キャリアグレード・サービス・プラットフォーム

小池　友岳　　白鳥　毅　　鈴木　俊範　　吉本　正明
安藤　智和　　水上　貴司　　尾崎　裕二

既存電話網の並存を前提とした現在のVoIPサービスから本格的なIPプロトコルをベースにした次世代IPネットワークに向かって，固定網も移動網もオールIPベースの統合サービスネットワーク構築に向けた動きが活発になってきた。

ITUでも次世代移動網（Beyond 3G）やNGN（Next Generation Network）でも本格的なオールIPネットワークシステムが検討されている。これらはさらに固定網，移動網を問わないシームレスなネットワークサービスを提供するIPベースネットワークへと進化するものと考える。

一方，この次世代IPネットワークを構成するサービスノードの構成方法も，いわゆるソフトスイッチ・アーキテクチャに基づき，さらにLinux[*1)1)]に代表されるオープンプラットフォームベースの汎用サーバを利用することがコスト，開発期間の短縮などの面から有利であり，一般化している。

しかし，システムはオープン化するものの，キャリアの責務として果たすべき「信頼性」は，既存の電話網が保持しているものと同じく非常に高いレベルとなる。すなわち，これまで実現していた既存電話システムが前提の「信頼性」をそのままIPサービスノード単独で実現することが求められる。

この要求に対し，OSDL[*2)2)]（Open Source Development Lab）のワーキンググループのひとつで，通信事業に利用するためのLinux仕様であるCGL（Carrier Grade Linux）では2004年4月にSpec3.0がドラフト公開され，さらなる検討が進んでいる。

また，SAF[3)]（Service Availability[*3)] Forum）によってミッション・クリティカル・システム開発のために標準化されたオープンAPIをキャリア向けシステムとして実装する動向もある。

さらに，電話交換サービスでは一つの交換装置で膨大な数の加入者へサービス提供する必要があり，加入者より発生する多様かつバースト的要求に十分応えなければならず，これに必要な「サービス性能」も「信頼性」同様，損なわれてはならない。

沖電気はこれまでCenterStage®[*4)]として，**図1**に示すような大規模VoIP用サーバシステムを開発し，現在キャリアで商用システムとして運用されている。

そこでさらに来るべき本格的なオールIPベースの大規模VoIP時代に求められる「信頼性」並びに「サービス性能」を実現し，幅広くキャリア・ミッションクリティカル・システムへ利用可能な「キャリアグレード・サービス・プラットフォーム」（沖CGSP）を開発している。

本稿では沖CGSPへの要求条件，プラットホームアーキテクチャ，信頼性／サービス性能確保技術，今後の展開について述べる。

キャリアグレード・サービス・プラットフォーム

キャリアがサービスを維持するために必要となる要件に以下の二点があることを述べた。

- 高信頼性条件
  サービスの継続性
- 高性能条件
  高トラヒックでも十分なパフォーマンスを確保

これを満たすプラットフォームを汎用的なオープンシ

TIPS【基本用語解説】

OSDL（Open Source Development Lab）
エンタープライズ分野や高機能サーバ向けのLinuxをベースとしたオープンソースプロジェクトを支援するための非営利団体（NPO）。

SAF（Service Availability Forum）
ハードウェア，OS，DB（インメモリ）などとインターフェイスの整合性をとるミドルウェアおよびAPIを標準化するためのフォーラム。サン・マイクロシステムズやヒューレット・パッカード，IBM，Nokia，モトローラ，NECなど32社が参加する。

SPOF（Single Point of Failure）
システムの構成要素が1つしかないために，その個所で障害が起きると業務が止まってしまう弱点。

*1）LinuxはLinus Torvaldsの米国およびその他の国における登録商標あるいは商標です。　*2）OSDLはOpen Source Development Labs, Inc. の商標です。
*3）Service AvailabilityはService Availability Forumの商標です。　*4）CenterStageは沖電気工業（株）の登録商標です。

