

映像配信サーバ向け ダイレクト ストレージ アクセス技術搭載 NPEngine™

NPEngine™ Equipped with Direct Storage Access Technology for Video Streaming Servers

山浦 隆博

山口 健作

橋 康達

■ YAMAURA Takahiro

■ YAMAGUCHI Kensaku

■ HASHI Yasumichi

近年のネットワークの広帯域化や視聴端末の普及に伴い、ネットワークを介した映像配信サービスが数多く展開されているが、従来の映像配信処理では、CPUやメモリなどの性能により同時配信できるユーザー数が制限されていた。

東芝は、この問題を解決するため、CPUやメモリを介さずにハードウェアでストレージからデータを読み出し、そのデータを直接ネットワークに送信するダイレクト ストレージ アクセス技術を搭載したNPEngine™を開発した。これにより、CPUやメモリの性能によらず映像配信処理を行うことができ、映像配信サーバ ExaEdge™ に適用することで、1台のサーバで同時に最大64,000セッションという多数のユーザーへの映像配信を実現した。

With the broad dissemination of wideband networks and mobile audiovisual terminals including smartphones and tablets in recent years, the number of video services provided via the network is increasing. However, the performance of conventional video streaming servers is limited by the speed of the central processing unit (CPU) and the bandwidth of the memory.

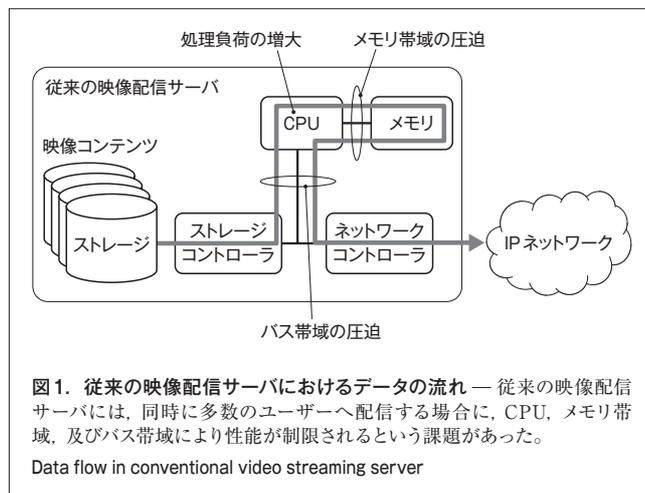
To overcome these problems, Toshiba has developed NPEngine™ equipped with a direct storage access technology that can directly transfer data from the storage to the network, bypassing the CPUs and memories in video streaming servers. This technology makes it possible to reduce the workload of CPUs and the bandwidth utilization of memories. We have applied it to the ExaEdge™ video streaming server, which can distribute video contents to a large number of users (up to 64,000).

1 まえがき

ネットワークの広帯域化と、テレビ (TV)、パソコン (PC)、スマートフォン、タブレットなどのネットワーク対応視聴端末の普及により、インターネットに代表されるようなIP (Internet Protocol) ネットワークを介した映像配信サービスが数多く展開されている。このようなサービスでは、1台で同時に多数のユーザーに配信できる映像配信サーバが求められているが、従来の映像配信サーバでは配信処理をCPU上で行っていたため、1台で映像配信できるユーザー数が限られていた。

東芝は、この問題を解決するため、CPUやメモリを介さずにハードウェアでストレージからデータを読み出し、そのデータを直接ネットワーク処理して送信するダイレクト ストレージ アクセス (DSA) 技術を搭載したNPEngine™を開発し、映像配信サーバ ExaEdge™ に適用した。ExaEdge™ は、映像コンテンツを記憶するストレージに多数のSSD (ソリッドステートドライブ) と、映像コンテンツを送信するための合計40 Gビット/sのネットワーク インタフェースを備えており、1台で多数のユーザーに映像を配信できる。

ここでは、まず従来の映像配信サーバにおける課題について述べた後、その課題を解決するDSA技術を搭載したNPEngine™の概要と、映像配信サーバ ExaEdge™ について述べる。



