広域 IP ネットワーク用 DHCP システムの信頼性向上方式

内藤克浩[†]　　西出誠[††]　　宮副英治[†††]

Katsuhiro NAITO[†], Makoto NISHIDE[††], and Eiji MIYAZOE[†††]

あらまし　DHCP (Dynamic Host Configuration Protocol) は商用 ISP (Internet Service Provider) において，適切なクライアントにネットワーク情報を通知する上で必要不可欠なシステムである．一方，無償あるいは低コストで入手可能は DHCP サーバープログラムは，商用 ISP などの広域 IP ネットワークでの実用に耐える信頼性と規模拡張性を有するものは少ない．本論文では，無償の DHCP サーバープログラムである ISC-DHCP の特性を明らかにする．そして，ISC-DHCP の基盤環境としてクラスタリング技術である DRBD (Duplicated Replicated Block Device) 及び Pacemaker を利用することにより，広域 IP ネットワークでの実用に耐える信頼性と規模拡張性を有する DHCP システムである「CREID」の提案を行う．実証実験より，提案 DHCP システムは商用 ISP などで要求される規模拡張性を有するとともに，商用運用で要求される信頼性も実現可能であることを示す．

キーワード　DHCP, 広域 IP ネットワーク, ISC-DHCP, DRBD, Pacemaker

1. はじめに

近年のネットワークの普及に伴い，DHCP (Dynamic Host Configuration Protocol) [1], [2] を用いたネットワーク設定の自動化は重要な機能となりつつある．特に，端末起動時に ISP (Internet Service Provider) 内にある管理システムとの通信を強制し，様々な設定情報を安全かつ確実に配布する DHCP の機能は，広域 IP ネットワークを運用する商用 ISP では必要不可欠なものである [3]〜[6].

広域 IP ネットワークの規模は飛躍低に拡大しており，特に商用 ISP では信頼性と規模拡張性の両立が必要不可欠となっている．結果として，商用 ISP 市場を対象としたハイエンド DHCP 製品では，信頼性及び規模拡張性が十分に改善されてきた [7]．一方で，ハイエンド DHCP 製品の淘汰と寡占化が進んだことにより，製品価格は高額化しており，特に設備投資資金が不足気味の中小 ISP では，製品導入コストが大きな負担になりつつある．

一般に DHCP システムの構築コストと信頼性はトレードオフの関係となる．そのため，商用のハイエンド DHCP 製品を利用した場合，高信頼かつ規模拡張性が高い DHCP システムの構築が可能であるが，設備投資コストの増大が問題となる．一方，無償またはローエンド DHCP 製品を利用した場合，商用 ISP で必要とされる信頼性及び規模拡張性を実現することが困難な状況である．特に，無償またはローエンド DHCP 製品における問題は，多数の端末が一斉に DHCP リクエスト処理を実行した際の DHCP サーバーの対応能力不足，DHCP サーバー機能のフェイルオーバー機能の不完全さによる信頼性の低下が問題となる [8].

本論文では，無償の DHCP ソフトウェアである ISC-DHCP [9] の信頼性と規模拡張性を補う手段として，クラスタリング技術である DRBD (Duplicated Replicated Block Device) [10] 及び Pacemaker [11] を用いる CREID システム [12] の提案を行う．ISC-DHCP は

[†] 三重大学 大学院工学研究科
Graduate school of Engineering, Mie University
[††] 株式会社ネットステップ
Net Step Inc.
[†††] オーエスエスブロードネット株式会社
OSS BroadNet Inc.

表 1　基礎測定でのハードウェア諸元

	DHCP Server	DHCP Client
CPU	Intel Xeon(R) E5620 2.40GHz	Intel Celeron G1101 2.26GHz
Memory	8GB	1GB
HDD	300GB 6G SAS 15000rpm	250GB SATA2 7200rpm
Network	1Gbps	
OS	Scientific Linux 6.1	
DHCP	ICS-DHCP 4.1-ESV-R2	

無償の DHCP ソフトウェアの事実上の標準ソフトウェアであり，商用 ISP でも必要となる IPv4 及び IPv6 のアドレス貸出が実現可能である．また，DRBD 及び Pacemaker は UNIX 用のクラスタリング実現ソフトウェアであり，複数の PC を用いたクラスタリングが実現可能である．

CREID システムは無償で公開されているソフトウェア群の上に構築されているにも関わらず，信頼性と規模拡張性が高い DHCP システムを実現可能である．結果として，設備投資の資金不足から信頼性の向上を断念せざるを得なかった中小 ISP 事業者が，大手 ISP 事業者と同水準の DHCP サービスをエンドユーザーに提供できるようになると考えられる．また，ISP 事業者の規模や資本力によらず，高信頼かつ規模拡張性が高い DHCP サービスの提供が可能となれば，広域 IP ネットワークの地域格差が是正され，ネットワークインフラが更に安定するものと考えられる．

2. ISC-DHCP の基礎性能

本章では，CREID で利用する ISC-DHCP の基礎特性を明らかにし，CREID システムを設計する上で重要となる点について考察を行う．本評価では，表 1 に示すハードウェアを利用することで，DHCPD を 1 プロセス実行した際の特性を評価した．また，各ハードウェアはクロスケーブルを利用して直接接続した．

2.1 IP プールレンジの総 IP アドレス数の影響

DHCP サーバーでは，貸出を行う IP アドレスを予め IP プールレンジに登録する．また，IP プールレンジが増加するほど，IP アドレスの貸出処理時間が増える傾向がある．本評価では，DHCP サーバーの負荷測定で利用される dhcperf [13] を利用し，Discover トランザクションの 5 シーケンス (Discover，Offer Request，Ack，Release) 及び，Renew トランザクションの 2 シーケンス (Request，Ack) を連続で実行し，それぞれの最大処理速度を計測した．

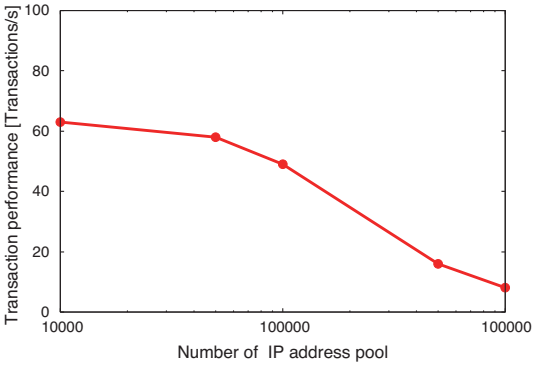

図 1　IP プールレンジの総 IP アドレス数に対する DHCP トランザクション性能.