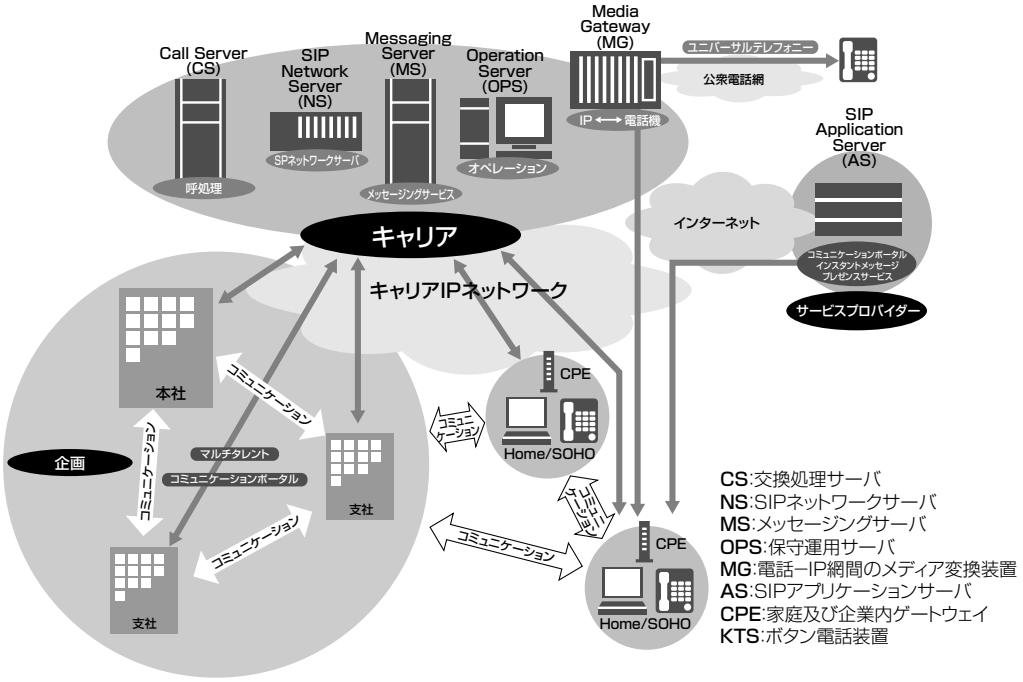


図1　CenterStageによるキャリアネットワーク

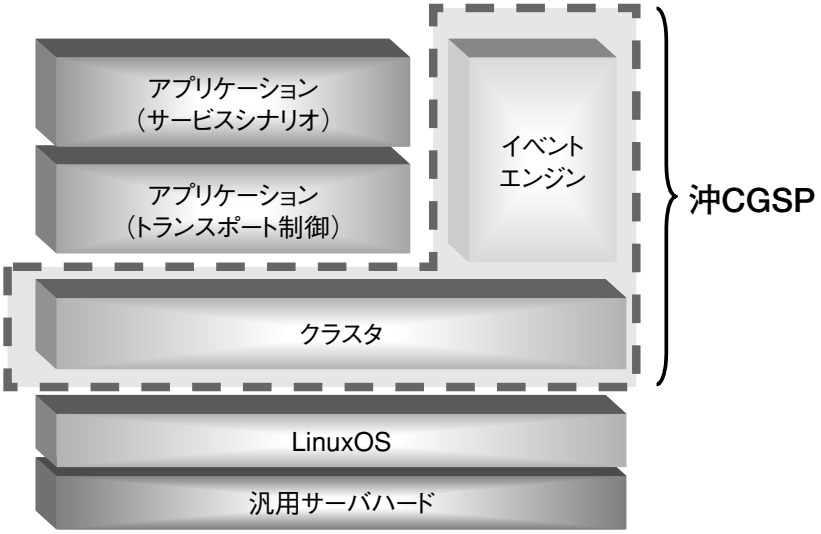


図2　沖CGSPのソフトアーキテクチャ

ステムであるLinuxを用いた「キャリアグレード・サービス・プラットフォーム（沖CGSP）」として検討している（**図2**）。本プラットフォームを実現するための要素技術を以降で説明する。

