200 Gbpsの超高速通信をリアルタイムにソフトウェアで解析する技術

Software Technology Offers Real-time Analysis of Super-fast Telecommunications at 200 Gbps

● 野村祐士　● 田村雅寿　● 小沢年弘

あらまし

スマートフォンやタブレットなどの端末の普及や，データセンター活用シーンの拡大により，クラウドサービスが大きく進展している。これにより，通信ネットワークを利用したサービスの品質向上がますます重要になっている。一方で，ネットワークのデータ流量増大や複雑なシステム構成などにより，遅い・つながらないといったサービス障害が発生することがある。このようなサービス障害を早期に発見して復旧させるためには，その原因がネットワーク品質にあるのか，またはアプリケーション側にあるのかを解析して切り分ける必要がある。このためには，通信パケットごとの詳細な挙動をリアルタイムに解析する必要がある。また，複雑な障害については一旦障害の証拠を蓄積しておき，事後に詳細な分析を行うことが求められている。このような要求に応えるため，富士通研究所は超高速なネットワークの通信パケットをリアルタイムに解析するとともに，蓄積・検索も高速に実現する技術を開発し，高価な専用ハードウェアが不要なソフトウェアとして汎用システムに実装した。

本稿では，高速通信を解析・蓄積・検索するためのソフトウェア技術，およびそのシステム構成について述べる。

Abstract

Smartphones and tablets have become widely and readily available, and the expanding scope of applications for datacenter use has stimulated the advancement of cloud services. Thus, it is becoming crucial to maintain the quality of IT-network-based services. Meanwhile, an ever-increasing volume of data traffic and further sophistication of system structures are leading to service disruptions such as slow data propagation speeds and difficulties in establishing network connections. For early detection and swift recovery of such service disruptions, it is essential to analyze and distinguish between their causes to know whether they are due to network (connection) quality or lie in application programs. This requires detailed, real-time behavioral analysis of each telecommunication packet. For complicated cases, evidence of the errors may need to be accumulated for later in-depth analyses. In order to meet these requirements, Fujitsu Laboratories has developed technology that facilitates real-time analyses of telecommunication packets on super-fast communication networks, as well as high-speed data accumulation and queries. The technology has been realized in the form of software that requires no expensive, purpose-specific hardware, and it is installed on a versatile system. This paper describes the software technology that enables to analyze, accumulate and search data from high-speed telecommunications. It also explains the system structure.

ま え が き

ネットワークを流れる通信量は年々増加し続けており，通信品質の監視や遅延などの発見は大きな課題になっている。従来，通信速度が20 Gbpsを超える通信回線では，高価なハードウェアを使用しても全ての通信の品質は解析できず，通信サービスを利用するユーザー全体に快適で安定したサービスを提供できなかった。

富士通研究所は，マルチコアCPUに適したパケット収集技術と，TCPなどのトランスポート層とHTTPなどのアプリケーション層それぞれの品質解析技術，および収集したパケットデータを高速に蓄積・検索する技術を開発し，[(1), (2)] ソフトウェアとしてシステムに実装した。これにより，汎用サーバ1台とソフトウェアだけで200 Gbpsの通信品質のリアルタイム解析が可能になった。また，パケットデータを蓄積する汎用サーバと組み合わせたシステムを構成することで，40 Gbpsでの蓄積・検索が可能になった。

このシステムをキャリアネットワークやデータセンターのネットワーク品質監視に適用することで，多くのユーザーに快適で安定したサービスを提供できる。

システム構成

図-1に示すように，本技術を実装したシステムは，最大で200 Gbpsの超高速通信でパケットデータを収集するストリーム解析部と，収集したパケットデータを蓄積するスケーラブル蓄積部から成る。ストリーム解析部は，パケットの到着時間を一つのクロックで管理できるように1台の汎用サーバで構成している。また，スケーラブル蓄積部は，蓄積容量を増やせるように複数の汎用サーバで構成している。

従来，高速のパケット収集に関しては，ハードウェアに搭載されたCPUなどの性能に依存して，単位時間あたりの受信可能な最大パケット数が頭打ちとなることがあった。また，メモリアクセス性能に依存してパケット収集，ネットワーク品質解析，およびアプリケーション品質解析のそれぞれの処理間のメモリコピーが間に合わないといった問題により，処理性能を向上させることが困難であった。

今回，200 Gbpsでの高速なパケット収集に加え，ネットワーク品質解析，およびアプリケーションの品質解析を高価な専用ハードウェアを使わずにソフトウェアだけで実現した。ネットワーク品質解析では，ネットワーク上のサービスやユーザーごとの詳細な通信量，パケットロス，ネットワーク遅延などネットワークが原因となる通信品質の問題が検出できる。また，アプリケーション品質解析では，アプリケーションごとの通信量の算出や，サービスを提供するアプリケーションの応答遅延などが原因となる問題が検出できる。これらの品質解析を統合することで，サービス遅延の原

