様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

# 科学研究費助成事業　　研究成果報告書

令和　５ 年　 ６ 月 ２８ 日現在

機関番号：６２６１５
研究種目：基盤研究(B)（一般）
研究期間：2020～2022
課題番号：２０Ｈ０４１６９

研究課題名（和文）振動同期を利用した分断同意手法に関する研究

研究課題名（英文）Study on Distributed Consensus by Using Synchronizing Vibration

研究代表者
佐藤　一郎（Satoh, Ichiro）

国立情報学研究所・情報社会相関研究系・教授

研究者番号：８０２８２８９６

交付決定額（研究期間全体）：（直接経費）　　13,600,000円

研究成果の概要（和文）：分散システムにおける分散合意手法として、ホタルの点滅や心筋の伸縮など自然界の同期メカニズムを利用する新手法を提案した。既存の分散同意手法はマルチキャスト通信と返信のフェーズを繰り返す一種の振動系と捉え、分散合意の過程を各コンピュータを同じ周期で通信を繰り返す振動子として扱う方法を提案した。この方法では各コンピュータには同じ周期を維持しつつ、通信タイミング（位相）は互いに重ならないよう割り当てられることで、タイミングによる衝突を低減できる。従って分散合意の実現において、通信タイミングが重複が少なく、連続的な分散合意を必要とする分散ストレージのような場合に効果が高い。

研究成果の学術的意義や社会的意義
分散システムの複数コンピュータそれぞれが共通のリズムを共有させる手法に相当して、その共有されたリズムが乱れない限りは、分散合意において起きがちな通信タイミングのズレによる衝突を低減することできる。この結果、分散合意の収束を高速化することができる。分散合意はクラウドコンピューティングを含めて、多様なシステムで利用されており、本研究の効果は広く利用が期待できる。

研究成果の概要（英文）：This work proposed a novel approach for enabling distributed consensus in distributed systems, utilizing synchronizing mechanisms in nature such as the flashing of fireflies or the contraction of heart muscles. We focus on the existing distributed consensus methods, which involve repeated phases of multicast communication and responses, as a kind of oscillating system, and propose a method that treated each computer as an oscillator that repeats communication at the same cycle. In the proposed approach, while maintaining the same cycle for each computer, communication timing (phase) is assigned so as not to overlap with each other, reducing timing collisions. Consequently, in achieving distributed consensus, there was less overlap in communication timing, making it highly effective in cases such as distributed storage that requires continuous distributed consensus.

研究分野：分散システム

キーワード：分散システム　分散合意　ミドルウェア

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９（共通）

１．研究開始当初の背景

分散システムの難しさは、通信ネットワークの遅延により、現時点の全域的な状態の把握と更新ができないことにある。一方で分散システム向けのアプリケーションは、複数コンピュータ間で一つの同じ状態を保持することが求められる。例えばDNSでは名前解決情報を複数コンピュータで保持し、さらに名前解決情報が同じであることが前提となる。また、クラウドコンピューティングでは、サーバの故障によるデータ損失を防ぐため、データの複製を複数サーバが保持するが、更新時には対象データの複製を保持するすべてのサーバにおいて更新させる必要がある。これを解決するのが分散同意であり、その機能は複数のコンピュータ間において、ある処理単位からみたとき、一つの同じ状態を保持できるようにする仕組みである。なお、分散合意は様々な分散アルゴリズム、例えば分散相互排除や全順序メッセージ配送などの基礎となっており、分散同意の改善は分散システムの様々な処理に利用されている。

しかし、既存の分散合意手法は、最も基本的となる２フェーズコミット(以降、2PC)を含めて、マルチキャスト通信とそれの返信を順番に繰り返している。このとき合意に参加する複数コンピュータがほぼ同時に新たな同意要求などを行うと、合意処理を最初からやり直す(手戻り)ことや、最悪、同意に至らない可能性もある(図１)。実際、この分散同意処理中の衝突は多発している。例えばクラウドコンピューティングは前述のように複数コンピュータがデータの複製を保持･管理するが､あるデータの更新が集中した場合､複数コンピュータが同時に更新要求を行い、前述の衝突状態に陥る。このため、分散合意における衝突を減らすことができれば、分散合意の性能は大きく向上することから、クラウドコンピューティングを含めて実システムにおいても極めて有用な進歩となる。

