

vrAIIn: vRANにおけるコンピューティングリソースと無線リソースのためのディープラーニングベースのオーケストレーション

AYALA-ROMERO Jose GARCIA-SAAVEDRA Andres COSTA PEREZ Xavier

要旨

本稿では、深層強化学習に基づくvRANのリソースオーケストレータであるvrAIInを、次の順に紹介します。最初に、オートエンコーダーを使用して、高次元コンテキストデータを潜在表現を介して構築します。次に、アクタークリティックのニューラルネットワーク構造とクラシファイアに基づくDDPG（深層決定方策勾配）アルゴリズムを使用して、コンテキストをリソース制御決定にマッピングします。その結果として、(1) CPUに依存しない方法に比べて、コンピューティング容量を最大30%削減、(2) 静的ポリシーに比べて、QoSターゲットを達成する確率を25%改善、(3) コンピューティング容量のプロビジョニングが少ない場合、最先端スキームに比べて25%のスループット向上、(4) 最適なオフラインオラクルに匹敵する性能を発揮、の4つを紹介します。



O-RAN/vRAN/深層強化学習/リソース制御/CAPEX/OPEX削減

1. はじめに

vRAN（仮想化無線アクセスネットワーク）は、移動体通信事業者がモバイルサービスの増え続ける需要に低コストで対応するため、鍵となる技術として広く認識されています。vRANは、ソフトウェア化されたRAP（リモートアクセスポイント）^{*}のスタックを、クラウドのコンピューティングインフラに集中させます。この処理は、通常CPUリソースが不足することがあるエッジで行われます。図1は、信号変調やエンコードなどの無線処理タスクを実行するために、共通のCPUプールを共有するvRAPのセットを示しています（薄灰矢印）。セット化されることにより、リソースのプーリング（集中化による）、簡単なアップデート実施（ソフトウェア化による）、安価な管理と制御（コモディティ化による）などのいくつかの利点がもたらされます。

したがって、vRANが学界や産業界の注目を集めているのは、驚くべきことではありません。公表されているRANソリューションの取り組みの例としては、主要通信事業者（AT&T、Verizon、China Mobileなど）、メーカー（Intel、Cisco、NECなど）、そして主要研究機関（スタ

ンフォード大学など）によって主導されるO-RANや楽天モバイルのvRANなどがあります。これらは、汎用プロセッシングプラットフォームとデカップリングされたBBU（ベースバンド装置）とRRU（リモート無線ユニット）に基づいた、完全にプログラマブルで仮想化されたオープンなRANソリューションに向けた取り組みになっています。

そうした状況にもかかわらず、現在のvRANが実現できるメリットは最適なものからほど遠く、そしてこれが大規模な展開を妨げています。特に、リアルタイムの負荷によるピーク需要に対応するため、現在実装されているvRANのほとんどが過剰なコンピューティング能力を持っているため、計算リソースのプーリングは効率的に行われていません。

vRANにおける動的なリソース割り当てには、次のような本質的な課題があります。

- (1) vRAPの計算動作は、無線チャネルの状況やユーザーの負荷要求など、制御できない可能性のある多くの要因に依存しています。具体的には、データビットレート負荷やSNR（信号対雑音比）のパターンなどのコンテキスト、帯域幅、MIMO設定などのRAP構成、及び計算リソースをプールして

* 文献では、BS（基地局）、eNB（eノードB）、gNB（gノードB）、AP（アクセスポイント）など、いろいろな無線スタックを指すためにいろいろな名称が使用されていますが、本稿では、概念を一般化するために一貫してRAPを使用しています。

