



6G/IOWN時代の高速なエンドエンド情報同期・連携技術 「In-Network Service Acceleration Platform」

In-Network Computing (INC) とは、従来端末およびクラウドが担っていた計算処理の一部をネットワークが代行し、ユーザを端末性能やサービス環境の制約から解放する技術です。本稿では、NTT 研究所が提案するINCとモバイルネットワークの融合アーキテクチャISAP (In-network Service Acceleration Platform) について解説します。

キーワード：#6G, #ネットワーク, #コンピューティング

はやし けんたろう ひらい しく
林 健太郎 / 平井 志久
 まつかわ たつや ばば ひろき
松川 達哉 / 馬場 宏基

NTTネットワークサービスシステム研究所

はじめに

ITU-R (International Telecommunication Union Radiocommunication Sector) による2030年以降のモバイルネットワークの性能目標やユースケースの勧告⁽¹⁾に代表されるように、さまざまな組織や企業が5G (第5世代移動通信システム) 以降の移動通信システム (Beyond 5G/6G) の検討を進めています。Beyond 5G/6Gの特徴として、5G以上の高速・大容量・低遅延の通信要件に加え、通信カパレッジの拡張、消費電力の低減、超多数接続、センシングなどが挙げられます⁽²⁾。これらの特性を組み合わせながら活用することで、

VR (Virtual Reality) やコネクテッドカー、遠隔手術など、多様なユースケースへの応用が期待されています。

一方で、上記のような先進的なサービスを利用するためには、ユーザ端末にも高い処理性能が必要となります。そのため、ユーザは利用するサービスごとに高価な専用端末を用意しなければならず、購入や管理の負担が強いられます。

本課題を解決できる有力な技術として、In-Network Computing (INC) が挙げられます。INCとは、従来は端末やクラウドが担っていた処理の一部を、ネットワーク内に存在するスイッチやルータなどの機器が代行する技術です。ここでは一例とし

て、ユーザが手元の端末でVRのサービスを利用する場合を考えます。従来の形態では、端末はクラウド上のサーバからVR空間の映像素材を受け取り、GPU (Graphic Processing Unit) 等を用いて映像を合成したのち、ディスプレイに表示する方式が一般的です (図1上)。これに対しINCでは、映像合成のような高度な処理を、経路内のネットワーク機器が代行して行う点が特徴です (図1下)。クラウドサーバが端末に向けて映像素材を送信すると、経路内のスイッチやルータはGPUを用いて合成やレンダリングなどの処理を行いつつ、データを伝達します。端末に必要な処理は、ネットワークから処理済みの映像を受け取り、

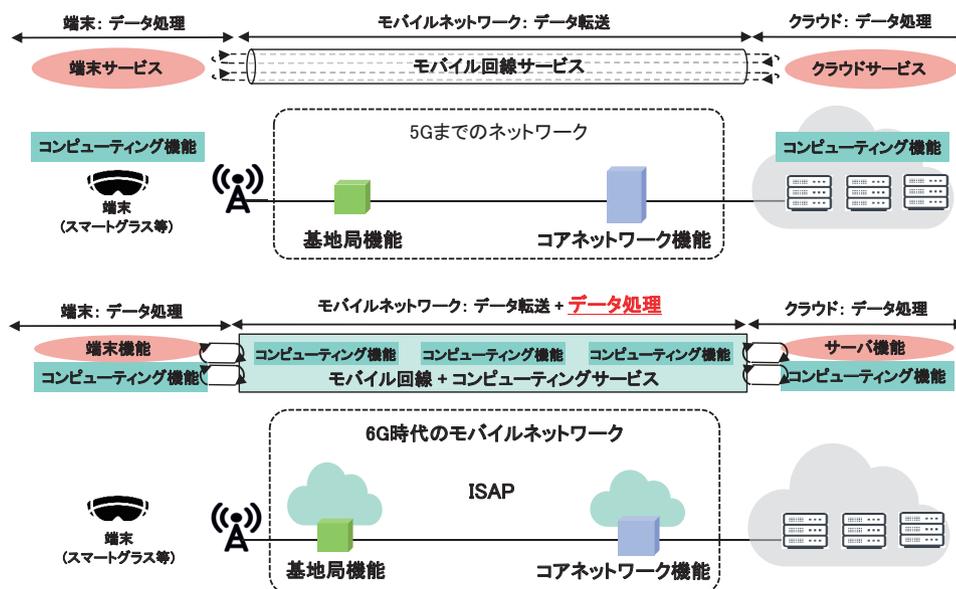


図1 従来のネットワークの機能分担 (上) とモバイルネットワークとコンピューティングの融合 (下)

