

IoT 端末を考慮したシグナリング制御による モバイルコアノードの資源利用の効率化

安達 智哉[†] 阿部 修也[†] 長谷川 剛^{††} 村田 正幸[†]

[†] 大阪大学大学院情報科学研究科 〒565-0871 大阪府吹田市山田丘 1-5

^{††} 東北大学電気通信研究所 〒980-0812 宮城県仙台市青葉区片平二丁目 1 番 1 号

E-mail: [†]{to-adachi,s-abe,murata}@ist.osaka-u.ac.jp, ^{††}hasegawa@riec.tohoku.ac.jp

あらまし モバイルネットワーク事業者は、自身が運用するモバイルコアネットワークのノード資源が枯渇しないように、収容端末台数や接続頻度に応じて資源割り当てを行う必要がある。一方、近年増加している IoT 端末は、接続される端末の台数を予測することは困難である。また、端末の通信開始時のシグナリング処理を軽減するために、RRC Connected Inactive と呼ばれる状態を導入し、端末情報をメモリに一時的に保存することが検討されている。これらのことから、今後、モバイルコアネットワークノードへの CPU やメモリ負荷が大きく変動することが予想され、効率的な資源割り当てが求められる。そこで本報告ではモバイルコアネットワークノードにおける CPU とメモリ間の負荷のオフロード手法を提案する。具体的には、ネットワークの負荷に合わせて端末の状態を制御することにより、モバイルコアネットワークノードの CPU およびメモリ負荷のバランスを調整し、収容可能な端末台数を最大化する。端末の状態制御は、端末がデータ送受信後にアイドル状態に遷移するまでの時間を制御することで実現する。提案手法により、モバイルコアネットワークノードの資源を增強することなく、収容可能な端末台数が最大で約 150% 向上することを示す。

キーワード モバイルコアネットワーク, M2M/IoT 通信, Long Term Evolution(LTE), RRC Connected Inactive

1. ま え が き

モバイルネットワーク事業者は、自身が運用するモバイルコアネットワークのノード資源が枯渇しないように、収容端末台数や接続頻度に応じて資源割り当てを行う必要がある。主なノード資源として、CPU およびメモリが挙げられる。CPU は、アタッチやデタッチ等のシグナリング処理を実行するために必要とされる資源である。一方メモリは、ベアラなどのセッション情報を保持するために必要とされる資源である。これらのノード資源は、モバイルネットワークにおける通信を実現するために必須であり、どちらか一方でも枯渇することは許されない。

一方、近年は IoT 端末の急激な増加が注目されている。IoT 端末は、スマートフォンのようなユーザ端末とは異なり、家電や自動車、電気メータ、センサなど様々な場所、様々な用途で使用される可能性があり、端末の台数およびその分布を予測することは困難である。そのため、多数の IoT 端末を収容するためにノード資源を過不足なく割り当てることは難しい。

また、IoT 端末は通信特性においてスマートフォン等の従来の端末とは異なり、データ送信に周期性や間欠性を持つという特徴がある。そのため、データの送信ごとにアイドル状態と接続状態を遷移することが予想される。その結果、端末のネットワーク接続やデータ送信に必要なシグナリングに関する通信や処理を行う、制御プレーンの輻輳が悪化すると考えら

れる。このような問題に対し、RRC Connected Inactive と呼ばれる IoT 端末の新たな状態を導入することによって、シグナリングの削減を目標とする研究が行われている [1, 2]。RRC Connected Inactive 状態とは、端末がネットワークから切り離された後も、モバイルコアネットワークでは、端末のセッション情報をメモリに保持している状態である。端末のセッション情報を破棄するアイドル状態とは異なり、モバイルコアネットワークノードのメモリ資源に負荷が発生する一方で、その端末が再び接続状態へ遷移する際に発生するシグナリングの一部を省略できるため、CPU 負荷の削減が可能となる。これにより、シグナリングを処理する際にモバイルコアネットワークノードに発生する CPU の負荷が削減され、制御プレーンの輻輳の抑制が期待できる。実際、文献 [1, 2] においては、RRC Connected Inactive を導入することにより、シグナリングオーバーヘッドの削減が可能であることが示されている。