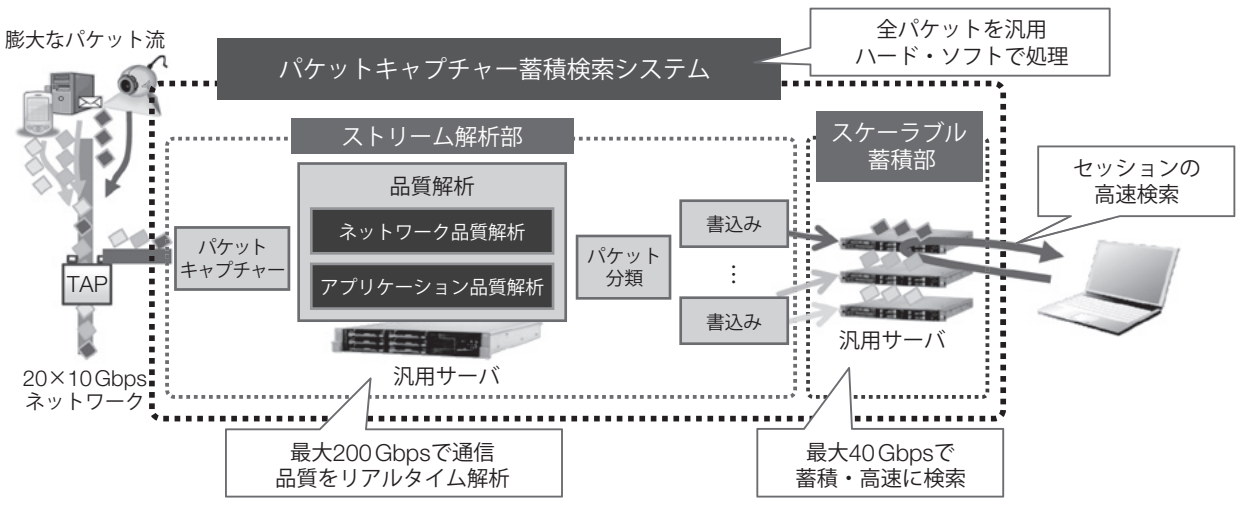

図-1　開発技術を実装したシステム構成

因と場所を特定できる。

開発技術の特長

本章では，開発した技術の特長を述べる。パケット収集高速化技術，およびメモリアクセス高速化技術を**図-2**に示す。

(1) パケット収集高速化技術

OSにおいてパケット受信（収集）処理は処理負荷が高く，システム全体の負荷を増加させる。特に受信処理の高速化には，負荷の高い割込み処理の効率的な削減が重要である。今回，パケットを受信する際に，パケット到着ごとに発生していた割込みを集約することで，処理回数を削減する技術を開発した。また，これをドライバとして実装することで，OSにおける処理負荷の削減に成功した。加えて，割込み処理を複数のCPUコアに負荷分散することで，更にパケット受信処理性能を向上させた。

(2) メモリアクセス高速化技術

現在の汎用サーバのメモリ帯域は200～400 Gbps程度であり，200 Gbpsのパケットデータを何度も読み書きして解析するには十分な帯域幅ではない。これが，通信品質解析の処理負荷が高い大きな要因である。従来，パケット解析処理は，パケット収集，ネットワーク品質解析，アプリケーション品質解析という三つの異なる処理を順番に実行しており，それぞれの処理間でパケットデータをコピーして受け渡ししていた。この方法では，例えば200 Gbpsのパケットデータの解析のためにコピーが3回発生すると，600 Gbpsメモリ帯域が必要となる。このため，メモリ帯域が不足して200 Gbpsでの解析が不可能になる。この問題に対して，パケット収集や品質解析の処理間でデータ参照の方法やタイミングを工夫し，パケットや解析データのコピーをせずに参照可能にするとともに，同時の書込みや参照中の領域への書込みが起こらないようにした。これにより，コピーや処理の排他制御を不要にした。

(3) 処理の並列化技術

複数のCPUコアを効果的に利用する並列化技術を**図-3**に示す。

現在のCPUコアは高速になってはいるが，1CPUコアで200 Gbpsのパケット解析処理を実行できるほど高速ではない。このため，いかに複数のCPUコアを効果的に利用して並列にパケット解析処理を実行できるかが重要である。多数のパケット解析処理が同時並行で実行される際には，解析処理性能がCPUコア数に比例して向上しないという問題が生じる。これは，従来は同一データ構造を複数の解析プロセスで共有する場合にロックなどの排他処理が必要であり，同時並行で解析処理ができないためである。これを解決するため，複数の