キャリアグレード高信頼化技術

ネットワークや装置がオープン化されても，その上でサービス・グレードを低下させることはできない。随時発生するネットワーク，ハードウェア，ソフトウェアに起因する障害に遭遇しても提供するサービスは継続しなくてはならず，既存電話交換機と同等としなければならない。

一般にキャリアサービスに必要な信頼性は6NINEs（99.9999%の稼働時間で年間の停止時間は32秒）とされ，市中製品のハイエンド・サーバの信頼性が4NINEs（99.99%の稼働時間で年間の停止時間は53分）程度であり，これを満たすにはネットワーク，ネットワーク装置，サーバ装置等すべてを冗長化することにより，SPOF（Single Point of Failure ）をなくし，より強固なシステムを構築する必要がある。

沖CGSPは汎用サーバ装置を電話交換ノードとしてクラスタリングし，さらに障害発生時の中断時間を極力短くすることにより，6NINEsの必要条件を満たすことに成功している。

(1) 高速フェールオーバを実現するクラスタリング

一般的なクラスタリングの方式には，共有ディスクを用いる「シェアド・ディスク方式」と，共有ディスクを用いない「シェアド・ナッシング方式」がある。前者の

「シェアド・ディスク方式」はデータに対して高い可用性を確保するために用いるが，共有ディスクを介して各ノードの状態を受渡しするため，障害時のフェールオーバ時間がディスクの性能に左右される。沖CGSPではサービスの継続性を重視し，後者の「シェアド・ナッシング方式」を採用し，さらにメモリ間で状態を引き継ぐことにより，より障害時のフェールオーバ時間を短くすることができる。

また，本プラットフォーム内のクラスタウェアはキャリア向けミッション・クリティカル・システムに特化したAPI（Application Programming Interface）をもち，アプリケーションからのフェールオーバ（系切替）要求に即時対応することができる。また，待機系のノードでもサービス開始に必要なリソースをクラスタウェアがアプリケーションと連携することにより確保することを可能としている。このため，プラットフォームより上位のアプリケーションの障害がトリガとなるフェールオーバの場合，サービス復旧に必要な時間は非常に短い。

(2) 仮想IP／MACアドレス

信号の送受信を実施するシステムのIPアドレスは1つとし，通信している対向システムからノードが切り替わったことが隠蔽される。このため，サービスで使用する仮想IPアドレスを付与しフェールオーバ時にノード間で引き継がれる。

また，IPアドレスが切り替わった際，信号を中継する上位ルータではARPテーブル（IPアドレスとMACアドレスの対照表）をフラッシュ／再設定する必要があり，それぞれのルータの性能に左右される。このため，仮想MACアドレスも付与しIPアドレスと同様に引き継ぐことにより，ルータのARPテーブルのフラッシュおよび再設定を不要とすることにより，ノードのフェールオーバ時間を短縮する（**図3**）。

(3) 他系メモリ転送による呼救済

電話交換システムのような複数の単純な非同期信号によって一つの呼を確立するシステムでは，それぞれのイベントにより遷移する呼の状態をノード内に保持しておく必要があり，コール・ステートフル・システムと呼ばれる。本プラットフォームは呼ごとに保持されるステートフル・データを待機系ノードへ転送する「他系メモリ転送機能」をアプリケーションに提供している。これにより，状態が遷移した時点でイベントドリブンにメモリエリアを転送するため，呼が発生し遷移したタイミングに発生した障害により系切り替えが発生しても，他系にてサービスを継続することができる。

(4) ライブパッチ

通常のオープンシステムではプロセスが起動している最中にプログラムを変更する場合，プロセスを再起動する必要があった。本プラットフォーム上で実現するアプリケーションはサービス提供している最中でもプログラムの関数ごとの変更を可能としている。このため，サービス稼働中に見つかったソフト問題や新たな機能追加をサー