このように、接続台数の予測が難しい IoT 端末の普及や、モバイルコアネットワークノードに与える負荷を変動させるような新たな状態の導入により、モバイルコアネットワークノードへの CPU やメモリ負荷が大きく変動することが予想される。そのため、モバイルネットワーク事業者は、これまで以上に効率的に資源割り当てを行う必要がある。

上述のような資源需要の予測が難しく、変動が激しいネットワークに対して、収容可能な端末台数の増加を目的とした既存研究には、スケールアウトの考え方をういたものが多い [3-7]。これらの研究では主に稼働するサーバやインスタンスの数を

資源の需要に応じて増減させることにより、ネットワークの変動に対応している。しかし、スケールアウトの方式で仮想マシンを増やす手法では、本来必要とされている資源量よりも多くの資源が供給される、オーバプロビジョニングが発生する問題がある。なぜなら、サーバやインスタンス一台あたりの資源構成は固定であることが一般的であり、偏った資源需要に対してスケールアウトを行った際には、本来増強する必要のない資源までプロビジョニングされるためである。文献 [4] では、IoT 端末を収容している MME を単純にスケールアウトした場合、一部の資源が不必要にプロビジョニングされる可能性があることを述べている。

また、CPU とメモリの資源を効率よく活用するために、Server Disaggregation を用いた研究も行われている [8–13]。Server Disaggregation とは、サーバの資源を分離することにより、それぞれの資源を個別に増強、更新できるシステムである。文献 [8] では、Server Disaggregation の考えをデータセンタに適用することで、CPU やメモリなどの資源を需要に合わせて自由に組み替えることが可能になり、資源の効率的な利用が可能であることを示している。しかし、Server Disaggregation は、年単位の長期的なサーバ更新のためのアーキテクチャであり、突発的な負荷変動に対応することを目的とした方式ではない。実際、文献 [13] では、時間スケールの細かい資源制御を行った場合、サーバ資源の再割り当て処理に伴う遅延やコスト面でのオーバーヘッドが大きくなると述べている。そのため、本報告で対象にするモバイルコアネットワークのように、突発的なトラフィックの増加が発生し、数分以下のオーダーで資源量の制御を行う必要がある環境においては Server Disaggregation は適していない。

そこで、本報告ではモバイルコアネットワークに特化した、柔軟かつ効率的な、CPU とメモリ間の負荷のオフロード手法を提案する。具体的には、モバイルコアネットワークの負荷に合わせて、端末の状態を制御することにより、メモリおよび CPU に与える負荷のバランスを調整させる。UE の状態の制御は、UE が最後にデータを送信したあと、Connected Inactive 状態からアイドル状態に移移するまでの時間を設定することで実現する。次に、数学的解析に基づき、提案手法の評価を行う。具体的には、端末の状態に応じて発生する、モバイルコアネットワークの負荷を数学的に導出する。そして、提案手法を用いて端末の状態を制御した場合に、収容可能な端末台数が最大約 150% 増加することを示す。

本報告の構成は以下の通りである。第 2 章では、本報告において評価対象となるモバイルコアネットワークの構成、シグナリング手順および端末の状態遷移について述べる。第 3 章では、数学的解析によって、モバイルコアネットワークに発生する CPU 負荷およびメモリ負荷を導出する。第 4 章では、評価対象となるモバイルコアネットワークのパラメータ設定を行う。第 5 章では、提案手法がモバイルコアネットワークの端末収容能力に与える影響を評価する。最後に第 6 章で本報告のまとめ、及び今後の課題について述べる。