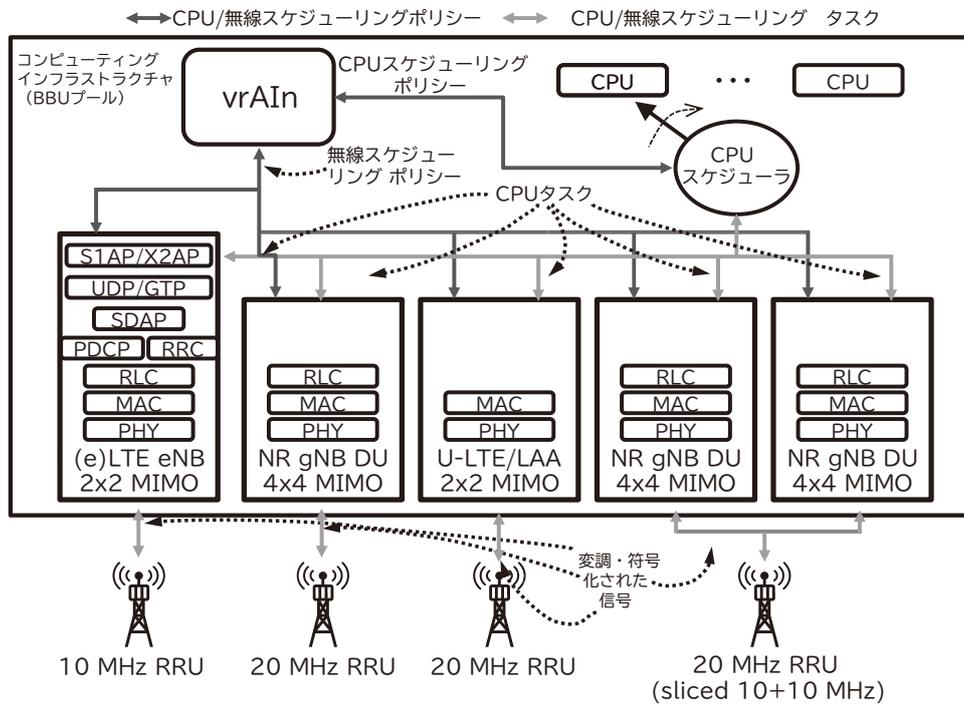


図1 vrAIInを活用したvRAN全体像

- いるインフラストラクチャに強く依存しています。
- (2) コンピューティング能力が不足した場合には、例えば、オーケストレーションの決定によりノードが一時的に過負荷になった場合、CPUの制御決定とスケジューリングやMCS（変調符号化方式）の選択などの無線制御決定はカップリングされません。

NECは、計算リソースと無線リソースの割り当てを統括する、AI制御のvRANオーケストレータであるvrAIInを提案します（図1、濃灰矢印）。

2. vrAIInのデザイン

vrAIInを活用したvRAN全体像（図1）は、フィードバック制御ループで構成されており、そこでは次のことが行われます。

- (1) コンテキスト情報（SNR及びデータ負荷パターン）の収集、エンコード
- (2) オーケストレータが、コンテキストをコンピューティング及び無線スケジューリングポリシーにマッピング

- (3) 報酬信号が意思決定を評価し、オーケストレータを微調整

こうしたことから、NECはリソース制御問題を、CB（文脈的バンドビット）問題として定式化しました。これは順次意思決定問題で、そこでは、各タイムステージ $n \in \mathbb{N}$ において、任意の特徴空間 $x^{(n)}$ から抽出されたコンテキストまたは特徴ベクトルをエージェントが観測して、アクション $a^{(n)}$ を選択し、フィードバックとして報酬信号 $r(x^{(n)}, a^{(n)})$ の受信が行われます。

コンテキスト空間：各ステージで、 T 個のコンテキストサンプルが収集されます。各サンプルには、すべてのvRAP上の全ユーザーのバッファサイズ、平均SNR、及びSNRの分散が含まれます。

アクション空間：NECのアクション空間は、CPU及び無線スケジューリングポリシーのすべてのペアから構成されます。 $c_i^{(n)} \in \mathcal{C}$ 、及び $m_i^{(n)} \in \mathcal{M}$ は、ステージ n においてvRAP _{i} に許可される最大計算時間共有（CPUスケジューリングポリシー）、及び最大MCS（無線スケジューリングポリシー）をそれぞれ示します。また、 $c_0^{(n)}$ は、コスト削減のために割り当てられていないCPUリソースの量を示します。したがっ

て、 $vRAP_i$ におけるリソース割り当てアクションは、ペア $a_i := \{c_i, m_i\}$ で構成され、システムアクションは $\mathbf{a} = (a_i)_{i \in \mathcal{P}} \in \mathcal{A} := \{(c_i \in \mathcal{C}, m_i \in \mathcal{M})\}_{i \in \mathcal{P}}$ となります。