2 従来の映像配信サーバの課題

映像配信サーバは、映像コンテンツをストレージに蓄積し、ユーザーからの視聴要求に対して、映像コンテンツをストレージから読み出し、映像配信規格に従ってIPネットワークに送信する。従来の映像配信サーバにおけるデータの流れを図1に示す。

従来の映像配信サーバでは、まず、CPUがストレージコントローラに指示を出し、ストレージ内に格納されている映像コ

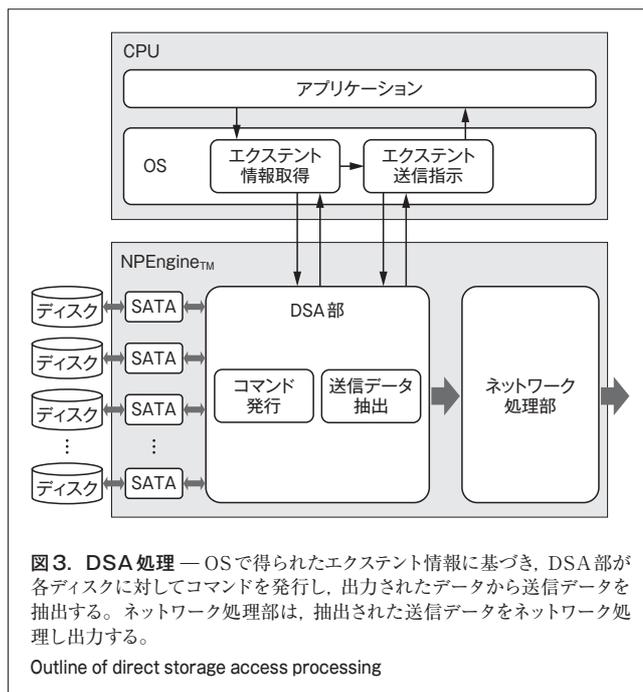
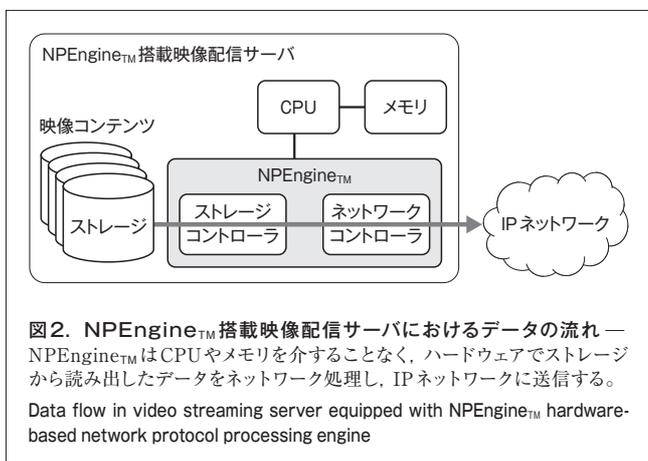
コンテンツファイルをメモリに読み出す。次に、CPUが送信すべきデータを抽出し、映像配信規格に従ってネットワーク処理を行い、ネットワークコントローラに対してデータをIPネットワークに送信するよう指示することで映像配信を実現する。

映像配信サーバのストレージとネットワークの部分では、近年の高速化により、多数のユーザーに対して十分に配信できる速度が実現されつつある。しかし、ストレージからデータを読み出すためのストレージ読み出し処理と、ネットワーク配信のためにパケット化するネットワーク処理を行うCPUでは、多数のユーザーへの配信を行うと高い処理負荷が発生する。また、メモリとCPUの間でも、多量のメモリアクセスが発生するため、メモリの入出力帯域が圧迫される。更に、読み出されたデータは、ストレージコントローラからメモリへ、メモリからネットワークコントローラへとデータ転送が行われるため、配信帯域の増加に伴い、バスの帯域も圧迫されるという問題があった。このため、多数のユーザーに大規模な配信を行う場合には、処理性能が高く、十分なメモリ帯域やバス帯域を備えた高性能なCPUを利用する必要があり、コストや消費電力の増大につながっていた。