2. モバイルコアネットワークアーキテクチャ

2.1 ネットワーク構成

2.2 シグナリング手順

2.3 UE のステート遷移

3. 解 析

3.1 CPU 負荷の算出

3.2 メモリ負荷の算出

3.3 収容可能な UE 台数の算出

4. パラメータ設定

5. 評価結果と考察

6. まとめと今後の課題

References

- [1] S. Hailu, M. Saily, and O. Tirkkonen, “RRC State Handling for 5G,” *IEEE Communications Magazine*, vol. 57, no. 1, pp. 106–113, Jan. 2019.
- [2] I. L. Da Silva, G. Mildh, M. Saily, and S. Hailu, “A Novel State Model for 5G Radio Access Networks,” in *Proceedings of 2016 IEEE International Conference on Communications Workshops (ICC)*, May 2016, pp. 632–637.
- [3] M. Shimizu, H. Nakazato, and H. Seshake, “Scale-Out Architecture for Service Order Processing Systems,” in *2013 IFIP/IEEE International Symposium on Integrated Network Management (IM 2013)*, May 2013, pp. 880–883.
- [4] P. C. Amogh, G. Veeramachaneni, A. K. Rangiseti, B. R. Tamma, and A. A. Franklin, “A Cloud Native Solution for Dynamic Auto Scaling of MME in LTE,” in *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Oct. 2017, pp. 1–7.
- [5] I. Alawe, Y. Hadjadj-Aoul, A. Ksentini, P. Bertin, and D. Darche, “On the Scalability of 5G Core Network: The AMF Case,” in *Proceedings of 2018 15th IEEE Annual Consumer Communications Networking Conference (CCNC)*, Jan. 2018, pp. 1–6.
- [6] Y. Ren, T. Phung-Duc, J. Chen, and Z. Yu, “Dynamic Auto Scaling Algorithm (DASA) for 5G Mobile Networks,” in *2016 IEEE Global Communications Conference (GLOBECOM)*, Dec. 2016, pp. 1–6.
- [7] C. H. T. Arteaga, F. Rissio, and O. M. C. Rendon, “An Adaptive Scaling Mechanism for Managing Performance Variations in Network Functions Virtualization: A Case Study in an NFV-Based EPC,” in *2017 13th International Conference on Network and Service Management (CNSM)*, Nov. 2017, pp. 1–7.
- [8] M. Mahloo, J. M. Soares, and A. Roozbeh, “Techno-Economic Framework for Cloud Infrastructure: A Cost Study of Resource Disaggregation,” in *Proceedings of 2017 Federated Conference on Computer Science and Information Systems (FedCSIS)*, Sep. 2017, pp. 733–742.
- [9] “Intel’s Disaggregated Server Rack,” Moor Insights Strategy, Technical Report (TR), Aug. 2013.
- [10] C. Devaki and L. Rainer, “Enhanced Back-off Timer Solution for GTP-C Overload Control,” Feb. 2016. [Online]. Available: <http://www.freepatentsonline.com/y2016/0057652.html>
- [11] B. Abali, R. J. Eickemeyer, H. Franke, C. Li, and M. Taubenblatt, “Disaggregated and Optically Interconnected Memory:

- When will it be cost effective?" *CoRR*, vol. abs/1503.01416, 2015. [Online]. Available: <http://arxiv.org/abs/1503.01416>
- [12] "Disaggregated Servers Drive Data Center Efficiency and Innovation," Intel Corporation, Technical Report (TR), Jun. 2017.
- [13] S. Legtchenko, H. Williams, K. Razavi, A. Donnelly, R. Black, A. Douglas, N. Cherie, D. Fryer, K. Mast, A. D. Brown, A. Klimovic, A. Slowey, and A. Rowstron, "Understanding Rack-Scale Disaggregated Storage," in *Proceedings of 9th USENIX Workshop on Hot Topics in Storage and File Systems (HotStorage 17)*. Santa Clara, CA: USENIX Association, 2017. [Online]. Available: <https://www.usenix.org/conference/hotstorage17/program/presentation/legtchenko>