報酬: vrAIInのデザインには次の2つの目標があります。

- (1) CPU容量が十分な場合、vRAPが所望のパフォーマンスが達成されている場合に限り、CPU使用量の面でオペレーションコストを最小化
- (2) そのようなパフォーマンスターゲットを満たすためのコンピューティング能力の不足がある場合、リソースの浪費を引き起こすデコードエラーを回避して、スループットを最大化し、遅延を最小化

この目的を達成するために、報酬関数を次のようにデザインしました。 q_{i,x_i,a_i} は、任意のスロットにおける与えられたコンテキスト x_i 及びアクション a_i に対して、 $vRAP_i$ のすべてのユーザーの集約バッファの占有状況をとらえるランダムな変数としました。QoS (クオリティオブサービス) 基準として、各 $vRAP$ の目標バッファサイズ Q_i を設定しました。この基準は、エンドユーザーが経験するレイテンシ (バッファ占有率が低いとレイテンシが小さくなる) とスループット (スループットが高いとバッファ占有率が低く保たれる) に密接に関連していることに注意してください。したがって、通信事業者は Q_i を設定することにより、ネットワークスライス間での差別化を実現するためなど、所望のQoSを選択することができます。 q_{i,x_i,a_i} が $vRAP_i$ に対して目標値以下である確率を $J_i(x_i, a_i) := \text{Prob}[q_{i,x_i,a_i} < Q_i]$ とし、報酬を次のように定義します。

$$r(\mathbf{x}, \mathbf{a}) := \sum_{i \in \mathcal{P}} J_i(x_i, a_i) - M\varepsilon_i - \lambda c_i$$

ここで、 ε_i は $vRAP_i$ のデコーディングエラー確率 (局所的に測定可能) であり、 M と λ はそれぞれ、デコーディングエラーの重み、及びリソース使用量とパフォーマンスとの間のトレードオフを決定するパラメータです。低い割

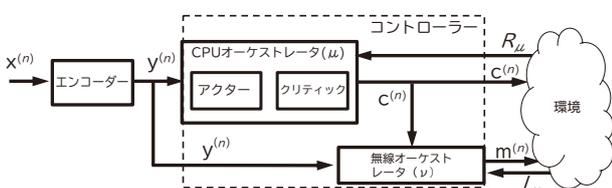


図2 vrAIInのリソースオーケストレータ

り当てによるデコーディングエラーを避けるために M を大きな値に設定し、QoS要件を満たしながら (コンピューティングリソースの使用を最小限に抑えることが目的) λ を小さな値に設定します。

こうしたCB問題を解決するために、NECは、図2に示されるシステムを設計しました。まず、SAE (スパースオートエンコーダ) を用いて、各コンテキスト $x^{(n)}$ を潜在表現にエンコードし、より低次元のベクトル $y^{(n)}$ を生み出します。次に、両方のアクションポリシーを、CPUと無線の2つのオーケストレータにデカップリングします。無線オーケストレータは、エンコードされたコンテキスト $y^{(n)}$ が与えられた場合、計算リソースポリシーを $c^{(n)}$ として高い信頼性をもって処理できる最大のMCSを提供する単純な分類器として構築されます。それとは逆に、CPUオーケストレータの設計は、図に示されるようにモデルフリーのアクタークリティックニューラルネットワーク構造を使用して、DDPG (深層決定方策勾配) アルゴリズムを用いて行われます。

3. 評価

現実世界のシナリオでvrAIInの性能を評価するために、図3(a)に示すように、197個のアクセスポイントが配置された欧州のある都市における、本番RANの配備に対してvrAIInのシミュレーションを行いました。図からわかるように、中心部 (大都市) にRAPがより密集しており、RANは周辺部で疎になっています (主に高速道路や小規模通勤圏郊外をカバー)。一般性を失うことなくトレーニングデータを活用するために、すべてのRAPにSISO 10-MHz LTE vRAPが使用され、第1章で紹介したLTE vRAPと同じ動作をするものと仮定しました。

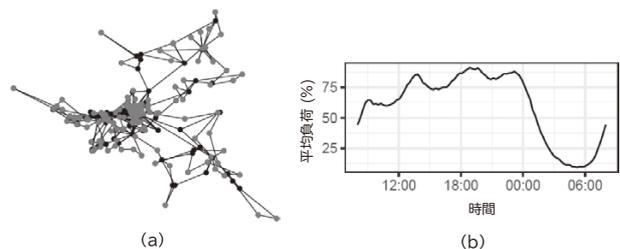


図3 (a) 欧州で運用中のRANの展開 — 灰色の点と黒い点は、それぞれ無線サイトとバックホール集約ノード、(b) 通常の平日の24時間にわたるトラフィック負荷パターン