図 1 は IP プールレンジの総 IP アドレス数に対する DHCP トランザクション性能を示す．結果より，DHCP のトランザクション性能は IP プールレンジの総 IP アドレス数の増加に伴い急激に劣化することが確認できる．また，ISC-DHCP を用いて広域 IP ネットワーク用 DHCP システムを構築する場合，IP プールレンジ数を適切な数に設定することが重要であることが確認できる．

2.2 登録 MAC アドレス数の影響

プロバイダーなどでは，加入者の MAC アドレス情報を予め DHCP システムに登録する運用を行うことがある．また，ISC-DHCP では，プロセス起動時に様々な設定情報を読み込むため，登録 MAC アドレス情報の増加に伴い，設定情報の読み込み時間も増加することが予想される．本評価では，プール IP アドレス数を 100,000 で固定して測定を行った．

図 2 は登録 MAC アドレス数に対するプロセス起動時間を示す．結果より，登録 MAC アドレス数に応じて DHCP のプロセス起動時間も増加することが確認できる．しかし，プロセス起動時間は数秒以内と短時間のため，DHCP クライアントのリトライ機能がある場合には大きな問題とはならないと考えられる．

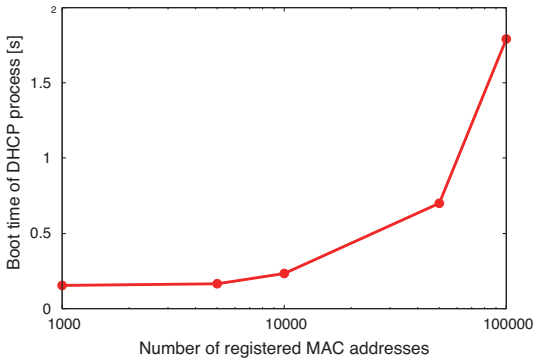

図 2　MAC アドレス数に対するプロセス起動時間.

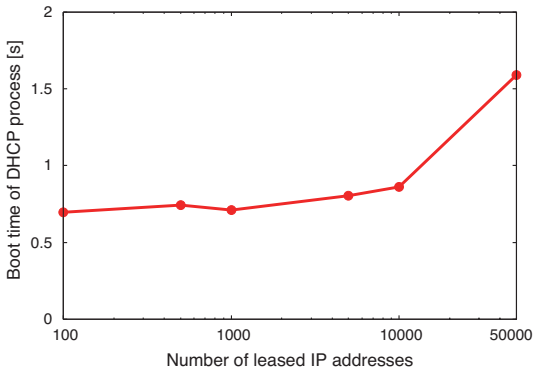

図 3　貸出 IP アドレス数に対するプロセス起動時間.

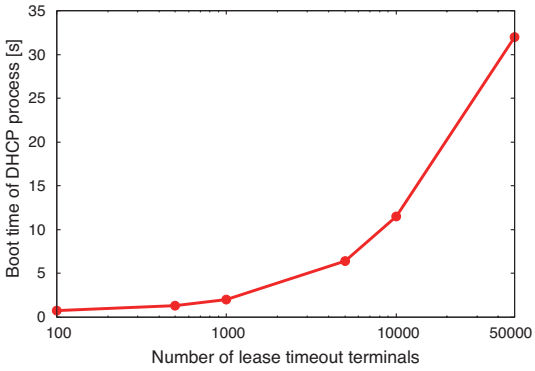

図 4　貸出時間が過ぎた IP アドレス数に対するプロセス起動時間.

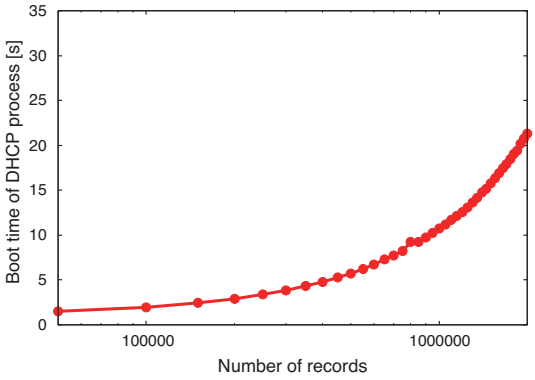

図 5　更新済みの貸出 IP アドレス数に対するプロセス起動時間.

2.3　リースファイルの最適化の影響

ISC-DHCP では端末への IP アドレスの貸出状況をリースファイルとして保存している．そのため，貸出端末数の増加に伴い，リースファイルの記録内容も増加する．結果として，プロセス起動時のリースファイルの読み込みに伴う遅延が予想される．本評価では，貸出 IP アドレスのレコード数を変化させた場合，貸出当 IP アドレスの貸出時間が過ぎたレコード数を変化させた場合，DISCOVER を繰り返したことにより，更新された貸出 IP アドレスのレコード数が変化した場合について検証を行った．本評価では，プール IP アドレス数を 100,000，登録 MAC アドレスを 50,000 で固定して測定を行った．

図 3 は貸出 IP アドレスのレコード数に対するプロセス起動時間を示す．結果より，貸出 IP アドレス数の増加に伴い，DHCP のプロセス起動時間も増加することが確認できる．しかし，プロセス起動時間は数秒以内と比較的短時間であり，DHCP クライアントのリトライ機能がある場合には大きな問題とはならないと考えられる．