２．研究の目的

分散システムにおける分散合意手法として、自然現象を利用した方法を提案していく。ホタルの点滅や心筋の伸縮などの自然界において振動系が同調するメカニズムを分散システムに導入する。これは既存の分散同意手法の多くが、同意に至るまでにマルチキャスト通信とそれに対する返信というフェーズを繰り返しており、ある種の振動系と捉えられる現象となる。例えば 2PC は、①フェーズ：調整役コンピュータがマルチキャスト通信で他コンピュータに状態変更の準備を求める､②フェーズ:準備ができた他のコンピュータは高々ひとつのコンピュータに返信する。③フェーズ：調整役は他コンピュータすべてから準備完了の旨を受け取った後、マルチキャスト通信で状態変更を実際に行わせる指示を送る。ところで分散システムを構成するコンピュータは非同期に動作しているが、本研究では自然界の同期メカニズムを分散システムに導入することにより、同意に参加する各コンピュータを同じ周期で通信を繰り返す振動子として扱う。さらに各コンピュータには同じ周期を維持させつつも、位相、つまり通信タイミングは互いに重ならないように割り当てる。これによりひとつの状態に関わる分散同意を行うとき、コンピュータは振動系としての周期は同じだが、常に位相は異なる、つまり通信タイミングは重ならない。特に更新頻度が高い分散ストレージのように分散同意を連続的に繰り返す場合、長期間にわたって衝突が回避できることになる。

３．研究の方法

研究では①自然界における同期現象を実分散システム上で再現し、次に②その再現した現象を分散システムに改変していく。ただし、自然界の同期現象をそのまま分散システムに実現できるわけではない。例えば心臓の脈動は、心臓を構成する個々の心筋細胞はそれ自身の周期で伸縮しているが、周囲の心筋細胞の伸縮を感知するとそれ自身の周期及び伸縮位相を僅かにずらす。これを互いに繰り返すことにより、最終的には各心筋細胞の伸縮振動はひとつの周期に収斂する（引き込み現象）。この振動周期の引き込みメカニズムを分散システムに導入した。ここで各通信には送信元コンピュータが設定した周期の情報などの情報を含むとする（心筋や蛍は位相も同調化する）。

１．分散同意の形成メンバーとなる各コンピュータは、それ自身の周期でマルチキャスト通信を他のメンバーに送信する。これは心筋細胞の伸縮や蛍の点滅の伝搬に相当する。

２．各コンピュータは他のコンピュータによるマルチキャスト通信を受け取ると、心筋細胞や蛍と同様に、送信元のコンピュータの周期と自らの周期の比較に応じて、自らの発信周期を僅かにずらす（多くの場合、前倒しすることで、同期化される周期を短くして、分散同意のスループット向上させる）。

提案方法では上記の、１．と２．を繰り返すことにより、心筋細胞や蛍の振動同期と同様に各コンピュータの送信はあるひとつの周期に収斂していくことを確かめた。

ところで、本研究の提案手法は、分散同意における手戻りの原因となる、同意要求などの衝突を未然に回避することを目的としている。ただし、一回の分散同意のコストを減らすためではなく、多数の分散同意を繰り返す分散システムを対象としている。提案方式は通信数が増えることがあるが、実システムでは通信数そのものの増加が直ちに性能劣化につながるわけではないからである。クラウドコンピューティングを含む、分散ストレージではデータ更新が頻発することから、分散同意を連続して行われることになり、本研究の効用は高いといえる。ところで提案方式は分散同意のフェーズは逐次的な実行となるが、ストレージの入出力において逐次化されてしまうので実質の影響は少ない。また分散同意を行うべき状態が複数あるときは、各コンピュータはそれぞれの状態ごとに独立に同期を行うことを想定しているが、複数の状態をまとめて通信回数を減らすこともできる。

以上により、各コンピュータは同期した周期で分散同意に関わる通信を行うことになった。各コンピュータがこの収斂した周期とそれの位相を維持している限り、同意要求などで衝突は起きないことになる。なお、分散合意手法において衝突により、手戻りなどの影響が生じる通信は、マルチキャスト通信によるものが大半である。その場合、位相をずらす対象は当該マルチキャスト通信だけとすることで、手戻りの範囲を最小化した。

なお、提案手法は複数コンピュータを強制的に同期するわけではない。このため、同期や位相が乱れる可能性は排除できない。あるコンピュータが、同一位相スロット内で、ひとつの状態に関する複数コンピュータから同意要求を受け取った場合、送信元のコンピュータを本来の位相スロットにおける送信を継続的に行えるまで同期対象から外す、または再度同期処理に行う。逆にコンピュータが、別のコンピュータからの通信を受信すべき位相スロット内で受信がなかった場合、その別のコンピュータは同期に失敗していると扱い、同期対象から外す。システム構成の変化及び故障に対処する。また、コンピュータや通信の性能などの影響により、２つ以上の周期に収斂する可能性がある。提案方法では分散アルゴリズムにおける多数決を含む、クォーラム

(Quorum)を満足するコンピュータの同意により全体の同意の代わりをするように、全体の同意の代用可能とする方法を検討する。既存のクォーラムでは対象コンピュータだが、本研究では同期する周期によるグループ化（本提案では時間的クォーラムと呼ぶ）を導入した。提案時の予想と違い、時間的クォーラムが多数となり過半数にならず、クォーラムの同意に至らない問題が起きた。