表 シミュレーションパラメータ

vRAP帯域幅	10MHz
サブフレーム時間	1ms
送信電力	46dBm
アンテナパターン	$A_H(\phi) = -\min 12 \frac{\phi^2}{\phi_{3dB}}, A_m$ $\phi_{3dB} = 70^\circ, A_m = 25\text{dB}$
アンテナ利得	14dBm
伝搬損失	$128.1 + 37.6 \cdot \log_{10}(R[\text{Km}])$, $R: \text{vRAP} \sim \text{UE 距離}$
シャドウフェーディング	対数正規分布 10dB標準偏差
熱雑音	-176dBm
vRAP数	197
平均UE数/vRAP	10
UE可動性	20% @100km/h, 80% @3Km/h

NECがカスタムしたシミュレータは、LTEのパフォーマンス評価のための3GPPガイドラインに従っており、そのパラメータの詳細は表の通りです。UE（ユーザー装置）が感知するSINR（信号対干渉雑音電力比）は、すべてのアクティブなRAPの干渉を集約して得られます。あるSINRに対して、UEのCQI（チャネル品質指標）を計算し、3GPP仕様に従ってこのCQIに関連する最大許容MCSを計算します。更に、UEに対しては、3GPPが推奨する通りUEとRAP間の最小距離35mを確保するように、ランダムモビリティモデルを実装しています。最後に、シミュレーション用に扱いやすいモデルでvRAPのコンピューティング動作を補足することが困難なため、データセットを使用してトレーニングされたDNN（ディープニューラルネットワーク）を実装して、コンピューティングリソース不足に起因するデコーディングエラーが発生する時期を判断しました。

実際の大規模RANの動作を模倣した合成トラフィックパターンを用いて、通常の1週間のシミュレーションを行いました。分析を簡素化するため、以下では平日の24時間に焦点を当て、図3(b)に示されるトラフィックプロファイル（システムの容量に関連する）を使用します。更に、vRAN全体で集約したコンピューティング能力は、vrAIInを使用しない場合の一日の負荷ピーク時にCPU不足によるエンコーディングデッドライン違反を起こさない最低限のCPUリソース量に合わせてであると仮定しています。これを「100%プロビジョニング」と呼びます。また、システ

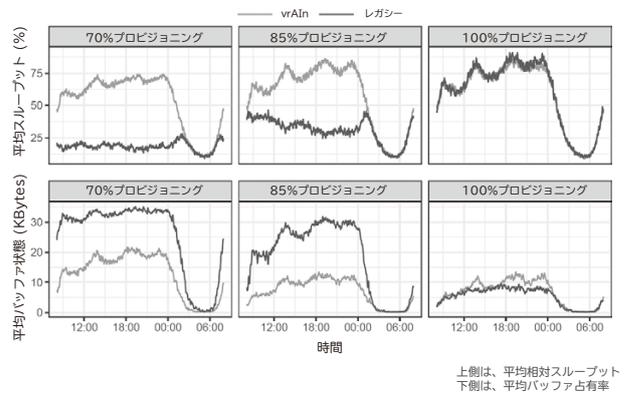


図4 ピーク負荷（「100%プロビジョニング」）に応じた計算能力と、その70%と85%分の計算能力（それぞれ「70%プロビジョニング」と「85%プロビジョニング」）についての、時間経過に伴う性能の変化