図 4 は貸出 IP アドレスの貸出時間が過ぎたレコード数に対するプロセス起動時間を示す．結果より，貸出時間が過ぎたレコードがリースファイルに含まれている場合，DHCP のプロセス起動時間は大幅に長期化することが確認できる．そのため，ISC-DHCP を 1 プロセスで実行する場合には，定期的な DHCP プロセスの再起動を通して，リースファイル内の貸出時間が過ぎたレコード数を一定量以下に抑える運用が重要と考えられる．

図 5 は更新された貸出 IP アドレスのレコード数に対するプロセス起動時間を示す．結果より，更新されたレコード数がリースファイルに含まれている場合，

DHCP のプロセス起動時間は大幅に長期化することが確認できる．そのため，図 4 と同様に，定期的な DHCP プロセスの再起動を通して，リースファイル内の貸出時間が過ぎたレコード数を一定値以下に抑える運用が重要と考えられる．

2.4 アプリケーションフェイルオーバー

IETF (Internet Engineering Task Force) によりドラフト化されている DHCP Failover Protocol [8] では，フェイルオーバーとロードバランスを同時に設計・定義しており，ISC-DHCP も本機能の実装を試みている [9]．しかし，同一レンジを共有する場合のロードバランス制御及び MCLT (Max Client Lease Time) による短間隔での DHCP 更新時における各 peer の自律的な状態判断の複雑さから，本機能の不安定な動作やバグの報告が後を絶たない状況である [15], [16]．

実際の広域 IP ネットワークでは，対象ネットワークを一定規模に分割する設計が一般的である．また，DHCP に対するロードバランシングの需要は比較的低く，本ドラフトが改訂される予定は今のところない．上記の通り，DHCP のフェイルオーバー機能は重要である一方で，現在利用可能な方式は商用 ISP において要求される信頼性の確保には不十分な状況である．

3. 広域 IP ネットワーク用 DHCP システムに求められる性能

3.1 広域 IP ネットワークの具体例

広域 IP ネットワークの具体例として，DOCSIS を利用したケーブル ISP 向け DHCP システムの信頼性要件について検討する．DOCSIS では，CM (Cable Modem) 及び CPE (Customer Premises Equipment) により DHCP システムへの信頼性要件が異なる．CM は起動時に DHCP Discover を実行することにより IP アドレスを DHCP サーバーから取得する．また，この際に DHCP オプションを利用することにより TFTP サーバーの IP アドレスと設定ファイル名を取得する．次に，CM は TFTP を利用した設定ファイルのダウンロード動作に移る．DHCP を用いた IP アドレス及び設定情報の取得と TFTP を用いた設定ファイルの取得は CM 単位で連続して実行されることで，CM のオンライン化によりトランザクションが完了する．このため，CM の一斉リブート時の再起動所要時間などの DHCP システムのリクエスト処理能力は，DHCP 及び TFTP のいずれか遅い方の処理性能が基準となり，相互作用により更なる能力低下が生じる仕組みである．そのため，DHCP 及び TFTP を分離，独立して取り扱うことは，商用のケーブル ISP での運用を想定した場合は適切ではない．

また，CM は起動からリース期間経過後，DHCP Renew を実行することによりリース期間を更新する．この際，CM に貸出される IP アドレスが変化した場合，DOCSIS では CM のリブートが発生するとともに，利用中のアプリケーションサービスが中断される．このため，DHCP サーバーには，フェイルオーバー時にも同一 CM には同一 IP アドレスの貸出を保証する，リース情報の動的な共有手段が求められる．

一方，CPE は TFTP 動作が不要のため，DHCP システムの処理能力は DHCP 単独の性能により決定される．また，CPE も起動からリース期間経過後，DHCP Renew を実行することによりリース期間を更新するが，リース期間の更新時に IP アドレスが変化した場合でも，CPE がリブートを行うことはない．そのため，CPE を対象とした DHCP サーバーでは動的なフェイルオーバーの仕組みは必ずしも必須ではなく，リース情報を動的に共有しないフェイルオーバー構成を利用可能である．

3.2 性能要件

ケーブル ISP の場合，CM 一斉リブートは，50,000CM を収容する CMTS の障害・復旧ないしは保守目的のリセットによるものが最大単位である．CM のリブート時は一時的なサービス中断が発生するため，可能であれば 1 分以内，遅くとも 5 分以内にリブートが完了する事が望ましい．この場合の所要性能は，DHCP Discover でそれぞれ 50,000/60=830 トランザクション/秒，50,000/60/5=167 トランザクション/秒となる．

3.3 可用性

DHCP システムを構成する DHCPD は，設定の変更時にシステム情報の最適化などに伴う処理遅延が発生する．CM が DHCP システムから IP アドレスを

獲得可能となるのは，DHCPD の設定の変更後となるため，DHCP サービスの可用性は重要である．

DHCP に関わる設定の変更に，DHCPD の再起動が必要な DHCP システムの場合，CATV ネットワークの運用において，DHCPD の再起動が発生するのは大きく以下の状況に大別できる．

- 日常業務での更新処理

CATV ネットワークでは，日常業務での DHCP サービスの設定更新などを頻繁に行っている．例えば，特定 CM の MAC アドレスに対する DHCP サービスのポリシー変更などが日常適時実施される．ここで，日常業務による 1 日あたりの更新頻度を P_{short} とし，再起動に必要となる秒数を T_{short} とする．

- メンテナンスでの更新処理

CATV ネットワークのメンテナンスに伴い，サブネット，IP レンジ，サブクラス，ポリシーなどの追加又は更新を行う必要がある．ここで，1 日あたりの更新頻度を P_{long} とし，再起動に必要となる秒数を T_{long} とする．

- システム障害

現用系サーバーで障害が発生した場合，予備系サーバーに切り替えを行う必要がある．ここで，1 日あたりの現用系サーバーの障害発生頻度を P_{fail} とし，予備系サーバーに切り替えるために必要となる秒数を T_{fail} とする．