3 DSA 技術を搭載した NP Engine™

そこで当社は、これまで開発してきたTCP/IP (Transmission Control Protocol/IP) 通信処理をハードウェアで行う NP Engine™^{(1), (2)}を拡張し、ストレージから読み出したデータを直接ネットワーク処理して送信する DSA 技術を搭載した NP Engine™を開発した。図2に示すように、NP Engine™では、専用ハードウェアでストレージの読み出し処理とネットワーク処理を行い、ストレージ内の映像コンテンツをサーバのCPUやメモリ、バスを介さず直接IPネットワークに配信する。

DSA 処理の概要を図3に示す。NP Engine™は、直接ストレージからデータを読み出す DSA 部と、映像配信規格に従ってネットワーク処理を行うネットワーク処理部から構成される。

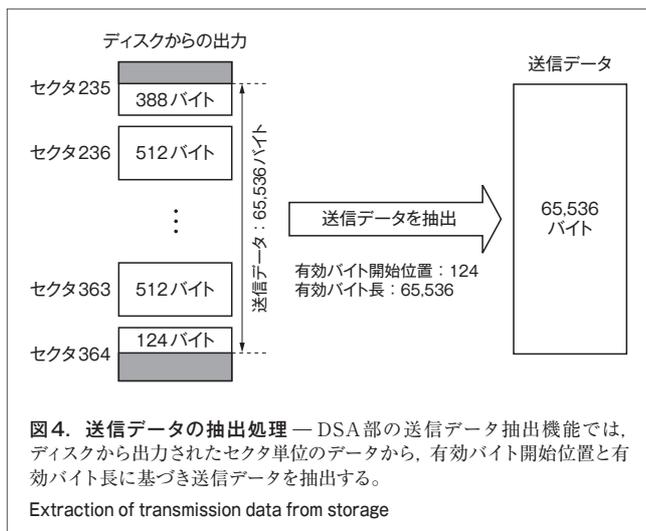


DSA部は、SATA (Serial Advanced Technology Attachment) インタフェースでディスクと接続されており、ディスクに対してデータ送信などのコマンドを発行する機能と、コマンド発行により出力されたデータから送信データを抽出する送信データ抽出機能を備えている。コマンド発行機能は、DSA部に接続されている複数のディスクに対して読み出しコマンドを振り分けることで、RAID 0 (Redundant Arrays of Independent (Inexpensive) Disks 0) 機能を実現している。これにより、CPU上で動作するOS (オペレーティングシステム) からは複数のディスクを一つのディスクとして扱える。

ディスクに対するファイルシステム処理はOSによって行われる。ファイルシステムでは、データはエクステントと呼ばれる可変長の連続した記憶領域の単位で管理されており、ファイルに対応するエクステントがディスクのどの部分に記録されているかを示すエクステント情報をOSからNP Engine™に渡すことで、以降のストレージ読み出し処理とネットワーク処理をNP Engine™でハードウェア処理して送信する。

まず、DSA 処理では、アプリケーションからファイル送信の指示があると、ファイルシステム処理機能によりOSがファイルのエクステント情報を取得する。このエクステント情報はディスクに格納されており、DSA部のコマンド発行機能を介してディスクから取得できる。

続いて、取得したエクステント情報に基づきエクステントを送信するようにDSA部に指示する。DSA部はコマンド発行機能を用いて各ディスクにデータの出力を指示するが、ディスクから出力されるデータはバイト単位ではなくセクタと呼ばれるより大きな単位で出力される。セクタは、例えば512バイトや



4,096バイトなどの大きさである。このため、**図4**に示すように、DSA部の送信データ抽出機能で、ディスクからセクタ単位で出力されるデータから、有効バイト開始位置と有効バイト長を用いて有効部分だけを取り出す処理を行う。これにより、ネットワーク処理部に対してバイト単位で送信するデータを出力する。このデータに対して、ネットワーク処理部で各プロトコルに従った処理を行い、ネットワークに送信する。