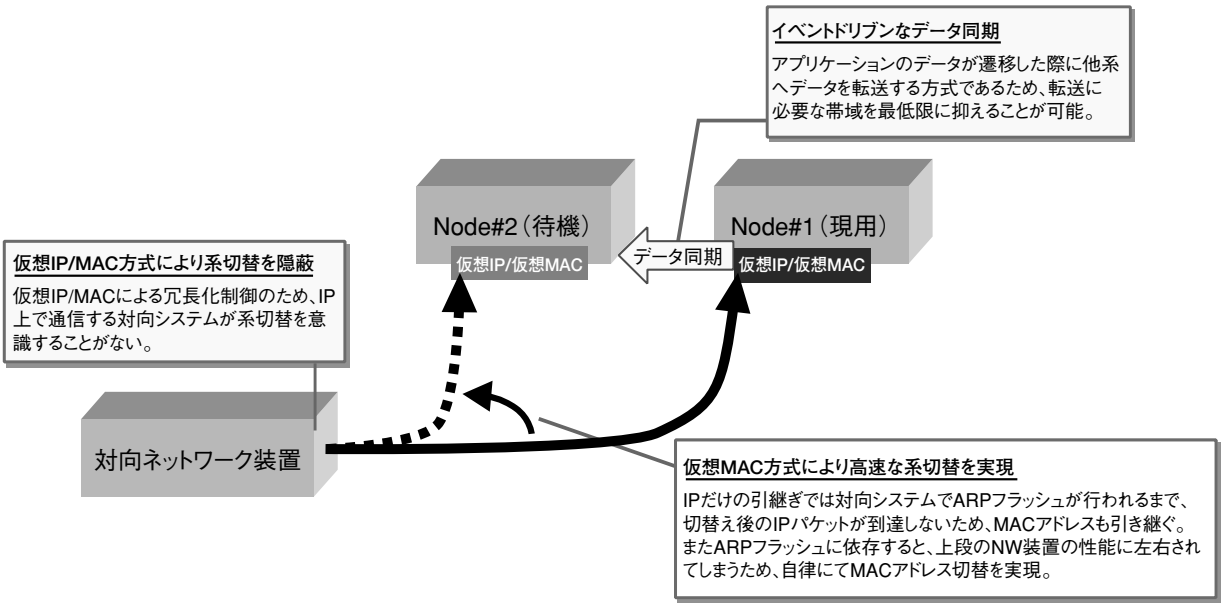


図3　クラスタ構成

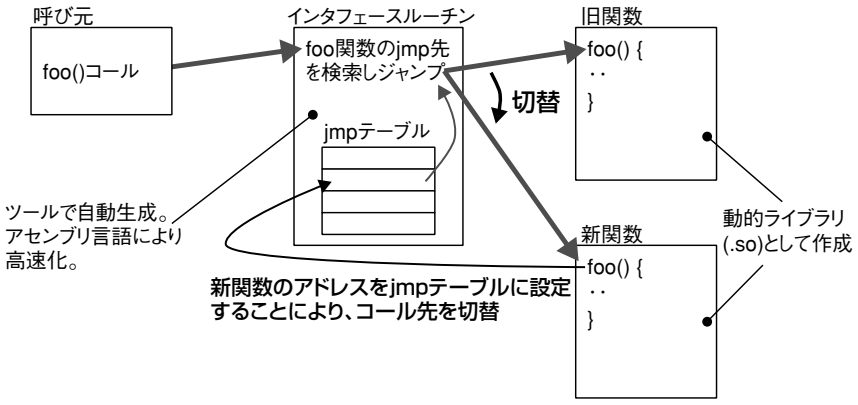


図4　ライブパッチ方式

ビス停止するとなく実施することができる。

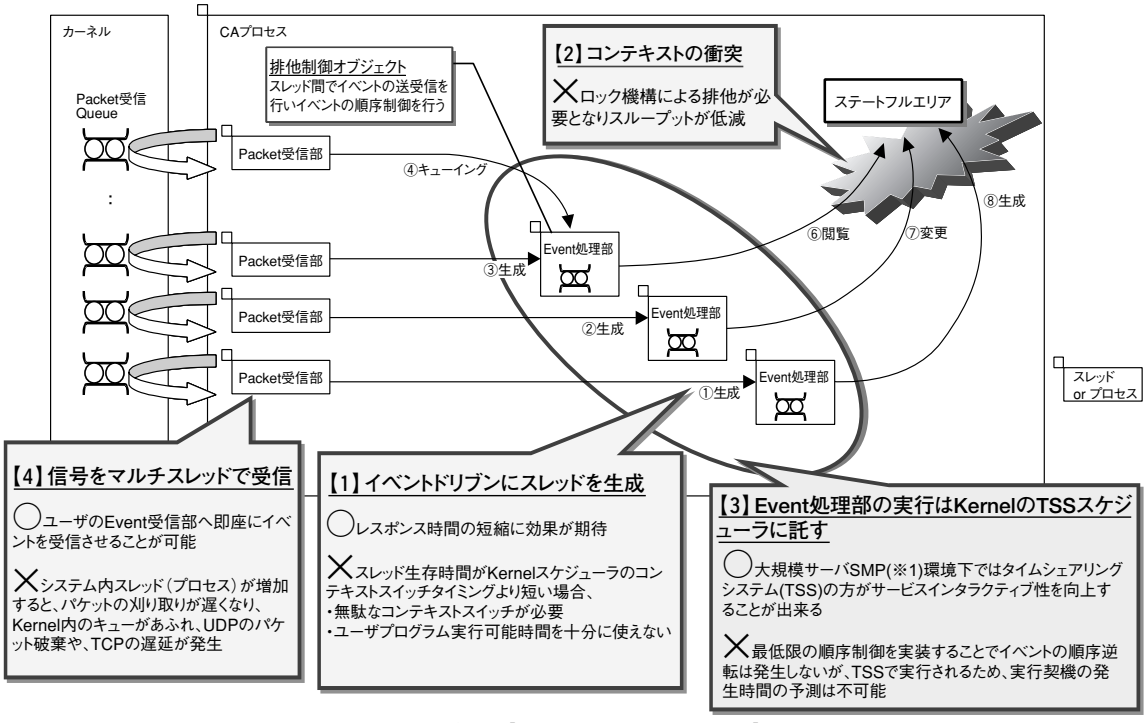
また，プログラム変更に割り当てられるコンテキストは変更の動作を割り込みが発生しないように行うことが可能であるため，サービスを実施している別のコンテキストと衝突することがなく，マルチプロセッサ，マルチコンテキストの環境でも安全なライブパッチを実現している（**図4**）。

キャリアグレード高性能化技術

電話交換サービスでは膨大な加入者をひとつの装置で一度に対応しなければならず，輻輳時にも十分にサービスを維持できるよう高い性能条件を確保する必要がある。

ところでこれまでコンピュータ技術の進歩により，非常に高性能CPU，並びにこのCPUを最大限使用することができるマルチタスク／スレッドOSが開発されてきた。一般的なマルチタスクOSでは，少ないCPUで複数のプロセス（もしくはスレッド）からの要求を同時に処理するため，各プロセスにCPU時間を分割し割り当てて制御するTSS（タイム・シェアリング・システム）スケジューラが実装されている。現在のコンピュータ・システムの多くはこのTSS（Time Sharing System）スケジューラを用いることで最大限に性能を発揮するよう設計されている。

しかし，電話交換システムのようなステートフルエリアを保持するシステムでは発生したイベントにコンテキストを割り付ける必要がある。同一呼へのイベントが輻輳した場合は，呼が保持する1つのメモリエリアへ複数のコンテキストが同時に書き込むことにより発生するデータ破壊が懸念される。このため，コンテキスト間で排他制御



