



Title	A Fully-distributed Paradigm for Self-sustaining Stream Processing on Autonomous Networks of Smart Things
Author(s)	Choochootkaew, Sunyanan
Citation	大阪大学, 2020, 博士論文
Version Type	
URL	https://hdl.handle.net/11094/76651
rights	
Note	やむを得ない事由があると学位審査研究科が承認したため、全文に代えてその内容の要約を公開しています。全文のご利用をご希望の場合は、大阪大学の博士論文についてをご参照ください。

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

Abstract of Thesis

Name (SUNYANAN CHOOCHOTKAEW)	
Title	A Fully-distributed Paradigm for Self-sustaining Stream Processing on Autonomous Networks of Smart Things (自律的なスマートデバイス間ネットワークにおける完全分散型ストリーム 処理の実現に関する研究)
<p>Abstract of Thesis</p> <p>Self-sustaining Stream Processing is a promising technology for borderless autonomous intelligent systems. It carries out on-the-fly continuous processing commands on real-time streaming data from sources for activating corresponding actions via actuators in dynamic environments without barriers from maintenance cost or dependency. It also allows non-programming-savvy users to create their services in a human-understandable language. This dissertation presents the developing solutions for self-sustaining systems with the following trace.</p> <p>Firstly, the dissertation starts with introducing the still-opening challenges along with the proposed solutions on the most-popular paradigm for stream processing systems, centralized. The chosen challenges are a link bottleneck from source networks and a task-resource matching in task distribution mechanism. For the first challenge, the data prioritization system has been proposed added up from the compression and pub/sub-content filtering techniques. By modifying the existing Multi-criteria Decision Analysis technique, the proposed system can evaluate and prioritize the data based on application-specific quality metrics with a lightweight calculation at the local gateway. The results on case-study smart buildings affirm the significance of multiple criteria from higher data value preservation under a very-limited communication link constraint. For the second challenge, the proposed solution comes with two principal mechanisms: demand estimation and resource selection. The demand estimation leverages a time-series model to estimate the next arrival demands regarding the best-fit distribution of the historical records. The resource selection adopts machine-learning techniques to determine whether the resource is good enough for working on a particular task with a specified latency tolerant. The results show a reduced overhead of task-reallocation after using the demand estimation and give more than 0.8 on the F-score measurement of the matching classification.</p> <p>Secondly, the dissertation introduces the self-sustaining concept to untie the stream processing technology from the maintenance cost and dependency on the autonomous intelligent systems. The idea is to draw full power from already-in-used devices. For realizing such a concept, the processing plan must not hinge on the device-connecting topology, and the processing element must be able to cover arbitrary processing tasks. On top of that, the distribution mechanism must not rely on any particular controllers. The proposed paradigm, named EdgeCEP, combines advantages of event specifications and relational-tuple-based processing techniques and meets those requirements. EdgeCEP presents a newly-defined language and processing element, separating the detection and processing to bring in the best efficient methodology. It distributes the task with a proactive optimization-based collaboration mechanism. The objective function is to minimize the communication cost of the complex flow, subjective to computation and communication resource constraints, defined at each device. The goodness of the distribution</p>	

result needs a reasonable tradeoff of the periodic information exchange overhead. The evaluation results show significant decrements of the flow volume in simulations and real-world deployment, compared to the centralized approach and other naive policies.

Thirdly, the dissertation addresses the adoption concern of EdgeCEP from restrictions on non-dedicated devices, specifically, an over-capability requirement and identity exposure from information exchange. Also, the uncertainty of non-dedicated networks could affect the service quality focus rather than just latency on each particular task. MicroEdge comes up with a less-coupling modular architecture and less-interrupting, fair-sharing solutions for the self-sustaining stream processing systems regarding task-specific mixed value metric. The modular architecture decouples primary role services, including processors, sensors, and actors, a client-synchronizing role service, coordinator, and a middle man core service by adopting the gRPC protocol. MicroEdge performs the distribution with three fundamental mechanisms: task selection, workload distribution, and overloading detection. As a goal of distribution, the novel metric named Value-of-Service (VoS) presents in terms of task-specific values. The task selection mechanism formulates a score function to encourage the high value of the VoS. The workload distribution leverages the most trivial balancing protocol, round-robin, with multiple bias level for achieving a low rejection rate and overhead in trial-and-error of blind balanced distributing. The overloading detection method applies a queuing theory and a time-series technique to detect the overloading state in the execution queue. The results on the assuming scenario and the real-world deployment show the superiority of the proposed MicroEdge to the default policies in terms of the service value and between-tasks fairness.