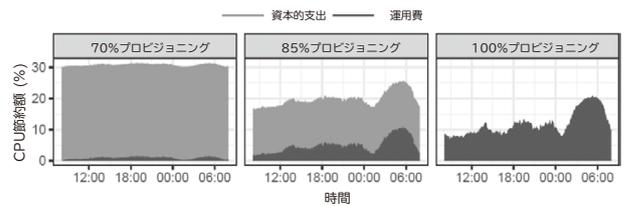


図5 ピーク需要に対して十分なCPUリソースプロビジョニングを行う場合の、計算に依存しない静的な方法に関連するvrAIInの節約額

ムがそのコンピューティング能力の70%および85%に抑えてプロビジョニングされた場合に、vrAIInの動作を調査し、資本コストの節約を可能にします。

3.1 ピーク時におけるコンピューティング能力のプロビジョニング

最初に、コンピューティング能力がピーク負荷に対応できるようにプロビジョニングされた場合のスループット性能、バッファ状態、コスト削減の変化に焦点を当ててみます（図4、図5右端のグラフ「100%プロビジョニング」）。図4から、vrAIInは従来の手法である「Legacy」とほぼ同じスループット性能を達成し、バッファサイズがわずかに大きい（最大5%）ことがわかります。これは、vrAIInがこのわずかに大きな遅延を、大幅なOPEX（運用費）の節約で相殺しているからです。これは図5（右端のグラフ「100%プロビジョニング」）で示されており、ここでは

vrAIInが10%から20%のOPEXの節約を実現しています。ただし、vrAIInがOPEXの節約よりも性能を優先するように設定された場合、遅延の差は消失することに注意してください。

3.2 コンピューティング能力のアンダープロビジョニング

次に、アンダープロビジョニング（プロビジョニングの割合を下げる）によって、積極的なCAPEX（資本的支出）削減を実現するケースについての分析を行います。ここでは、前述の区別方針についてコンピューティング能力の可用性を85%と70%へ下げてみます。これらはそれぞれ、「85%プロビジョニング」と「70%プロビジョニング」として、図4と図5に示されています。これら2つのシナリオのスループットとバッファ状態の進化を図4に示します（左端と中央のプロット）。この図から、vrAIInは、従来の計算非依存アプローチと比較して、高い性能を維持しながら積極的なCAPEX削減を実現することが明らかになります。具体的には、平均して、またピーク需要（18時～0時）の間に、コンピューティング能力をピーク負荷時の85%と70%へとアンダープロビジョニングする場合に、vrAIInは従来のスキームに比べてそれぞれ最大で50%と55%のスループット向上をもたらします。その理由は、vrAIInが無線と計算のスケジューリングポリシーを共同で最適化できる能力を持つために、計算能力不足によるエンコーディングデッドライン違反がほぼ発生しないことと遅延をトレードオフにして、vrAIInが時間軸に沿って需要によりよく対応できるようになるためです。これとは対照的に、従来の手法によるコンピューティング能力の可用性に依存しない動作では、瞬間的なピーク需要の際にエンコーディングデッドライン違反が頻繁に起こるため、大幅なスループットの低下が生じます。これにより再送信に大量の時間が浪費され、感知されるレイテンシがより大きくなるという連鎖効果が生じます。平均的には、85%及び70%のコンピューティングプロビジョニングに対して、vrAIInに比べて160%及び73%以上のバッファ占有率が同じピーク需要期間に発生することになります。

4. むすび

本稿の最後に、vrAIInは制約のあるコンピューティング環境に適応すること、つまり性能を最大化することを述べておきます。これは、「70%プロビジョニング」と「85%

プロビジョニング」の性能指標を「100%プロビジョニング」の指標と比較することによって明らかになります。特に、前述の同じピーク需要期間中、vrAIInは「85%プロビジョニング」ではスループット性能に損失がなく、平均して「70%プロビジョニング」でも最大でスループット損失が10%以下になります。これに対して、従来の手法では、同じ条件下で最大で65%から80%という相当なスループット損失が生じます。レイテンシの観点で見ると、vrAIInでは「100%プロビジョニング」のバッファ占有率に対して「85%プロビジョニング」の場合に平均バッファ占有率が1%増加するだけであり（従来の手法では282%増加）、また75%プロビジョニングの場合に平均バッファ占有率が73%増加するだけです（従来の手法では337%増加）。

- * Intelは、アメリカ合衆国および / またはその他の国における Intel Corporation またはその子会社の商標です。
- * Cisco, Cisco Systems、およびCiscoロゴは米国およびその他の国におけるCisco Systems, Inc.の商標または登録商標です。
- * O-RAN ALLIANCE、O-RANの名称とそのロゴは、O-RAN ALLIANCE e.V.の商標または登録商標です。
- * 楽天モバイルは、楽天グループ株式会社又はその関連会社の商標です。
- * その他記述された社名、製品名などは、該当する各社の商標または登録商標です。

