



DHCP Failover: Requirements of a High-Performance System

A white paper by Incognito Software
April, 2006

DHCP Failover: Requirements of a High-Performance System

Introduction	2
DHCP Failover Overview	2
Characteristics of Robust, High-Performance DHCP Implementations.....	3
<i>Maximum Reliability with 1:1 Failover</i>	3
<i>Efficient IP Address Space Utilization and Server Communications</i>	3
<i>Ease of Server Configuration</i>	5
<i>Integration with Other Broadband Provisioning Functions</i>	5
Comparison of System Functionality.....	5

Reliable DHCP systems are key to deploying successful VoIP services.

Introduction

Reliable DHCP systems are key to deploying successful VoIP services. VoIP requires a minimum of 5-nines, 99.999% uptime, which is why DHCP failover has to work flawlessly.

The IETF has released a DHCP Failover Protocol draft document, although it has not yet reached RFC status.

The IETF draft represents a set of minimum requirements for DHCP failover. For maximum performance, broadband providers need to go beyond the draft specifications. Implementations of DHCP failover vary widely, and even subtle differences can have a huge impact on reliability.

This paper discusses a high-performance implementation of DHCP failover, especially critical for broadband environments.

DHCP Failover Overview

DHCP failover involves both a primary DHCP server and a secondary backup DHCP server, which takes over DHCP duties should the primary fail.

During “normal operation,” only the primary server performs IP address assignment while the secondary server is in standby mode. The primary server issues IP address leases, but the secondary does not. The primary server constantly transfers IP address data to the secondary server to ensure data synchronization. Also, the two servers regularly poll each other, using stay-alive “heartbeat” messages, to determine their status. An interruption in this heartbeat is deemed a failure situation.

The secondary server must react immediately and assume the role of assigning IP addresses.

If the primary server shuts down or stops communicating with the secondary server, failover mode is initiated. The secondary server must react immediately and assume the role of assigning IP addresses. The secondary server must also attempt to re-establish communications with the primary. When the primary comes back

online, it should resume its role as primary after being updated by the secondary with the most recent IP address data.

If the primary server loses communications with the secondary server, the primary continues to allocate IP addresses while trying to re-establish communications with the secondary.

Characteristics of Robust, High-Performance DHCP Implementations

As a critical part of a broadband service infrastructure, DHCP failover implementations have to meet broadband needs for robustness and efficiency.

As a critical part of a broadband service infrastructure, DHCP failover implementations have to meet broadband needs for robustness and efficiency. There are several key features that differentiate a high-performance DHCP failover system: a 1:1 relationship between the primary and secondary servers, efficient use of IP address space and server communications, ease of server configuration, and integration with other device provisioning functions.

Maximum Reliability with 1:1 Failover

The IETF draft specifies n:1 failover and load-balancing, which involves one secondary as backup for multiple primary servers. The limitation with this approach is that the failure of multiple primary servers may have a wide-reaching, catastrophic impact on the entire network.

A higher-performance system dedicates one secondary for each primary, a 1:1 failover relationship, where the secondary is an exact replica of the primary, including configuration information. This scheme is a "true hot standby" system, which ensures that no performance penalties occur during failover.

Efficient IP Address Space Utilization and Server Communications

Part of the IETF draft discusses possible solutions to "two challenging scenarios" that test the robustness of a DHCP failover system. However, these solutions significantly reduce a network's available IP address space due to the reservation of large blocks of IP addresses and MCLT restrictions, and also slow down communications.

While the IETF draft ensures there will be no contention between servers, there is a large price to pay – up to 50% of the IP addresses may have to be reserved specifically for the secondary server alone, where they sit idle most of the time, not available for broadband service deployments.

A high-performance scheme uses 100% of the available IP address space for broadband service deployments.

An alternative, higher-performance scheme is possible – it uses 100% of available IP addresses for broadband service deployments, offers "high-water mark" alerts to avoid IP address depletion, and leverages a faster communications channel for synchronization.

A more efficient approach is to have both servers use the same address pools, but during failover mode, the secondary server processes addresses in an order opposite to that of the primary.

- Primary-Secondary Server Contention: *“IETF Challenging Scenario#1 – DHCP servers can't communicate to each other and both are allocating IP addresses to clients.”*

The primary and secondary servers may be functioning correctly but cannot communicate with each other due to an unrelated network failure. In this situation, both servers attempt to allocate the same IP addresses, a contention issue.

To avoid this scenario, the IETF Draft suggests using “disjoint IP address pools,” which essentially cuts the number of available IP addresses in half by devoting specific address ranges to the secondary server. The IETF approach requires network operators to develop critical assumptions about the arrival rate of new DHCP clients during each failover period and the mean-time-to-restoration for the primary server. This determines the amount of IP address space to be reserved for the secondary server. Underestimating these parameters can result in the primary being offline and the secondary being unable to issue leases because its address pool is exhausted. Overestimating these parameters can create a persistent and very wasteful reservation of routable IP address space.