- リースファイル最適化処理

ISC-DHCP では，DHCPD の再起動時にリースファイルの最適化を行う．この最適化処理は，リースファイルサイズの増大に伴って長期化するため，定期的な DHCPD の再起動はサービスの長期中断を防ぐ上で重要である．ここで，1 日あたりのリースファイルの最適化頻度を P_{opt} とし，再起動に必要となる秒数を T_{opt} とする．

DHCP クライアントに貸出されていた IP アドレスのリース時間が切れた場合，DHCP クライアントは DHCP サーバーに対して，リース時間の延長要求を行う．多くの DHCP クライアントでは，DHCP サーバーからの返信がない場合には，再送処理を行う．また，DHCP サーバーが利用不能と判断した時点で，利用していた IP アドレスの解放を行う．ここで，リース更新の際に，DHCP サーバーが利用不能と判断するまでの時間を $T_{timeout}$ とすると，上記の DHCPD の再起動に伴う DHCP サービスの停止時間は $T_{timeout}$ 短くなる．また，更新処理及びシステム障害時の DHCPD の再起動では，リースファイルの最適化も同時に行われる．そこで，各項目について，リースファイルの最適化時間 T_{opt} を含めた再起動に必要となる秒数を，$T'_{short}, T'_{long}, T'_{fail}$ とする．また，各項目の CM のタイムアウト時間を考慮した実質的なサービス中断時間は以下の通りとなる．

$$T''_{short} = \begin{cases} 0 & (T'_{short} \leqq T_{timeout}) \\ T'_{short} - T_{timeout} & (T'_{short} > T_{timeout}) \end{cases} \tag{1}$$

$$T''_{long} = \begin{cases} 0 & (T'_{long} \leqq T_{timeout}) \\ T'_{long} - T_{timeout} & (T'_{long} > T_{timeout}) \end{cases} \tag{2}$$

$$T''_{fail} = \begin{cases} 0 & (T'_{fail} \leqq T_{timeout}) \\ T'_{fail} - T_{timeout} & (T'_{fail} > T_{timeout}) \end{cases} \tag{3}$$

$$T''_{opt} = \begin{cases} 0 & (T_{opt} \leqq T_{timeout}) \\ T_{opt} - T_{timeout} & (T_{opt} > T_{timeout}) \end{cases} \tag{4}$$

また，上記理由に伴う，DHCP サービスの 1 日あたりのサービス停止秒数 T_{down} は以下の式で表せる．

$$\begin{aligned} T_{down} \ = \ & P_{short}T''_{short} + P_{long}T''_{long} \\ & + P_{fail}T''_{fail} + P_{opt}T''_{opt} \end{aligned} \tag{5}$$

DHCP サービスの運用では，CM はリース更新時間毎に獲得しているアドレスの更新処理を行う．各 CM は相関なく起動する場合を考えると，DHCP サーバーが管理する CM 数を N_{cm}，リース時間を T_{lease} 秒とした場合，1 秒あたりのリース更新を実施する CM 数 N_{lease} は以下の式で表せる．

$$N_{lease} = N_{cm}/T_{lease} \tag{6}$$

次に，CM のサービス停止につながる DHCP のサービス停止期間中にリース更新を実施する CM 数 N_{down} は以下の式で表せる．

$$N_{down} = N_{lease}T_{down} \tag{7}$$

また，CM1 台あたりのサービス可用性に着目すると，可用率は以下の式で表される．

$$P_{available} = 1 - N_{down}/N_{cm} \tag{8}$$

4. CREID の提案

4.1 CREID の概要

CATV ネットワーク向けの商用 DHCP 製品として,CNR(Cisco Network Registrar) が市場に普及している．CNR は Cisco Systems 社が開発・販売するソフトウェア製品であり,DHCP・TFTP・DNS サーバー・CLI コマンドインタフェースの nrcmd および, その他のツール類により構成される．上位の業務系アプリケーションは nrcmd を介して, クライアントの登録・設定変更・削除や各サーバーの設定・制御等の各種操作を行う．一方で CREID は, DHCP が ISC-DHCP,TFTP が Advanced TFTP, DNS が BIND と，無償オープンソースソフトウェアにより構成されるコストパフォーマンスに優れたソフトウェア製品であり, 付属の nrcmd エミュレータが CNR 互換の CLI コマンドインタフェースを提供する．CREID の nrcmd エミュレータにより, 業務系アプリケーションの既存プログラムコードに改修・変更なく, DHCP 等の基幹サーバー群を CNR から CREID に容易に置換できる．CREID クラスターは DRBD と Pacemaker によるウォームスタンバイのフェイルオーバーが設定された 2 台の Linux PC サーバーにより構成され,1 クラスターで最大 100,000 の DHCP クライアントを収容する．CATV ネットワークの場合,CREID の収容対象となる DHCP クライアントの種類は,CM, EMTA(Embedded Media Terminal Adapter), STB(Set Top Box) 及び CPE(Customer Premises Equipment) すなわちパソコンなど加入者宅内の情報端末である．CREID の主要機能は,DHCP や TFTP など,CATV ネットワーク以外の分野でもデファクトの機能・プロトコルにより構成されているため, 潜在的な応用分野は CATV に限らず幅広い.

4.2 CREID のシステム構成と仕様

図 6 に CREID のシステムモデルを示す．図 6 は CATV ネットワークを想定した例であり，CMTS (Cable Modem Termination System) 配下の CM に対して DHCP の機能を用いた IP アドレスの貸出と TFTP サーバーなどの補足情報の通知を行う．また，CATV ネットワークでは CM の MAC アドレスを用いて不正な CM への IP アドレス貸出を回避している．そのため，管理サーバーを利用して DHCP サーバー用の MAC アドレスリストの管理を行う.