ディスプレイに表示するだけとなります。結果として、高度なサービスを簡素な端末機能で提供することが可能です。

NTT 研究所では、INCとモバイルネットワークを融合した将来のネットワークアーキテクチャとして ISAP (In-network Service Acceleration Platform) の提案・実証を行っています。本稿では、ISAPの構成や特徴、ユースケースを解説します。

ISAPのアーキテクチャ

ISAPは、モバイルネットワーク上で動作するINC基盤です。ネットワークと連携してコンピューティング機能の分散処理を制御・管理することで、ネットワーク内のデータ処理・転送を高速化します。ISAPのアーキテクチャを図2に示します。以降では、ISAPの特徴を3つの観点から説明します。

■イベント駆動型リソース配備

ISAP上ではさまざまなアプリケーションの利用が想定されるとともに、各アプリケーションが必要とするリソースや処理機能も多種多様です。したがって、すべての処理機能をネットワーク内に固定的に配備する方法は非効率といえます。

そこでISAPでは、ユーザがアプリケーションを起動している期間中のみ、必要

な量のリソースを確保する方式をとります。より具体的には、ネットワークへの端末位置登録や通信セッション、ハンドオーバーなどの制御イベントのほか、アプリケーションの認証やサービス起動のイベント、サイバースペース上の行動などさまざまな情報を収集します。これらを解析し、必要な時間、場所に必要量の計算リソースをアプリケーションに割り当てます。本方式により、多様なアプリケーションや処理に対して、効率良くリソースを利用することが可能となります。

■専用ハードウェアを用いたアクセラレータ間チェイニング

ISAPが想定するコネクテッドカー、ロボティクス、VR、遠隔手術などのアプリケーションは、レンダリング、AI (人工知能) 画像解析、暗号計算をはじめとする高度な演算処理を伴います。加えて、ネットワークを介して処理の結果を端末に転送しなければなりません。したがって、アプリケーション処理と転送処理の両者を高速、低遅延に完了する必要があります。

そこでISAPでは、GPUやFPGA (Field Programmable Gate Array)、DPU (Data Processing Unit) /SmartNIC (Smart Network Interface Card) などの専用アクセラレータを活用します。より具体的には、各種アプリケーション処理や

ネットワーク接続の機能をマイクロサービス化するとともに、それぞれに適切なアクセラレータを割り当てます。一例として、AI解析や3Dレンダリング処理機能にはGPUを、GTP (General Packet Radio Service Tunneling Protocol) encaps-decap機能やRTP (Real-time Transport Protocol) ストリーミング送受信機能にはDPU/SmartNICを割り当てる形態が考えられます。さらに、これらマイクロサービス化されたアプリケーションやネットワークの機能を、ハードウェアアクセラレータチェイニングにより連鎖的に接続することで、CPU (Central Processing Unit) を介さない処理を実現します。

以上により、ISAPではネットワーク内のアプリケーション処理、ならびに処理結果の転送を高速・低遅延に行います。さらに、機械の遠隔制御信号を確定的な遅延・周期で転送するなど、アプリケーション・ネットワークの遅延やロスに起因する誤動作などを防ぐことができるので、サービスの堅牢性向上にもつながります。