A more efficient approach is to have both servers use the same address pools, but during failover mode, the secondary server processes addresses in an order opposite to that of the primary. In other words, the secondary always allocates the highest IP address possible (starting at the top of the range and working its way down to the bottom) and the primary always allocates the lowest IP address possible (bottom to top). This helps to ensure that the two servers do not allocate duplicate IP addresses. Every active IP address lease should also indicate which server the lease was allocated from.

As well, if customizable high-water marks are available in the DHCP system, they can help prevent the network from running out of IP address space. When a user-specified threshold is reached, the system can trigger an emergency notification to network administrators via e-mail, SNMP traps, or online ‘pop-up’ messages.

- Primary-Secondary Server Synchronization: *“IETF Challenging Scenario#2 – primary server crashes before secondary is updated, and the requesting DHCP client goes offline and won't be discovered by a ping.”*

The primary may crash before it can update the secondary with the most recent changes, resulting in a lack of data synchronization between servers.

The approach presented in the IETF Draft has two performance-related issues. The IETF draft suggests a single TCP connection for both data transfers and “keep-alive” heartbeat messages between primary and secondary servers, which can clog communications and cause false startup of DHCP failover. Also, the IETF uses a delay mechanism, called Maximum Client Lead

To ensure the most up-to-date data synchronization between servers, there is a fast, efficient, low-overhead method that uses a less complex transport layer protocol, UDP, and a data format that has virtually zero overhead.

Time (MCLT), which is the length of time an IP address lease may be renewed by either server without contacting the other. The longer this time is, the longer it will take the functioning server to recover IP addresses. The IETF approach reserves IP addresses for a prolonged period of time, thereby reducing the available IP address pool size.

To ensure the most up-to-date data synchronization between servers, there is a faster, more efficient, lower-overhead method that uses a separate, less complex transport layer protocol, UDP, and a data format that has virtually zero overhead. UDP is also used for heartbeat messages but in a different thread to eliminate any blockages. (During initial synchronization of the servers, the primary server sends configuration information and IP address lease data to the secondary using a TCP connection.)

Ease of Server Configuration

A DHCP failover system should simplify server configuration with a GUI and built-in rules to prevent errors.

To keep network operations efficient and productive, a DHCP failover system should offer an easy-to-use interface. It should simplify server configuration with a GUI and built-in rules to prevent configuration errors. It also needs to combine monitoring, reporting, and emergency alert features into a single user interface. Over 200 statistics may be monitored for failover operation including event times, errors, number of messages sent/received, and status.

In contrast, the IETF document describes manual configuration of the secondary server using a basic command line interface.

To achieve the fastest return on investment, there needs to be tight integration between the DHCP failover system and additional services such as TFTP and DNS.

Integration with Other Broadband Provisioning Functions

To achieve the fastest return on investment, there needs to be tight integration between the DHCP failover system and additional services such as TFTP, DNS, and ToD for end-to-end support of a wide variety of broadband provisioning requirements.

Comparison of System Functionality

The following table summarizes some differences in functionality between Incognito Software's DHCP failover system and other implementations.

Feature	Incognito Software	Other Implementations
Initial Setup	<p>2 Steps with a GUI: (1) Enter the IP address of the secondary server, and (2) Click the "Initiate Failover" button. All further configuration and synchronization is automatic.</p> <p>A command line interface is also available.</p>	<p>10 steps with a command line interface: (1) Create a copy of the primary server's configuration file, (2) Transfer the file to the secondary server, (3) Import the file into the secondary, (4) Confirm the primary and secondary have identical configurations, (5) Compare text output, looking for discrepancies between the primary and secondary configurations, (6) If there are errors, resolve them and return to Step 1, or (7) Save the new configuration, (8) Reload the secondary server, (9) Wait several minutes, and (10) Verify that the primary and secondary servers are synchronized.</p>
Failover Monitoring	<p>(1) GUI tree view, (2) emails to administrators, (3) SNMP traps, (4) notification to currently logged-in consoles, or (5) the command line interface.</p>	<p>(1) GUI menu options, (2) SNMP traps, or (3) the command line interface.</p>
Utilization of IP Address Space	<p>100% utilization of IP address space is possible by using bottom-up allocation for the primary and top-down allocation for the secondary. Two levels of high-water marks are set on both servers to monitor address pool levels and avoid conflicts.</p>	<p>Large blocks of IP address space are reserved only for the secondary server, where they go unused unless the secondary is issuing leases. MCLT reserve addresses for a particular period of time after their leases have expired.</p>
Failover Communications	<p>TCP and multiple UDP threads for data synchronization and heartbeat information with virtually no overhead and no bottlenecks.</p>	<p>A single TCP connection for all data synchronization and heartbeat information, which can lead to delays and false failover.</p>
Changes to the DHCP service configuration on the secondary server	<p>Changes are permitted to the DHCP service configuration on the secondary server during failover or during normal operation. When the primary is ready and/or restored, all changes are automatically sent to the primary during synchronization.</p>	<p>Not supported.</p>

Contact:

Incognito Software Inc.
 Phone: 604.688.4332 or US/Canada toll free 800.877.1856
 Fax: 604.688.4339
 Email: sales@incognito.com
 Web: <http://www.incognito.com>