CREID の大きな特徴は，DHCP プロセスへの貸出 IP アドレス数の増加によるオーバーヘッドを削減するため，複数の DHCP プロセスを用いて DHCP の機能を提供する点である．そのため，各 DHCP プロセスは担当する MAC リストを所持するものとする．また，DHCP サーバーのフェイルオーバー対策として，DRBD 及び Pacemaker を利用したクラスタリングシステム上に DHCP サーバーを構築する．そのため，現用系サーバーに加えて予備系サーバーを準備し，両サーバーは専用のネットワークを利用して直接同期が可能な設計を行う．結果として，現用系サーバーの障害発生時には，現用系サーバーの MAC リスト及び IP アドレスのリース状況の情報を用いて予備系サーバーが稼働可能である.

5. 数 値 例

5.1 評 価 環 境

提案システムである CREID の基礎特性を評価する

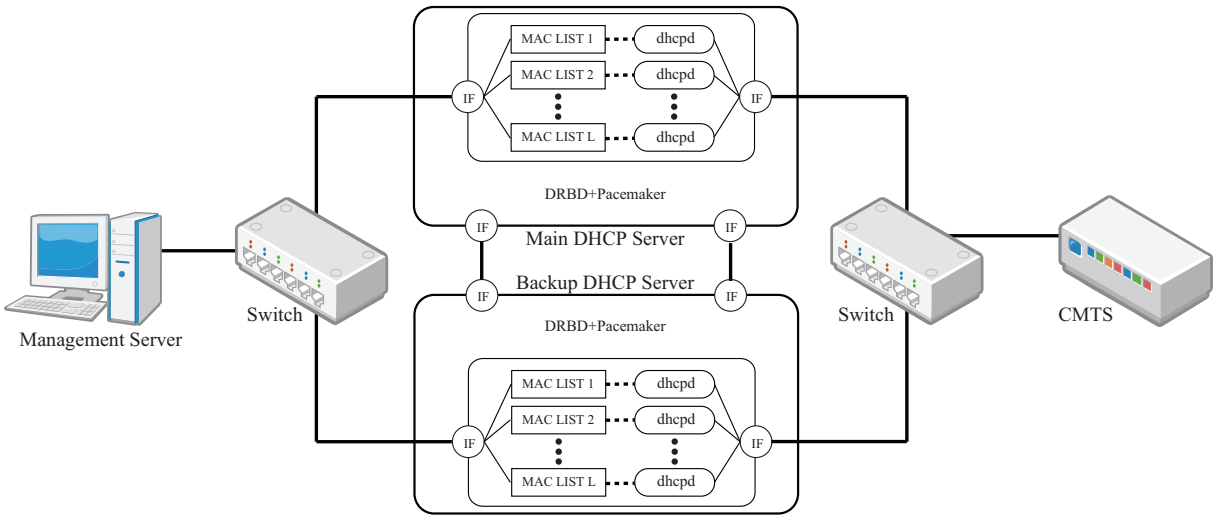

図 6　CREID のシステムモデル.

表 2　CREID システムの測定ハードウェア諸元

	DHCP Server	DHCP Client
CPU	Intel Xeon(R) E5620 2.40GHz	Intel Celeron G1101 2.26GHz
Memory	8GB	1GB
HDD	300GB 6G SAS 15000rpm	250GB SATA2 7200rpm
Network	1Gbps	
OS	Scientific Linux 6.1	
DHCP	ICS-DHCP 4.1-ESV-R2	
Pacemaker	1.0.11	
Heartbeat	3.0.5	
DRBD	8.4.1	

ため，図 6 に示すモデルを表 2 に示す装置を利用して構築した．構築システムでは，各機器間をクロスケーブルを用いて直接接続を行うことで，スイッチングハブなどによる特性の影響を排除した．

5.2　クラスタリング性能

CREID ではサーバーのフェイルオーバー対策として，DRBD 及び Pacemaker を利用したクラスタリングシステム上に DHCP サーバーを構築している．一般にクラスタリングシステム上にシステムを構築した場合，クラスタリング層の処理が増加することによるオーバーヘッドが発生する．特に，CREID では DHCP サーバーによる IP アドレス貸出時にディスクアクセスが発生するため，IO 性能は重要となる．そこで，クラスタリングシステム間のファイル書き込み性能の実装を実施した．

測定では，同期オプション付で DRBD 同期ディレクトリをマウントし，プライマリサーバにおいて 10MB ファイル 60 個を DRBD 同期ディレクトリにコピーを行った際の経過時間を測定した．

結果では，DRBD を利用しない場合のファイル入出力性能は 56.34Mbps であるのに対し，DRBD を利用した場合のファイル入出力性能は 3.81Mbps であった．DRBD を利用することによるスループットの低下は大きいが，DHCP プロセスが処理するリースファイル及び MAC リストファイルには，テキスト情報しか含まれておらず，DHCP プロセスのファイル入出力には十分なファイル入出力性能と考えられる．

5.3　並列 DHCP プロセスのトランザクション性能

ISC-DHCP では複数の DHCP プロセスを起動することによるトランザクションの並列処理が実現可能である．本評価では，DHCP プロセス数に応じた環境を 8 種類 (1，10，20，25，40，50，80，100) 用意し，全 DHCP プロセスの登録クライアント数合計が常に 100,000IP アドレスとなるように登録した．一方でプール IP レンジは，全 DHCP プロセスの合計が常に 200,000IP アドレス分の論理空間となるように登録した．

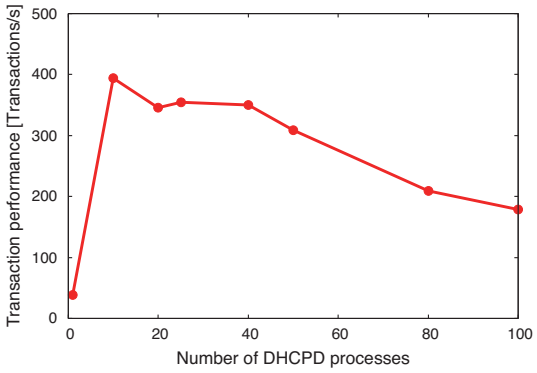

図 7　DHCP プロセス数に対する DHCP Discover トランザクション性能.