■システム強靱化のための障害検知・見える化

ネットワークとコンピューティングの融合は、システム全体の複雑化をもたらします。システムの複雑化によって問題となる

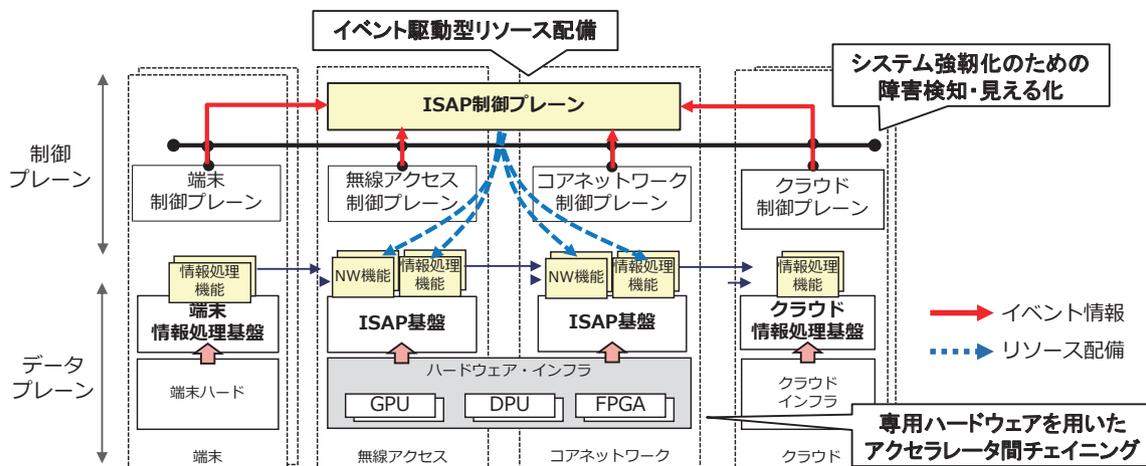


図2 ISAPのアーキテクチャ

のが、障害発生時の影響範囲の拡大です。3GPP (3rd Generation Partnership Project) でも移動体通信網のコアネットワークの強化は課題として提起され、今後の対策が議論されています。

そこでISAPは、ネットワーク内の障害や異常を検知し、迅速に緩和するための仕組みを持っています。障害箇所や異常原因の見える化、障害予兆の検知により、信号輻輳の影響拡大の抑制や、信号輻輳自体の防止により、堅牢なネットワークシステムを実現しています。詳細な情報については、ロバストネットワークに関する記事⁽³⁾を参照ください。

実装と評価

■実装

ISAPの実現可能性と効果を検証するために、複数のユースケースと併せて実装を行いました。システムの構成を図3に示します。ユースケースとして、AIによる映像ストリーム解析と、メタバースの2つを採用しました。

前述の「イベント駆動型リソース配備」については、ISAPを5Gのコアネットワークやクラウド上のアプリケーションと接続して実現しました。より具体的には、ユーザ端末の起動やネットワークへの登録、移動などの情報を収集するために、3GPPの標準インターフェース⁽⁴⁾を用いてISAPと5Gのコアネットワークを接続しました。さらにISAPとアプリケーションの間を接続し、アプリケーションの開始や状態遷移などの情報を収集可能としました。収集した情報を基に、ISAPはアプリケーションの起動を検知し、AI解析や映像レンダリングの機能をネットワークへオンデマンドに配備します。

「専用ハードウェアを用いたアクセラレータ間チェイニング」については、デバイスプラグイン機能とカスタムリソース機能を用いて実現しました。AI解析や3Dレンダリング処理などのアプリケーション機能のほか、GTPヘッダ処理やRTPストリーミング送受信などの各種ネットワーク機能をコンテナ化したうえで、各コンテナに適切なアクセラレータを仮想化して割り当てま

した。さらに、コンテナ間の連鎖により、CPUを介さない高速処理を実現しました。

以上の技術により、ユーザやアプリケーションイベントと連動し、ハードウェアリソースの柔軟な制御・割り当てが可能であることを実証しました。

■評価

(1) イベント駆動型リソース配備の効果
ISAPの第1の特徴であるイベント駆動型リソース配備について有効性を検証しました。評価では、多数のユーザが同時にさまざまなアプリケーションをランダムに利用した場合の総メモリ使用量を検証しました。各アプリケーションが要求するリソース設計は、実際のコンテナ機能をベースとして設計しました。実験から、固定量のメモリを割り当てる方式と比較して、ISAPでは効率的に計算リソースを利用できていることを確認しました。これは、個々のユーザが必要ときに計算リソースを割り当てるISAPの特徴によるものです。