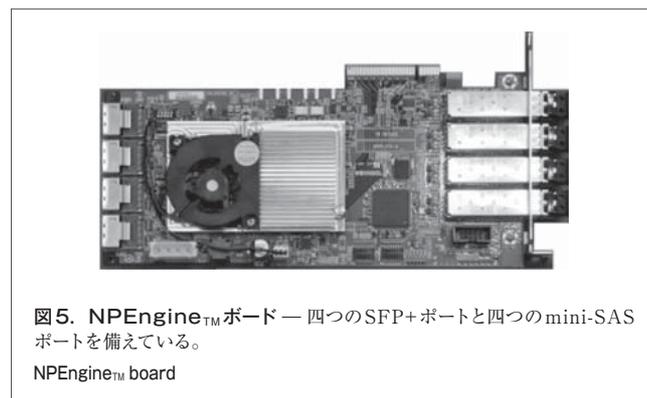
ネットワーク処理部は、プロトコルの規定によりフロー制御（データの取りこぼしを防ぐためのデータ送信制御）を行う。このため、DSA部でもネットワーク処理部の送信に合わせながらディスクに対してデータの出力を指示し、送信指示のあったエクステントの送信が完了するまでこれらの処理を繰り返す。

このように、DSA部とネットワーク処理部を備えたNPEngine™を用いてストレージから読み出したデータを送信することで、CPUはエクステント情報の取得とエクステントの送信を指示するだけで、以後のディスクに対するコマンド発行やフロー制御などの処理は全てハードウェアで処理されるため、CPUの負荷を軽減できる。また、直接ストレージからデータを読み出し、そのデータをハードウェアでネットワーク処理して、IPネットワークへ配信できるため、メモリやバスの性能に依存することなく多数のユーザーへ同時配信ができる。

4 映像配信サーバExaEdge™

前章で述べたNPEngine™を搭載し、従来の映像配信サーバの課題を解決した映像配信サーバExaEdge™を開発した。ExaEdge™は、多数のユーザーへの配信を行うため、40 Gビット/sの帯域でネットワークと接続されるインタフェースと、16台のストレージと接続されるインタフェースを備えている。

ExaEdge™は、**図5**に示すNPEngine™ボードを搭載している。NPEngine™ボードは、ネットワークを接続するインタ



フェースとして、SFP+ (Small Form-Factor Pluggable Plus) を4ポート備えており、それぞれ10 Gigabit Ethernetでネットワークに映像コンテンツを送信できる。40 Gビット/sという配信能力は大手の配信業者にとっても十分な性能である。また、インタフェースに40 Gigabit Ethernetを用いず、10 Gigabit Ethernetを四つのインタフェースとしていることで、比較的安価な10 Gigabit Ethernet機器と接続できる。

また、NPEngine™ボードは、ストレージを接続するインタフェースとしてmini-SAS (Serial Attached SCSI (Small Computer System Interface)) 規格のインタフェースを4ポート備えており、SATA 3.0対応のSSDを最大16台接続できる。このインタフェースは、**図6**に示すように前面のディスクアレイと接続されており、ここに映像コンテンツを蓄積できる。16台のSSDを用いることで、40 Gビット/sでの配信に耐えるディスク性能を確保している。

これらのインタフェースを備えるNPEngine™ボードを用いて、映像コンテンツをハードウェアによりストレージから読み出して直接ネットワークに出力することで、CPUやメモリ、バスの帯域に依存することなく配信できる。ExaEdge™を用いることで、従来は複数台のサーバで行っていた多数のユーザーへの映像配信を1台で行え、消費電力などのコストを削減できる。

ExaEdge™は、RTP (Real-time Transport Protocol) とHTTP (Hypertext Transfer Protocol) の2種類の映像配信規格に対応している。以下では、それぞれの特長について述べる。

4.1 RTP 配信の特長

RTP 配信では、IPTV (Internet Protocol TV) 規定のVOD (Video on Demand) 仕様⁽³⁾に準拠した配信ができる。IPTV 規定では、一時停止や早送り、巻戻しなどといったトリックプレイへの対応が規定されている。トリックプレイは、映像コンテンツ中のIフレーム(他のフレームの情報をういず符号化されたフレーム)を抽出して送信することで実現される。

従来の映像配信サーバでは、あらかじめトリックプレイ用にIフレームを抽出したデータを、別データとしてストレージに生成しておき、トリックプレイを実現するものがあった。しかし、これらの別データに起因して、利用できるストレージの容量が減ってしまうという問題があった。