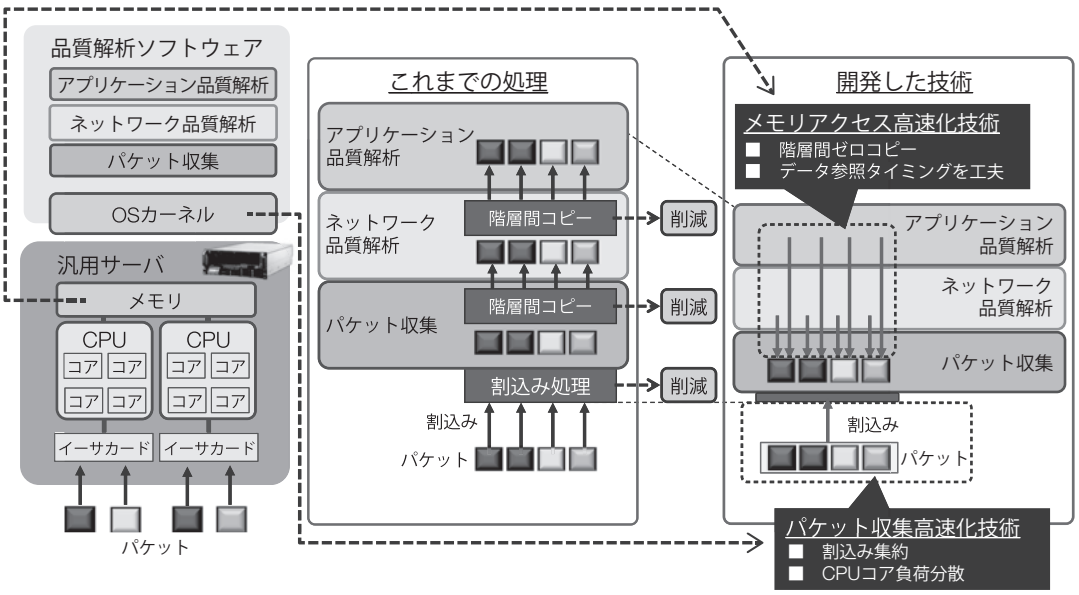

図-2　パケット収集高速化技術とメモリアクセス高速化技術

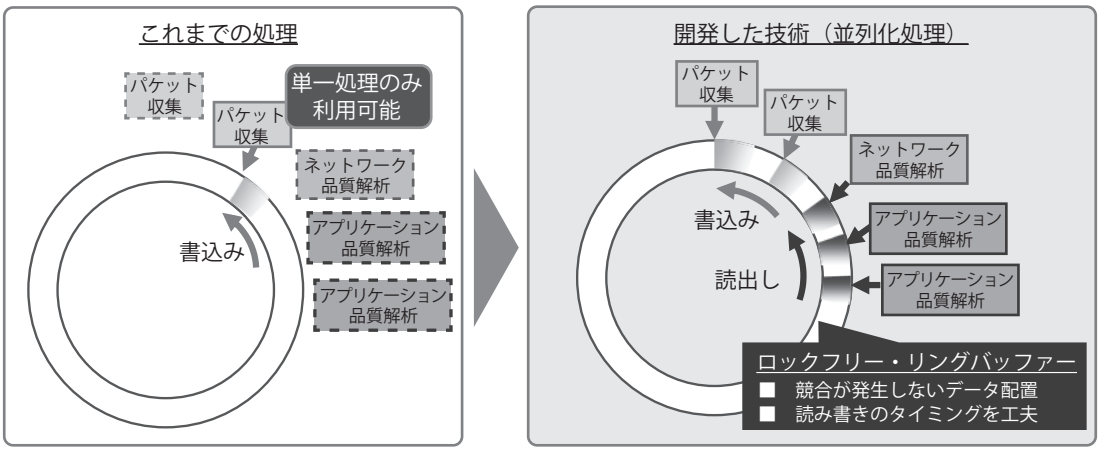

図-3　複数CPUコアを効率的に利用する並列化処理

CPUコア上で動作する多数の解析プロセスから，排他制御不要で同一のロックフリー・リングバッファー(注)にアクセスできる技術を開発した。具体的には，図-3のように，複数プロセスが同一メモリ領域を参照しない場合には，排他制御をしないようにデータ構造を工夫している。これにより，複数のCPUコアが競合しないようにし，CPUコア数の増加に伴い解析処理性能をリニアに向上させることを可能にした。

スケーラブル高スループット蓄積機構

超高速なネットワークの通信パケットを収集してリアルタイムに解析する技術と組み合わせて，40 Gbpsまでの流量のパケットを蓄積し，高速な検索を可能にする技術を開発した。この技術も，汎用のサーバとスイッチのみを用いて，複数の汎用サーバを協調して動作させることで，高い拡張性と信頼性を実現している。