(2) 専用ハードウェアを用いたアクセラレータ間チェイニングの効果
次に、アプリケーション処理機能とネッ

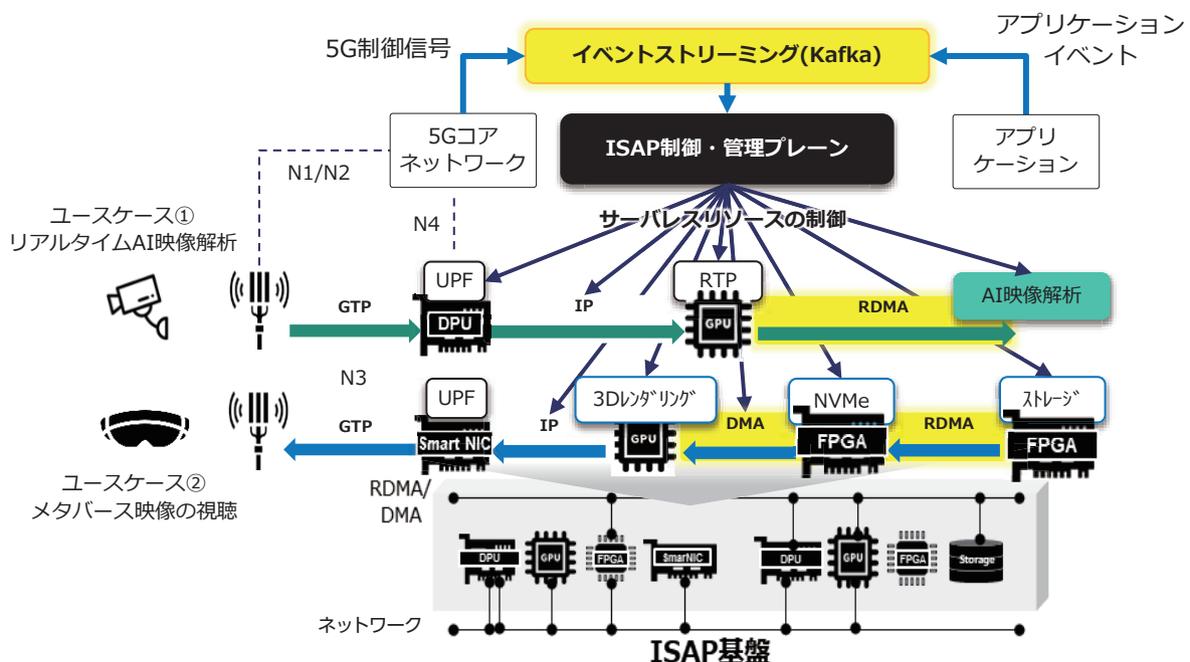


図3 実装構成

トワーク接続機能をアクセラレータにオフロードし、各機能間を連鎖的に接続する方式の効果を検証しました。アプリケーション処理機能については、図3のAI映像解析の系を対象に、ネットワーク接続性能については、図3の3D映像レンダリングにおけるUPF (User Plane Function) とトンネリング処理を対象に測定を行いました。

測定結果から、パケット転送性能では遅延で97%、ジッタで99%の改善がみられました。データ処理性能では、遅延で28%、ジッタで95%の低減を確認しました。加えて、従来方式と比較して、高性能な解析や3D映像のスムーズな再生・操作が可能であることを確認しました。

ISAP実証実験の取り組み

NTTネットワークサービスシステム研

究所では、ISAPの社会実装に向け、Beyond 5G/6G時代の多様なユースケースに適用したさまざまな実証実験に取り組んでいます。ここではその一例を紹介いたします。

■AI映像解析

昨今の生成AI技術の発展に代表されるように、AI技術はさまざまな社会課題を解決する手段として期待されています。しかし、サービス事業者がAIを組み込んだソリューションをユーザに提供するためには、一般的に「データ収集・前処理」「モデル学習・評価」「モデル展開・運用」といったフレームワークを構築する必要があります。特に、パラメータ数が大規模なモデルの学習や展開には、多数のGPUなどで構成されるAIクラスタを必要とする場合があります。

そこで、通常、高度な計算処理基盤が必

要となるAIソリューション・サービスを、ISAPで管理されるGPU/DPUリソースを活用し実証しました。この例では、前述したような、DPUとGPUが連携したデータ処理の低遅延・低ジッタ性により、4K非圧縮映像のような高いデータ処理レートが求められるユースケースにおいても、非常に滑らかで高精度なAI映像解析を実現しました(図4(a))。ISAPにより通信事業者の局舎に配備されたGPU/DPUリソースを効率的に利用することで、低遅延・低ジッタ性が求められるミッションクリティカルなサービス要件を含む、さまざまな事業者の要件に合わせて迅速かつ柔軟にAIモデルを展開することが可能です。

