The scalability factors will change for different platforms and configurations. It should be noted that for CPU intensive traffic or applications like encrypted traffic or deep-packet inspection by ID/P IP Clustering provides higher scaling factors.

Conclusion

In today's global economy companies have started to realize that 'access and interaction' is the name of the game for progressive businesses. There is a critical need for networking devices to maintain availability. Minutes of downtime are measured in the tens of thousands of dollars for many enterprises. To provide secure communications to customers and employees companies have to deploy new security applications. With the deployment of these technologies companies might be compromising network performance.

With the Nokia's reliable, low cost, high performing, security appliances high availability can be provided without compromising performance through IP Clustering. Nokia IP Clustering provides (N+1) availability and performance scalability unlike other solutions in the market today.

For More Information

Nokia Inc.
102 Corporate Park Drive
White Plains, NY 10604 USA
www.nokia.com
Americas
Tel: 1 877 997 9199
Email: mobile.business.na@nokia.com
Asia Pacific
Tel: +65 6588 3364
Email: mobile.business.apac@nokia.com
Europe, Middle East, and Africa
France: +33 170 708 166
UK: +44 161 601 8908
Email: mobile.business.emea@nokia.com

About Nokia

Nokia is the world leader in mobile communications, driving the growth and sustainability of the broader mobility industry. Nokia is dedicated to enhancing people's lives and productivity by providing easy-to-use and secure products like mobile phones, and solutions for imaging, games, media, mobile network operators, and businesses. Nokia is a broadly held company with listings on five major exchanges.

For more information, please visit <http://www.nokia.com/forbusiness>.