The dissertation emphasizes the significance of developing self-sustaining stream processing systems. Rather than coming to replace the dedicated systems like a centralized cloud or local edge servers, it is for unbounded growth of developing applications on the networks of smart things. The proposed systems deal with the challenges of dynamic, heterogeneous, and non-dedicated participating devices and processing task variety towards the self-sustaining systems. This dissertation is the foremost research towards the freedom era of application development over smart thing networks.

論文審査の結果の要旨及び担当者

氏 名 (Sunyanan Choochootkaew)			
	(職)	氏 名	
論文審査担当者	主 査	教授	東野 輝夫
	副 査	教授	村田 正幸
	副 査	教授	渡辺 尚
	副 査	教授	長谷川 亨
	副 査	教授	松岡 茂登

論文審査の結果の要旨

自律型ストリーム処理は、インテリジェントシステムを自律的に実行するための有望な技術の一つとして注目されている。その実現には、プログラミングに精通していないユーザーが簡易な言語で記述したサービスとストリーミングデータの集合に対して、ストリーム実行の時間制約や依存関係、ネットワークの利用可能帯域の動的な変動を考慮し、アクチュエータを介して対応するアクションを自律的に連続処理する仕組みを考案する必要がある。本論文では、以下の三項目に対し、自律型ストリーム処理システム開発のための新たな手法を提案している。

最初の項目では、集中型ストリーム処理において、特にストリーム情報源からのネットワークのリンクボトルネック問題およびタスク配分におけるタスクと資源のマッチング問題を扱っている。提案手法では、ネットワーク上に配置されているローカルゲートウェイにおいて、アプリケーション固有の品質要求に基づいてストリーム処理の優先順位付けを実現する手法を提案している。また、タスクの資源需要推定とそれに基づく資源選択のメカニズムも提案している。資源需要推定を行った後のタスクの再割り当てによりオーバーヘッド削減を実現し、かつマッチング分類のF値が0.8以上となったことを示している。

第二の項目では、分散ストリーム処理において、メンテナンスコストを低減し、かつストリーム間依存関係を意識させることなくストリーム処理を分散実行するための概念を提案している。提案手法であるEdgeCEPは、イベント仕様とリレーショナルタプルベースの処理技術の利点を組み合わせた分散ストリーム処理機構である。ストリーム処理の性能確保には、オーバーヘッドとなる定期的なノード間の情報交換をできるだけ削減する必要があるため、それを考慮しながら、各ノードが協調しながらも独立かつプロアクティブにタスク分散の最適化を実現するメカニズムを設計している。シミュレーションや実環境での評価実験において、集中型アプローチやその他の単純なポリシーと比較してオーバーヘッドの大幅な減少を示している。

第三の項目では、ストリーム処理専用には提供されていないデバイスを活用する場合のネットワーク利用帯域制約やデバイスの信用度の不確実性といった課題に対処するための手法を提案し、MicroEdgeと呼ばれる新たなパラダイムを考案している。MicroEdgeは、自律性を高め、公平性を考慮した自律型ストリーム処理システム向けのアーキテクチャである。MicroEdgeでは、タスクの選択、負荷の分散、および過負荷の検出という3つの基本的な機能を有し、Value-of-Service (VoS) とよばれる新たなメトリックを提案している。提案するタスク選択メカニズムでは、VoSの高いタスクが選択されるようなスコア関数を定式化している。実環境における実験の結果、サービス価値とタスク間の公平性の観点から、MicroEdgeがデフォルトポリシーよりも優れていることを示している。

本論文ではこれらの三項目の研究を進めることで、自律型ストリーム処理システムを開発することの重要性を強調するとともに、その効率的な実現のための新たな手法を提案している。集中型クラウドやローカルエッジサーバーなどの固定の資源を用いた専用システムを開発するのではなく、モノのインターネット (Internet of Things) 上で、動的かつ異種・専用でないデバイスを用いて自律的にサービス価値の高いストリーム処理システムを開発する有効な技術となると考えられる。

よって、本論文は博士 (情報科学) の学位論文として価値のあるものと認める。