※1）SMP【Symmetric Multiple Processor】
「対称型マルチプロセッサ」の略。複数のCPUが同等な立場で処理を分担するマルチプロセッサ手法

図5　一般的なマルチスレッドシステム概要図

する必要がある。一般的にはMutex（mutual-exclusion lock：相互排他ロック）に代表されるようなOSが提供している機能で実現するが，実際にコンテキストの衝突が発生した場合，排他されたコンテキストの処理が実行中のコンテキストの次に行なわれる処理の保証はされず，イベントの伝達遅延や最悪の場合デッドロックに陥る（**図5**）。

また，各イベント処理で必要となるCPU使用時間がクォンタム（スケジューラが各プロセスに割り当てる単位CPU時間）に満たない場合は，即座に他のプロセスにCPU時間を明け渡すこととなる。この時，処理していたプロセスの情報をロードしたCPU内のキャッシュを一旦破棄し他のプロセスの情報をロードするキャッシュ・バウンスが発生する。キャッシュ・バウンスが多発するとユーザ・プロセスが使用できるCPU時間が減少し，結果としてサービス性能の低減につながる。これが汎用OSをステートフルシステムに適用した場合におけるマルチスレッドの限界である。

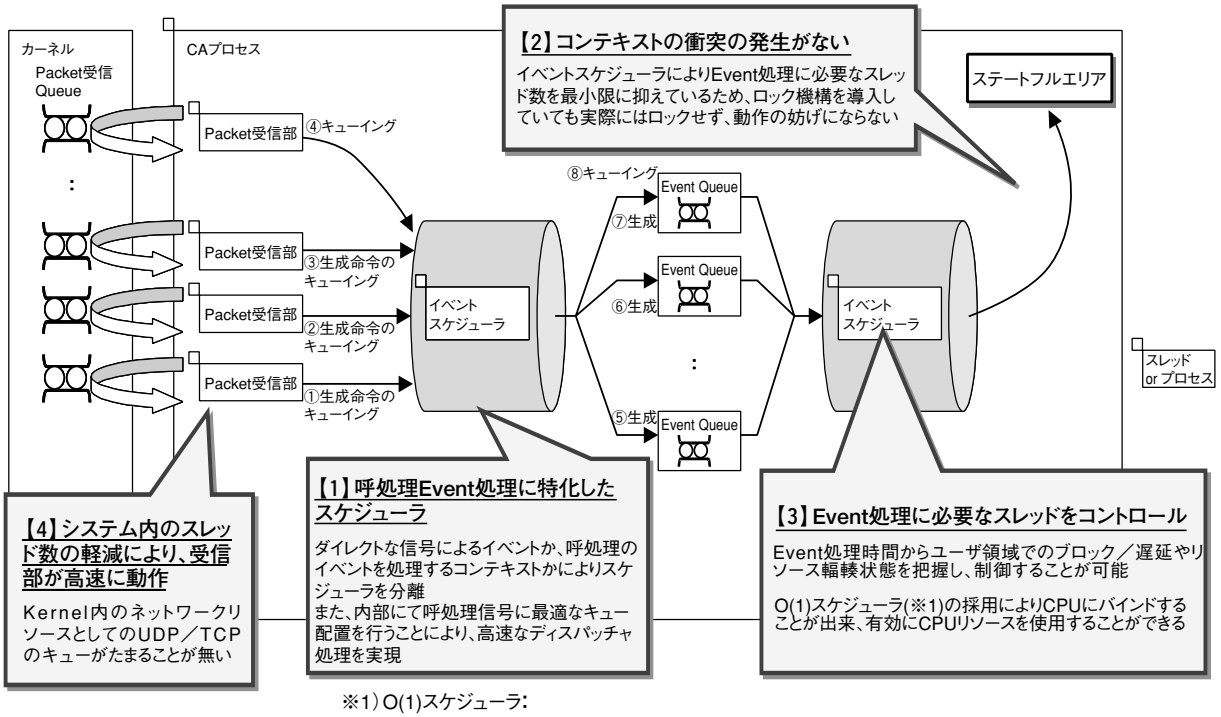
これを解決するため，沖CGSPでは輻輳的に発生する呼処理イベントを高速に処理する「呼処理イベントエンジン」をアプリケーションに提供している。呼処理イベントエンジンの機能要素の一部分を以下に紹介する。