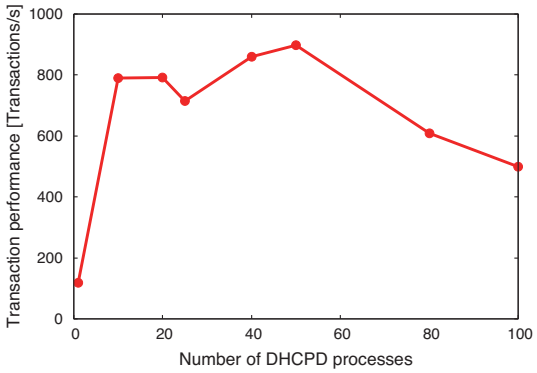

図 8　DHCP プロセス数に対する DHCP Renew トランザクション性能.

図 7 は DHCP プロセス数に対する Discover トランザクション性能を示す．図 7 より，ISC-DHCP の Discover トランザクション性能は並列プロセスを実行することにより大幅に改善可能であることが確認できる．なお，測定に用いたハードウェア環境では，最大並列プロセス数は 50 までが有効であり，その後は並列化によるオーバーヘッドにより特性が劣化することも確認された．

図 8 は DHCP プロセス数に対する Renew トランザクション性能を示す．図 8 より，ISC-DHCP の Renew トランザクション性能も並列プロセスを実行することにより大幅に改善可能であることが確認できる．特に，Discover トランザクションとは異なり，並列プロセス数が増加した場合にも特性の劣化が少ないことが確認できる．これは，Renew トランザクションは Discover トランザクションと比較してデータ入出力などへの負荷が少ないためと考えられる．なお，DHCP のトランザクション性能の飽和はデータ入出力などによるオーバーヘッドが要因として考えられるため，ディスクの高速化によりトランザクション性能を更に改善することも可能と考えられる．一般に，商用 DHCP 製品では，TFTP (Trivial File Transfer Protocol) や DOCSIS (Data Over Cable Service Interface Specifications) 特有のオーバーヘッドを除いた DHCP トランザクション性能は概ね 100〜300 トランザクション/秒程度である [14]．結果より，複数の DHCP プロセスを同時に実行することにより，商用 DHCP 製品と同様のトランザクション性能を実現可能であることが確認できる．

5.4　プロセス起動時間

CREID で利用する ISC ー DHCP の基礎特性を検証した結果，特にリースファイル内に多数のレコードが存在する場合に，プロセス起動時間が大幅に増加することが確認された．本検証では，CREID におけるプロセス起動時間を評価することで，長期に渡る DHCP サービスの停止が発生しないことを確認する．なお，プール IP アドレス数及び登録 MAC アドレス数として最大 100,000 を想定した．

図 9 に登録 MAC アドレス数に対する起動時間を示す．結果より，想定する登録 MAC アドレス数の範囲では，大きなプロセス起動時間の増加は見られていない．また，プロセス起動時間は 200ms 以下と非常に短時間であることが確認できる．

図 10 は貸出 IP アドレスのレコード数に対するプロセス起動時間を示す．結果より，貸出 IP アドレス数の増加に伴う DHCP プロセスの起動時間の増加は非常に小さいことが確認できる．また，プロセス起動時間は 200ms 以下と非常に短時間であることが確認できる．

図 11 は貸出 IP アドレスの貸出時間が過ぎたレコード数に対するプロセス起動時間を示す．結果より，ISC-DHCP を単一プロセスで実行した際には，プロセス起動時間が大幅に増加していたが，CREID では，貸出時間が過ぎたレコード数が増加しても，プロセス起動時間を非常に短時間に抑えていることが確認できる．

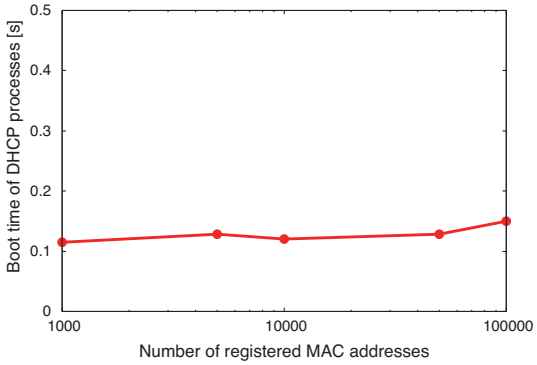

図 9　登録 MAC アドレス数に対する起動時間.

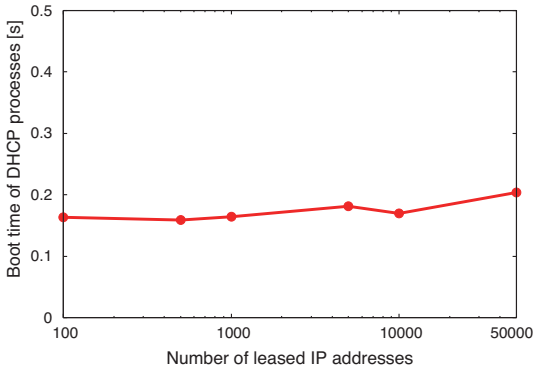

図 10　貸出 IP アドレス数に対するプロセス起動時間.

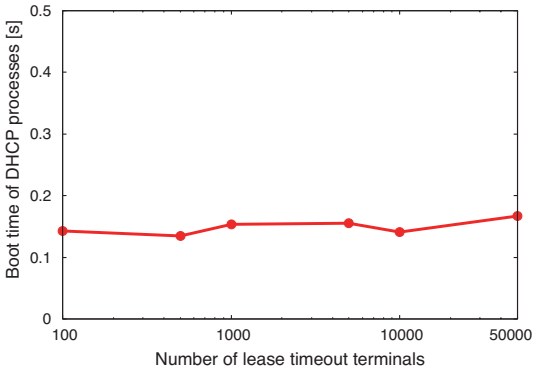

図 11　貸出時間が過ぎた IP アドレス数に対するプロセス起動時間.