■メタバース

近年、3Dコンピュータグラフィックス技術やモバイルをはじめとする通信・端末技術の発展、またリモートワークの普及に

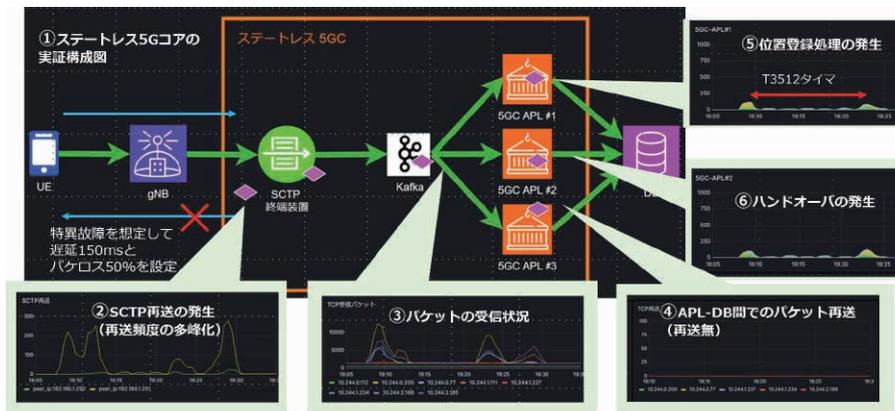
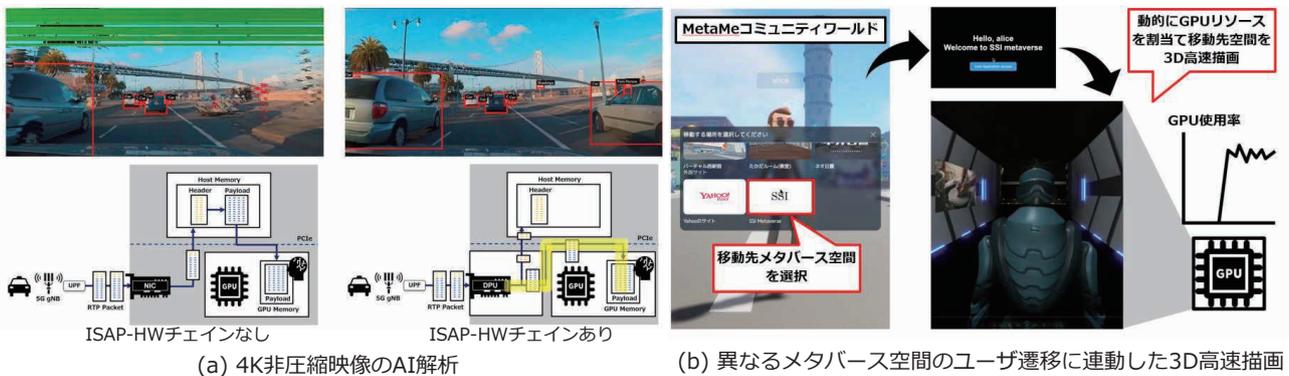


図4 実証実験

より、ユーザどうしがオンライン上でコミュニケーションを取る機会が増えたことで、インターネット上の仮想空間である「メタバース」が注目を集めています。大規模なユーザを収容するメタバースプラットフォームも登場しており、今後ますますの発展・普及が予想されます。

この例では、ユーザが複数のメタバース空間を行き来する、今後のユースケースを想定した実証を実施しました⁽⁵⁾。具体的には、さまざまなメタバース空間に移動可能なMetaMe^{®(6)}のコミュニティワールドと連携し、ユーザが異なるメタバース空間に移動する際に、当該ユーザ端末に表示される移動先の新たな空間を、ISAPを活用して動的に高速描画する機能を確認しました(図4(b))。ISAPにより、ユーザ端末の5Gアクセス等のネットワークへの接続状態や、サービス事業者により提供されるクラウド側のメタバース環境情報を連携させることで、端末スペックやアクセス・サービス環境によらない、フレキシブルなサービス体験の創出につながります。

■ネットワークスライシング

5Gでは、映像配信、自動運転やロボット遠隔制御、IoTなどの複数の異なるサービス要件を満たすために、共用化された物理ネットワークインフラから仮想的にネットワークリソースを切り出す、「ネットワークスライシング」が導入され、通信事業者のSLA監視・保証の手段として注目されています。