（1）呼処理イベント排他制御

前述したコール・ステートフル・システムでは，呼状態を幾つかの呼オブジェクトとして保持している。それら呼オブジェクトをグループ化し，発生するイベントの優先／順序制御を実現する。さらにグループごとにコンテキスト（スレッド）を割り当てるため，呼オブジェクト内のステートフルエリアへアクセスする際，コンテキスト間の衝突が発生することがない。このため，不要なロックによる処理遅延が発生しない。

（2）呼処理イベントスケジューラ

入力／生成された一つの呼処理イベントを処理するために必要なCPU時間はイベントごとに異なる。一般的に呼処理イベントに必要なCPU時間はOSのスケジューラ・クォンタムよりも短いものがその大多数を占める。この場合，先に述べたようなキャッシュ・バウンスが多発し，複数のイベントからなるサービス処理にCPUを最大限に使用できない。このため，本プラットフォームでは「呼処理イベントスケジューラ」（**図6**）を設け，OSのスケジューラ・クォンタムより処理時間が短い複数のイベントを束ね，1イベント処理が終了しても次のイベントが起動するためのコンテキスト・スイッチが発生しないよう，呼処理イベントに特化したアプリケーション層で独自にス



※1）O(1)スケジューラ:
SMPやひとつのチップに複数のCPUコアを載せるたマルチコアプロセッサにに有効なLinuxにおけるスケジューラ
LinuxKernel2.6から採用されている。

図6　イベントスケジューラを用いたシステム概要図

ケジューリングする。これにより，少ないコンテキスト（スレッド）数で輻輳するイベントを処理することができ，信号受信部などイベント処理以外のコンテキストの動作契機も多くなり，高速な呼処理イベント処理を実現している。

実際にコール・ステートフル・システムに本呼処理イベントエンジン機能を実装することにより，2〜3倍の性能向上が図れることを確認している。

今後の計画／展望

今後キャリア向けミッション・クリティカル・システムCenterStage はフルIPネットワークへ移行していく通信キャリアのサービスや，既存通信装置をオープンシステムへのリプレースに対応していくことが計画されている。キャリアが求める信頼性と性能条件は非常に高いものであるが，プラットフォーム基盤として沖CGSPを使用することにより，その要件に十分対応できる基盤を築いていく。 ◆◆

■参考文献

1）Linuxコンファレンス Linux Kernel2.6 概説
http://www-6.ibm.com/jp/linux/event/2004/matsuri/data/a_05.pdf
2）OSDL キャリアグレード Linux
http://www.osdl.jp/osdl/lab_activities/carrier_grade_linux/
3）Service Availability Forum
http://www.saforum.org/home

●筆者紹介

小池友岳：Tomotake Koike. IPソリューションカンパニー ソリューション開発本部 キャリアネットワークソリューション開発部
白鳥毅：Tsuyoshi Shiratori. IPソリューションカンパニー ソリューション開発本部 キャリアネットワークソリューション開発部
鈴木俊範：Toshinori Suzuki. ネットワークシステムカンパニー メガキャリアビジネス本部 ソリューションSE第三部
吉本正明：Masaaki Yoshimoto. IPソリューションカンパニー ソリューション開発本部 キャリアネットワークソリューション開発部
安藤智和：Tomokazu Ando. IPソリューションカンパニー ソリューション開発本部 キャリアネットワークソリューション開発部
水上貴司：Takashi Mizukami. IPソリューションカンパニー ソリューション開発本部 キャリアネットワークソリューション開発部
尾崎裕二：Yuji Ozaki. IPソリューションカンパニー ソリューション開発本部