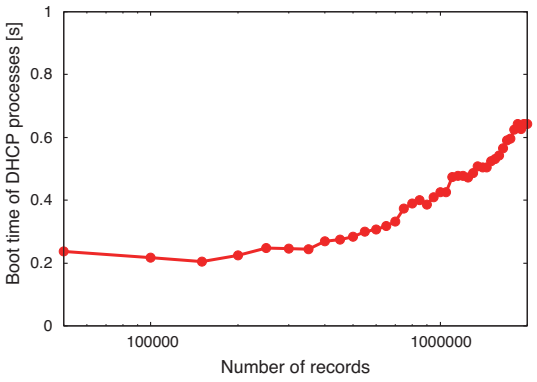

図 12　更新済みの貸出 IP アドレス数に対するプロセス起動時間.

図 12 は更新された貸出 IP アドレスのレコード数に対するプロセス起動時間を示す．結果より，図 11 と同様に，更新された貸出 IP アドレスのレコード数が増加した場合にも，プロセス起動時間を 1 s 以下と短時間に抑えていることが確認できる．また，上記の結果より，CREID を利用することにより，規模拡張性の高いシステムを実現可能であることが確認できる．

5.5　並列 DHCP プロセスのシステム負荷

本節では，DHCP プロセスを並列で実行することによるシステム負荷について検討を行う．

表 3 は DHCP プロセス数に対する UDP パケットの受信状態を示す．結果より，受信エラー及びバッファ溢れによるパケット損失は発生していないことが確認できる．CREID では，ネットワークを用いてクラスタリングを実現しており，パケット損失は DHCP 機能の大きなオーバーヘッドになる可能性があるが，本検証による特性劣化につながるパケット損失は発生しないことが確認された．

表 4 に DHCP プロセス数に対するメモリ使用量とスワップ使用量を示す．結果より，プロセス数の増加に対してメモリ使用量も増加するが，その増加度合いは緩やかであることが確認できる．また，スワップメモリは使用されていないことも確認されるため，十分な物理メモリの搭載のみで複数の DHCP プロセスの起動が実現可能であることが確認できる．

表 5 に DHCP プロセス数に対する CPU 負荷を示す．なお，r 値超過回数は DHCP 負荷試験中に r 値が 12(プロセッサ数× 3) を以上となった回数，ユーザー使用率超過回数は，DHCP 負荷試験中に 50%以上となった回数である．

結果より，プロセス数の増加に伴い，特に r 値最大値が顕著に増加するが，r 値最大値に比べて r 値の平

表 3　DHCP プロセス数に対する UDP パケットの受信状態

プロセス数	1	10	20	25	40	50	80	100
受信エラー数	0	0	0	0	0	0	0	0
バッファ溢れ	0	0	0	0	0	0	0	0

表 4　DHCP プロセス数に対するメモリ使用量とスワップ使用量

プロセス数	1	10	20	25	40	50	80	100
メモリ使用量 (KB)	830,144	1,003,468	1,141,088	1,214,228	1,424,456	1,577,352	1,999,016	2,289,708
スワップメモリ使用量 (KB)	0	0	0	0	0	0	0	0

均値は一桁程度小さく，CPU の負荷は常時高くないことが確認できる．また，b 値最大値も同様に増加するが，超過の度合いは r 値最大値に比較して十分に小さい．常時連続稼動するサーバー系プログラムでは，CPU 負荷が瞬間的に大きい場合であっても，平均負荷が十分小さく一定値で推移している限り，安定動作を期待できるため，本システム構成は安定動作を期待できると考えられる．

また ISC-DHCP の場合，ディスク I/O は主にログ出力とリースファイルへの追記で利用され，プロセス数増加によりログの合計出力量が増え bo 値が上昇している．しかし，システム性能への影響は概ね通常のプロセス挙動の範疇であり，想定するプロセス数においては安定動作を期待できると考えられる．

5.6　サービス可用性

3 章及び 5 章での評価実験結果に基づいて，CREID の可用性について検証を行う．検証では，表 6 に示すよう，50,000 台の DHCP クライアント接続時を想定し，DHCP のリース時間は 24 時間とした．また，1 日あたり 10 回のメンテナンスに伴うプロセス再起動，3ヶ月に 1 回のシステム設定変更に伴うプロセス再起動，1 日あたり 1 回のリースファイル最適化のための再起動を想定した．なお，DHCP のリース時間は 24 時間のため，リースファイル内の貸出 IP アドレス情報は 24 時間で 50,000 レコード更新される．また，メンテナンスに伴うプロセス再起動とリースファイル最適化のための再起動の合計 11 回を想定すると，1 回のプロセス再起動時には $50,000/11 = 4,545$ ほどの更新済みレコードが含まれると考えられる．図 12 より，50,000 の場合のプロセス起動時間は 0.2 s ほどのため，プロセス起動時間である $T_{short}, T_{long}, T_{optimization}$ は 0.25 s を想定する．

次に，システム障害の発生頻度として，年に 1 回を想定する．なお，CREID では現用系サーバーの動作を定期的に監視することにより，障害発生の検出を行っている．そこで，現用系サーバーの障害発生後から予備系サーバーのプロセス起動が開始するまでの切り替え時間を 10 回測定したところ，平均値は約 23.2 s であった．また，現用系サーバーで利用している設定ファイル及びリースファイルなどは予備系サーバーでのプロセス起動時に読み込まれるが，プロセス起動時間は現用系サーバーと同様に 0.2 s ほどかかることが考えられる．そのため，本検証では，システム障害時のサービス停止期間として 23.5 s を想定した．

DHCP クライアントが貸出 IP アドレスの更新処理を行う際に DHCP サービスが停止している場合，DHCP クライアントは一定時間後に更新処理を改めて行う．また，更新処理を改めて行うまでは，利用していた貸出 IP アドレスの利用を継続するため，通信サービスの停止は発生しない．CREID でも利用している ISC-DHCP では，10 s 後に更新処理を改めて実施するが，更新処理を改めて行うまでの期間は設定及び実装に依存する．そこで，本検証では，より厳しい条件として $T_{timeout}$ が 0 s を仮定する．