この例では、NTTドコモが検討を進めている、パブリッククラウド上に展開された5GC(5G Core network)と連携し、ネットワークスライシングによるエリアや時間指定、利用用途に応じた最適なエンド・ツー・エンドネットワークを提供する実証を実施しました⁽⁷⁾。ISAPによるユーザ端末の5Gネットワーク接続状態に応じたイベント駆動型リソース制御とハードウェアアクセラレーション・チェイニング機能を活用し、通常モバイル区間(端末～モバイル無線基地局～モバイルコア機能)だけでは

実現できない、アプリ・サービスドメインまで含めたエンド・ツー・エンドネットワークスライシングを実現しました。これにより、イベント会場や災害地域など通信トラフィックが集中する特定スポットに対して、オンデマンドに高品質なネットワーク・サービスの提供が可能となります。

■システム強靱化のための障害検知・見える化

5Gコアにおける障害の拡大・長時間化を抑制するための仕組みとして、障害の検知・見える化を5Gコア内部で実施し、故障の措置やリソースの制御にフィードバックする仕組みを検討しました。5Gコアをステートレスアプリケーション方式とし、障害の影響が他のアプリケーションやデータベースに波及しないようにしました。さらに、SCTP(Stream Control Transmission Protocol) 終端装置、メッセージ配信基盤(Kafka)、コントロールプレーン処理装置(5GCAPL)、共有データベースで構成される実証環境を構築しました(図4(c)の①)。この構成では、5GCAPLを停止させた場合にも、分担して処理を実施します。データベースからユーザ端末の接続状態の情報を取得することで、処理を問題なく継続できることが確認できました。

次に特異な障害の発生を想定し、通信の片方向にのみパケットの損失や遅延が極端に大きい条件を与え、位置登録処理が失敗する状況を再現しました(図4(d)の③)。位置登録処理が完了しないことで、ユーザは通信が開始できない状況となります。ネットワークレイヤにおいてはSCTPやTCP(Transmission Control Protocol)の再送が発生し、メッセージの滞留や上位レイヤの再送が生じます。このように、メッセージやパケットなどの挙動を把握することが、原因箇所の絞り込みや措置判断に有効であることを確認しました。

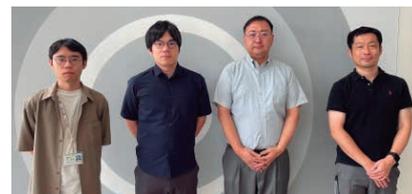
今後の展望

本稿では、インクルーシブコアの主要な

技術要素の1つであるISAPについて、アーキテクチャ、システム実装・評価、数々の実証実験を通じた技術のフィジビリティ確認について紹介しました。NTTネットワークサービスシステム研究所では、ユーザが端末やサービスに左右されず、いつでもどこでも使いたいサービスを安心・安全に利用できるような新しいサービスを実現するアーキテクチャを検討しています。今後は、ISAPに関連する細かな技術要素を国際的な標準化団体およびオープンソースコミュニティへ提案していきます。そして、6G/IOWNの本格導入が予定されている2030年に、標準仕様として社会に広く実装されることをめざします。

■参考文献

- (1) <https://www.itu.int/rec/R-REC-M.2160-0-202311-I/en>
- (2) https://www.docomo.ne.jp/corporate/technology/whitepaper_6g/
- (3) https://www.rd.ntt/ns/inclusivecore/whitepaper_ver1.html
- (4) <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3144>
- (5) <https://group.ntt/jp/newsrelease/2024/02/21/240221a.html>
- (6) <https://official.metame.ne.jp/>
- (7) https://www.docomo.ne.jp/binary/pdf/info/news_release/topics_240115_03.pdf



(左から) 林 健太郎/ 平井 志久/
松川 達哉/ 馬場 宏基

NTTネットワークサービスシステム研究所では、6GにおけるINCの実現に向けて、膨大な計算リソース等を提供可能なネットワーク・情報処理基盤を実現するアーキテクチャの研究開発を行っています。

◆問い合わせ先

NTTネットワークサービスシステム研究所
ネットワークアーキテクチャプロジェクト
アーキテクチャ方式グループ
E-mail inclusive-core@ntt.com