今回開発したスケーラブル蓄積部は，ディスク容量の拡張性と並列書込み性能を確保するため，複数のサーバをネットワークで接続して協調動作させた分散型のストレージとして構成した。このため，キャプチャーしたデータは，そのネットワークを介して蓄積されることとなる。

● 高速蓄積

通信パケットを全て蓄積するには，次々に送られてくるデータの流れのペースに遅れないようにする必要がある。HDDの蓄積性能が最も高いのは，データをシーケンシャルに書き込んでいった場合である。後述する高速検索との両立のため，パケットを到着した順に書き込むのではなく，パケットの解析結果に応じて数Mバイト程度の大きさにまとめてから書き込んでいる。

パケットデータの蓄積は，古いデータを削除しながら新しいデータを蓄積することになる。古いデータの削除時に，HDDの領域管理情報の更新のためにアクセスが発生すると，データ書込みのシーケンシャル性が阻害され，蓄積性能が悪くなる。そこで，HDDの領域管理情報をメモリ上の管理領域に保持している。その際，書込みデータのサイズを固定長にして，この管理領域の処理を単純化している。データを書き込む際のHDD領域の割当ては，空き領域の分布に関わらず，アドレスが増加する方向に割り当てていく。そして，終端に達したら再度先頭から領域割当てを行うことで，シーケンシャル性を維持している。

● 高速検索

蓄積された大量の通信パケット群から，目的としているデータを高速に検索する性能も重要となる。検索を高速に行うために，パケットの到着時刻に加え，検索条件に応じた加工をして蓄積している。パケットデータ解析において典型的に用いられる検索のキーとして，時刻と送信元アドレスがある。このうち，送信元アドレスによる検索への最適化として，全パケットを時系列で書き込ん

(注)　データの上書きを可能にするデータ格納方式の一つがリングバッファーであり，リングバッファーのうち排他制御を不要とするバッファーアクセス方式のこと。

でいくのではなく，送信元アドレスの適当な範囲（これを送信元グループと呼ぶ）と，パケット到着時間の範囲の二つの軸でグルーピングして書き込む。これにより，送信元アドレスや時刻を指定した検索において，検索範囲をあるグループに限定できるとともに，グループごとに並列に検索することで高速検索が可能となる。

送信元アドレスによるグルーピングに先立って，パケットの属性情報のうち，送信先/元アドレス，送信先/元ポート番号，プロトコル番号が一致するパケット群をセッションとして抽出する。抽出されたセッションに対しては，セッション単位で属性情報をまとめて保持する。これにより，パケット単位で属性情報を保持する場合と比較して，必要とする容量を大幅に削減できる。そして，このセッション属性情報によって検索することで，例えば，通話時における端末ごとの通信セッションといった論理的な単位で高速な検索が可能となる。

むすび

本稿では，通信パケットをリアルタイムに解析しつつ，証拠としてのパケットデータを蓄積し，事後に迅速に分析できるように高速に検索する技術について述べた。

本開発技術により，従来の10倍の性能となる200 Gbpsの超高速ネットワークに流れる大量のパケットをモニタしながら，ネットワーク品質およびアプリケーション品質のリアルタイムな解析を汎用ハードウェアとソフトウェアだけで実現できる。したがって，低コストでサービス品質劣化の原因や場所の特定が可能となる。また，送信元などの履歴を残せるため，サイバー攻撃の送信元や被害を受けたサーバを容易に把握できる。更に，アプリケーションごとに必要な通信量が分かるため，必要に応じて適切なインフラを構築できる。

これらにより，ネットワーク品質の向上，データセンターの運用安定化やセキュリティ強化，トラブルの早期解決，および迅速なサービス提供などに活用できる。

今後は，通信データの蓄積と検索処理性能の更なる向上と実証実験を進めていく予定である。

参考文献

(1) 田村雅寿ほか：高速ネットワークを流れるパケットデータの蓄積・活用に向けた分散ストレージの提案. 第6回データ工学と情報マネジメントに関するフォーラム（DEIM 2014），D2-3，2014年3月.
(2) M. Tamura et al.：Distributed object storage toward storage and usage of packet data in a high-speed network. Asia Pacific Network Operations and Management Symposium（APNOMS）.

著者紹介

野村祐士（のむら　ゆうじ）
ソフトウェア研究所
運用管理ソフトウェアプロジェクト 所属
現在，ICTシステムの運用管理に関する研究開発に従事。

小沢年弘（おざわ　としひろ）
コンピュータシステム研究所
メディアサーバプロジェクト 所属
現在，次世代ストレージシステムに関する研究開発に従事。

田村雅寿（たむら　まさひさ）
コンピュータシステム研究所
メディアサーバプロジェクト 所属
現在，次世代ストレージシステムに関する研究開発に従事。