上記条件について，3 章の式を用いて可用性率の計算を行うと，1 日あたりの DHCP サービス停止期間 T_{down} は 5.3 s ほどとなり，DHCP サービス停止期間中に貸出 IP アドレスの更新を行う DHCP クライアント数 N_{down} は 0.44 台ほどとなる．結果として，1 台あたりの DHCP サービスの提供可能性を示す可用率 $P_{available}$ は 99.9991%となり，CREID は高い可用性を実現可能であることが確認できる．

表 5　DHCP プロセス数に対する CPU 負荷

プロセス数	1	10	20	25	40	50	80	100
r 値超過回数	0	0	1	2	1	9	9	4
r 値最大値	1	7	19	22	25	34	58	66
r 値平均値	0	1	1	2	1	7	12	7
ユーザー使用率回数	0	0	0	0	0	0	0	0
ユーザー使用率 (%) 最大値	7	4	5	6	7	9	13	15
ユーザー使用率 (%) 平均	6	3	4	4	6	7	11	14
b 値超過回数	0	0	0	0	0	0	0	0
b 値最大値	3	2	3	3	2	3	3	5
b 値平均値	1	1	1	2	1	1	2	2
CPU wait 率超過回数	0	0	0	0	0	0	0	0
CPU wait 率 (%) 最大値	9	19	18	17	17	19	17	17
CPU wait 率 (%) 平均値	6	16	16	16	15	17	15	14
bi 増分	1,202	746	220	452	208	630	179	347
bo 増分	28,902	96,571	105,216	111,363	124,097	137,204	161,086	184,383

表 6　可用性検証諸元

Number of CMs	50,000
Lease period of DHCP	24 hours
P_{short}	10 times / day
T_{short}	0.25 s
P_{long}	1 time / 3 months
T_{long}	0.25 s
P_{fail}	1 time / year
T_{fail}	23.5 s
$P_{optimization}$	1 time / day
$T_{optimization}$	0.25 s
$T_{timeout}$	0 s

6. ま と め

本論文で提案した CREID システムでは，無償の DHCP ソフトウェアであり，商用 ISP でも必要となる IPv4 及び IPv6 のアドレス貸出が実現可能な ISC-DHCP を基盤ソフトウェアとして採用した．そして，ISC-DHCP の信頼性及び規模拡張性を改善する手段として，クラスタリング技術である DRBD 及び Pacemaker を用いた．提案システムの実証実験を行うことにより，50,000 台を超える端末を収容する大規模ネットワークにおいても，DHCP のリース更新性能では 850 トランザクション/秒，可用性では 99.999%以上と，広域 IP ネットワークでの商用利用に耐える信頼性を達成できる事が確かめられた．

謝辞　機材及び資料を提供して頂いた Casa Systems, Inc. 及びオーエスエスブロードネット株式会社に感謝する．

文　献

[1] R. Droms, “Dynamic Host Configuration Protocol,” IETF RFC2131, March 1997.
[2] R. Droms, J. Bound, B. Volz, T. Lemon, C. Perkins, and M. Carney, “Dynamic Host Configuration Protocol for IPv6 (DHCPv6),” IETF RFC3315, July 2003.
[3] S. Alexander and R. Droms, “DHCP Options and BOOTP Vendor Extensions,” IETF RFC2132, March 1997.
[4] J. Littlefield, “Vendor-Identifying Vendor Options for Dynamic Host Configuration Protocol version 4 (DHCPv4),” IETF RFC3925, October 2004.
[5] Cable Television Laboratories Inc., “Operations Support System Interface Specification,” Cable Television Laboratories, Inc., CM-SP-OSSIv3.0-I07-080522, May 2008.
[6] Cable Television Laboratories Inc., “CableLabs’ DHCP Options Registry,” Cable Television Laboratories, Inc., CL-SP-CANN-DHCP-Reg-I02-080306, March 2008.
[7] Cisco “Cisco Network Registrar,” http://www.cisco.com, retrieved: January 2012.
[8] R. Droms, K. Kinnear, M. Stapp, B. Volz, S. Gonczi, G. Rabil, M. Dooley, and A. Kapur, “Draft, DHCP Failover Protocol,” IETF INTERNET DRAFT, March 2003.
[9] Internet Systems Consortium, ISC-DHCP, http://www.isc.org, retrieved: January 2012.
[10] B. Hellman, F. Haas, P. Reisner, and L. Ellenberg, “DRBD ユーザーズガイド日本語版,” LINBIT HA Sol. GmbH, May 2011.
[11] A. Beekhof, “Pacemaker 1.0 Configuration Explained,” http://www.clusterlabs.org, retrieved: January 2012.
[12] 宮副英治, “CREID 設定管理ガイド,” オーエスエスブロードネット (株), January 2012.
[13] dhcperf, http://www.nominum.com, retrieved: Jan-

uary 2012.
[14] http://cn.teldevice.co.jp/product/infoblox/ib_spec.html, retrieved: January 2012.
[15] http://paulroberts69.wordpress.com/2011/10/27/isc-dhcp-failover-is-just-too-complex/, retrieved: January 2012.
[16] http://www.accumuli.com/using-infoblox-dhcp-failover-part-1-i-3232.php, retrieved: January 2012.

内藤克浩

1999 年 慶大・理工・電気卒．2004 年 名大大学院博士課程後期課程修了．同年, 三重大・工・電気電子・助手．2007 年 三重大・工・電気電子・助教 2011 年 カリフォルニア大学ロサンゼルス校・客員研究員 博士（工学）無線ネットワーク, ITS, CATV システムの研究に従事．IEICE, IPSJ, IEEE 各会員．

西出誠

1997 年 岡山理科大・理・応用物理卒．オーエスエスブロードネット勤務, 2008 年 ネットステップ代表取締役就任, 現職．

宮副英治

1993 年 東北大・工・資源卒．富士通, General Instrument (現 Motorola) 勤務．2001 年 オーエスエスブロードネット代表取締役就任, 現職．