執筆者プロフィール

AYALA-ROMERO Jose	GARCIA-SAAVEDRA Andres
NEC Laboratories Europe 6G Networks Senior Researcher	NEC Laboratories Europe 6G Networks Principal Research Scientist

COSTA PEREZ Xavier

NEC Laboratories Europe
6G Networks
Manager

NEC 技報のご案内

NEC 技報の論文をご覧いただきありがとうございます。
ご興味がありましたら、関連する他の論文もご一読ください。

NEC技報WEBサイトはこちら

NEC技報 (日本語)

NEC Technical Journal (英語)

Vol.75 No.1 オープンネットワーク技術特集

～オープンかつグリーンな社会を支えるネットワーク技術と先進ソリューション～

オープンネットワーク技術特集よせて
NECのオープンネットワークに向けた技術開発と提供ソリューション

◆ 特集論文

Open RANとそれを支える仮想化技術

Open RANがもたらすイノベーション
モバイルネットワークにおける消費エネルギー削減
自己構成型スマートサーフェス
Nuberu: 共有プラットフォームによる高信頼性のRAN仮想化
vrAln: vRANにおけるコンピューティングリソースと無線リソースのためのディープラーニングベースのオーケストレーション

5G/Beyond 5Gに向けた無線技術

グリーン社会の実現に向けたNECにおける5G/Beyond 5G基地局のエネルギー効率化技術開発
双方向トランシーバアーキテクチャを備えたミリ波ビームフォーミングICとアンテナモジュール技術
5G/6G屋内ワイヤレス通信向け1ビットアウトフェーシング変調による光ファイバ無線システム
空間分割多重を用いた28GHz帯マルチユーザー分散Massive MIMO
28GHz帯マルチユーザー分散MIMOシステムを用いたOTFS変調信号のOTA測定
Sub6GHz帯アクティブアンテナシステムにおける空間多重性能の改善
トランジスタ非線形モデルを使用しないブラックボックスドハティ増幅器の設計手法
最大8マルチユーザー多重化を実現する39GHz帯256素子ハイブリッドビームフォーミングMassive MIMO

オープンAPN (オープン光・オール光) の実現への取り組み

APN実現に向けたNECの取り組み～Openな光ネットワーク実現に向けて～
APN実現に向けたNECの取り組み～APN製品(WXシリーズ)の特長～
APN実現に向けたNECの取り組み～フィールドトライアル～
オールフォトニクスネットワークを支えるシリコンフォトニクス光源による波長変換技術
NEC Open Networksを支える光デバイス技術～800G超の光伝送技術～

コア&パリアーネットワークへの取り組み

カーボンニュートラルな社会の実現に向けたデータプレーン制御を支える技術
5G時代の人々の暮らしを支えるNECのネットワークスライシング技術
Beyond 5G、IoT、AIを活用したDX推進を支えるアプリケーションアウェアICT制御技術
通信事業者向け5Gコアネットワークにおけるパブリッククラウド活用

高度なネットワークサービスを提供する自動化・セキュア化への取り組み

OSSにおける運用完全自動化へのNECの取り組み
利用者の要件に基づくネットワークの自律運用技術とセキュリティ対応の取り組み
情報通信ネットワークの安全性を向上するセキュリティトランスペアレンシー確保技術
ネットワーク機器のサプライチェーン管理強化に向けた取り組み

ネットワーク活用ソリューションとそれを支える技術

通信事業者向け測位ソリューション
5Gのポテンシャルを最大限に引き出すトラフィック制御ソリューション(TMS)
ローカル5G向け小型一体型基地局「UNIVERGE RV1200」及びマネージドサービス
産業DXを支えるローカル5G活用によるパーティカルサービス
ローカル5G、LAN/RAN融合ソリューション

グローバル5G xHaulトランスポートソリューション

トランスポートネットワークの高度化を実現するxHaulソリューション・スイート
xHaulトランスフォーメーションサービス
xHaulトランスポート自動化ソリューション
5G/Beyond 5Gにおける固定無線トランスポート技術
Beyond 5Gに向けたSDN/自動化
高効率・大容量無線伝送を実現するOAMモード多重伝送方式

Beyond 5G/6Gに向けて

Beyond 5G時代に向けた取り組み

◆ NEC Information

2022年度C&C賞表彰式典開催



Vol.75 No.1
(2023年6月)

特集TOP