## ４．研究成果

前述の提案手法を実際の分散システム上の汎用的ミドルウェアとして設計・実装していった。これはアプリケーションに対してはマルチキャスト通信及びその返信に関わる API として提供されるが、内部的には UDP ユニキャスト及びマルチキャストを利用して通信を実現するとともに、提案方式の周期同調及び位相配置機構を組み込み、通信を振動現象として実行することができる。同実装を通じて、提案手法の性能などを実際の分散システム上で評価した。例えば実アプリケーションとともにミドルウェア上に 2PC に基づく分散同意機構を作り、衝突が実際に減らせることを実システムの同意要求状況を再現しながら実験・確認を行った。

提案方式を前提にした新たな分散同意アルゴリズムの設計と実装を行う。また同意参加数や通信遅延などについて多様な設定とともに実験を行い、アルゴリズムの特性を実証的に調べていく。なお、Zookeeper などのオープンソースの分散データ管理ミドルウェアへの移植も検討したが、Zookeeper 自体がマスターデータ用サーバとレプリカ用サーバの処理が同期的な相互作用を前提にした実装となっているなど、提案方式とは乖離が大きいことから実施しなかった。一方で分散同意と原子ブロードキャスト（全順序マルチキャスト）は強い相関があることから、原子ブロードキャストを多用する状態機械(State Machine)については本提案の導入を試みた。

当初予定では想定していなかったが、本成果を情報学以外、例えば物理学や生物学に広げることも重要であり、特に自然科学分野における相互振動系に関わる研究は位相式などの数理モデルが基礎となっていることから、提案手法及びそれによる分散アルゴリズムの数理的なモデルの構築を行うことがありえる。いくつか準備は行ったが、その構築は将来への課題となる。

５．主な発表論文等

〔雑誌論文〕　計2件（うち査読付論文　2件／うち国際共著　0件／うちオープンアクセス　2件）

| | |
|---|---|
| １．著者名<br>Risa Kimura, Tatsuo Nakajima | ４．巻<br>1 |
| ２．論文標題<br>Collectively Sharing People's Visual and Auditory Capabilities: Exploring Opportunities and Pitfalls. | ５．発行年<br>2020年 |
| ３．雑誌名<br>SN Computer Science, | ６．最初と最後の頁<br>298-298 |
| 掲載論文のDOI（デジタルオブジェクト識別子）<br>なし | 査読の有無<br>有 |
| オープンアクセス<br>オープンアクセスとしている（また、その予定である） | 国際共著<br>－ |

| | |
|---|---|
| １．著者名<br>Risa Kimura, Tatsuo Nakajima | ４．巻<br>3(2) |
| ２．論文標題<br>A Digital Platform for Sharing Collective Human Hearing | ５．発行年<br>2022年 |
| ３．雑誌名<br>Journal of Data Intelligence | ６．最初と最後の頁<br>232, 251 |
| 掲載論文のDOI（デジタルオブジェクト識別子）<br>なし | 査読の有無<br>有 |
| オープンアクセス<br>オープンアクセスとしている（また、その予定である） | 国際共著<br>－ |

〔学会発表〕　計4件（うち招待講演　0件／うち国際学会　4件）

| |
|---|
| １．発表者名<br>Ichiro Satoh |
| ２．発表標題<br>5G-enabled Edge Computing for MapReduce-based Data Pre-processing |
| ３．学会等名<br>Fifth International Conference on Fog and Mobile Edge Computing, FMEC 2020（国際学会） |
| ４．発表年<br>2020年 |

| |
|---|
| １．発表者名<br>Ichiro Satoh |
| ２．発表標題<br>Context-Aware Information for Smart Retailers |
| ３．学会等名<br>7th International Conference, DCAI 2020（国際学会） |
| ４．発表年<br>2020年 |

| |
|---|
| １．発表者名<br>Ichiro Satoh |
| ２．発表標題<br>An Integration of Packet Routing and Data Processing in Sensor Networks. |
| ３．学会等名<br>13th International Symposium on Ambient Intelligence（国際学会） |
| ４．発表年<br>2020年 |

| |
|---|
| １．発表者名<br>Ichiro Satoh |
| ２．発表標題<br>Configurable Protocol for IoT Systems |
| ３．学会等名<br>9th International Conference on Internet of Things: Systems, Management and Security, IOTSMS 2022（国際学会） |
| ４．発表年<br>2022年 |

〔図書〕　計1件

| | |
|---|---|
| １．著者名<br>佐藤一郎 | ４．発行年<br>2021年 |
| ２．出版社<br>サイエンス社 | ５．総ページ数<br>165 |
| ３．書名<br>コンピュータのしくみ | |

〔産業財産権〕

〔その他〕

―

６．研究組織

| | 氏名<br>（ローマ字氏名）<br>（研究者番号） | 所属研究機関・部局・職<br>（機関番号） | 備考 |
|---|---|---|---|
| 研究分担者 | 中島　達夫<br>(Nakajiama Tatsuo)<br>(10251977) | 早稲田大学・理工学術院・教授<br>(32689) | |

７．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

８．本研究に関連して実施した国際共同研究の実施状況

| 共同研究相手国 | 相手方研究機関 |
| --- | --- |