ExaEdge™では、ハードウェアでリアルタイムに映像コンテンツからIフレームを抽出して送信する機能を備えている。これにより、あらかじめ別データを生成しておく必要がなく、ストレージ容量をむだにすることなく利用できる。

また、RTP 配信ではネットワーク機器や受信端末でのバッファあふれを防ぐため、ユーザーの受信するレートが常に一定となるよう、パケットを等間隔で送信する制御が求められる。通常のソフトウェア処理では、多数のユーザーへの配信を行うと、短い時間間隔でパケット送信を制御できないため、パケットの送信間隔が乱れてしまう。ExaEdge™では、パケット送信の制御をハードウェアで行うため、送信レートを常に一定に保ち、安定した配信が実現できる。

RTP 配信では、四つのネットワークインタフェースを用いた構成により、合計で最大16,000セッション、36 Gビット/s以上の送信レートで配信できる。

4.2 HTTP 配信の特長

HTTPによる映像配信では、NP Engine™を、OSが提供する標準的なネットワーク制御のためのAPI (Application Programming Interface)であるSocket APIから利用できる。これにより、Apache HTTP Server™^(注1)などのHTTP配信アプリケーションソフトウェアでは、ソースコードを改変することなく、NP Engine™を利用した映像配信を行える。

HTTP 配信では四つのネットワークインタフェースを用いた構成により、合計で最大64,000セッション、36 Gビット/s以上の送信レートで配信できる。

5 あとがき

DSA 技術を搭載したNP Engine™及び映像配信サーバExaEdge™を開発した。ExaEdge™は、映像コンテンツを蓄

(注1) Apache及びApache HTTP Serverは、The Apache Software Foundationの商標。

(注2) NAB及びNAB Showは、National Association of Broadcastersの登録商標。

積するストレージとして16台のSSDを用い、NP Engine™でSSDから読み出した映像コンテンツをネットワーク処理し、直接IPネットワークに送信する。これにより、従来の映像配信サーバで問題となっていたCPUやメモリ、バスなどの性能によらず、1台で多数のユーザーに対して配信できる。ExaEdge™では、RTP 配信で最大16,000セッション、HTTP 配信で最大64,000セッションの配信を実現できる。

NP Engine™は、2012年4月に米国ラスベガスで開催されたNAB Show^{®(注2)}で、米国放送業界専門誌TV Technologyから技術的ブレークスルーを切り開いたとして、Mario Awardを受賞した。

今後は、ネットワークのいっそうの広帯域化や、4K2K(約4,000×2,000画素)など映像コンテンツの高解像度化が進み、ネットワークを流れる映像の配信数やコンテンツ容量は更に増加すると思われる。当社は、DSA技術という映像配信における基盤を確立したことにより、ストレージインタフェースやネットワークインタフェースをより高速な規格に対応させることで、今後の配信数の増加やコンテンツ容量の増加といったトレンドにもすばやく対応できると考えている。

文 献

- (1) 田中信吾 他. 超高速TCP/IP通信ハードウェア処理エンジンNP Engine™. 東芝レビュー. 65, 6, 2010, p.40 - 43.
- (2) 田中信吾 他. Gigabit/10Gigabit Ethernetに対応した高効率TCP/IPオフロードエンジン. 情報処理学会論文誌. 52, 12, 2011, p.3715 - 3728.
- (3) IPTV Forum Japan. "IPTV 規定 VOD仕様 1.2版". <<https://www.iptvforum.jp/download/>>, (参照2012-12-18).



山浦 隆博 YAMAURA Takahiro

研究開発センター ネットワークシステムラボラトリー。
通信プロトコル処理の研究・開発に従事。電子情報通信学会、IEEE 会員。
Network System Lab.



山口 健作 YAMAGUCHI Kensaku

研究開発センター ネットワークシステムラボラトリー研究主務。
情報通信機器の開発に従事。電子情報通信学会会員。
Network System Lab.



橋 康達 HASHI Yasumichi

社会インフラシステム社 府中事業所 放送・ネットワークシステム部。通信システム機器の開発に従事。
Fuchu Complex