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 ２１ 日現在

機関番号：１３９０４
研究種目：若手研究
研究期間：2018～2022
課題番号：１８Ｋ１８０２９

研究課題名（和文）パブリッククラウドにおける地理的分散BFTレプリケーションのレプリカ動的再配置

研究課題名（英文）Adaptive Replica Relocation of Geo-replicated State Machines on Public Cloud

研究代表者
中村　純哉（Nakamura, Junya）

豊橋技術科学大学・情報メディア基盤センター・准教授

研究者番号：６０７３９７４６

交付決定額（研究期間全体）：（直接経費）　　2,200,000円

研究成果の概要（和文）：地理的分散BFTレプリケーションは，状態機械として定義されるサービスを地理的に分散された複数のレプリカに複製する．レプリカ同士が協調して動作することにより，サービスにビザンチン故障耐性を実現できる．本研究課題では，与えられた条件下で最適なレイテンシを実現するレプリカ配置を決定する手法と，サービスの状態を遠隔のレプリカに効率的に転送する状態転送手法を開発した．これにより，レプリケーション環境の動的な変動に対応できる効率的な地理的分散BFTレプリケーションが実現できるようになった．

研究成果の学術的意義や社会的意義
地理的分散BFTレプリケーションはパブリッククラウドの発達によって容易に実現できるようになったが，レプリカ間の通信帯域の変動などの影響を受けて，その性能は常に変化する．良好なレプリケーション性能を維持するためには，定期的に最適なレプリカ配置を計算し，計算結果に基づいてレプリカを移動させる必要があるが，これらの問題は既存研究ではほとんど考慮されてこなかった．本研究では新たに2つの手法を考案して問題を解決し，地理的分散BFTレプリケーションの実用性を向上した．

研究成果の概要（英文）：Geographical BFT replication replicates a service, which is defined as a state machine, to multiple geographically distributed replicas. The replicas work in cooperation with each other and provide Byzantine fault tolerance to the service. In this research project, we developed the following two methods to enable efficient geographically distributed BFT replication that can respond to dynamic changes in the replication environment. The first method determines replica deployment that achieves optimal latency under given conditions. The second method efficiently transfers service state to remote replicas.

研究分野：分散システム

キーワード：BFTレプリケーション　ビザンチン故障　地理的分散　パブリッククラウド　ビザンチン合意

様　式　Ｃ－１９，Ｆ－１９－１，Ｚ－１９（共通）

１．研究開始当初の背景

インターネット技術の発展によって，オンラインバンキングやオンライントレーディングなど，重要で価値を持つサービスが提供されるようになった．しかしこれらのサービスは，攻撃者にとっても魅力的であることから，日々クラッカーからの攻撃にさらされている．これらの攻撃は分散システムの分野ではビザンチン故障としてモデル化される．サーバクライアントモデルで提供されるサービスにおいて，ビザンチン故障から被害を防ぐための技術として，Byzantine Fault Tolerance（BFT）[1]というレプリケーション手法がある．そこではオリジナルのサービスを複数のレプリカに複製し，すべてのレプリカが同じ処理を行う（図 1）．全レプリカが常に同じ状態を維持し続けることで，たとえ一部のレプリカが故障してもその影響をマスクし，安定したサービスの提供を続けることができる．

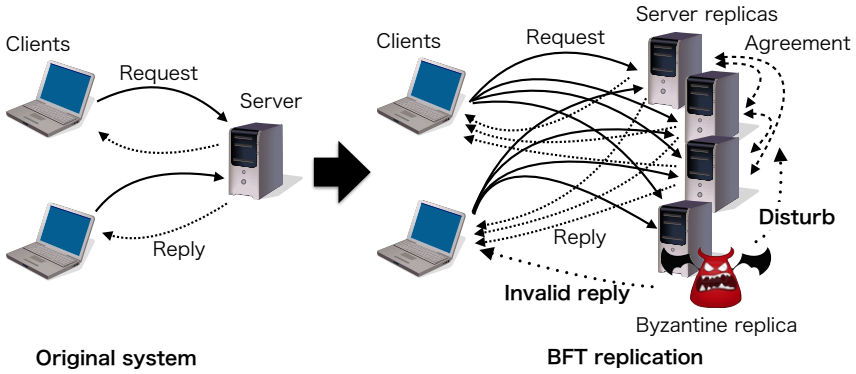

図 1. BFT レプリケーション

近年の仮想化技術の発展は，パブリッククラウドという新しい計算機の実行環境を出現させた．Amazon Web Services や Google Compute Engine に代表されるパブリッククラウドでは，ユーザの要望に応じて自由に仮想計算機や仮想ネットワークを世界中に設置し，利用することができる．これらのサービスでは，仮想計算機等の制御を外部のプログラムから自動で行うための API が用意されている．これらの機能を活用することで，従来の物理計算機やネットワークでは行うことのできなかった，新しい柔軟な計算機システムが運用できると期待されている．

複数の大陸にまたがって構成される BFT レプリケーションは，地理的分散 BFT レプリケーションと呼ばれる（図 2）．地理的分散レプリケーションは，地域ごとの負荷分散や，大地震や津波などの大規模災害に対する耐障害性など，様々な用途で用いられる．地理的分散 BFT レプリケーションは，パブリッククラウドサービスの発達によって，従来と比べて非常に低コストで実現可能になった．

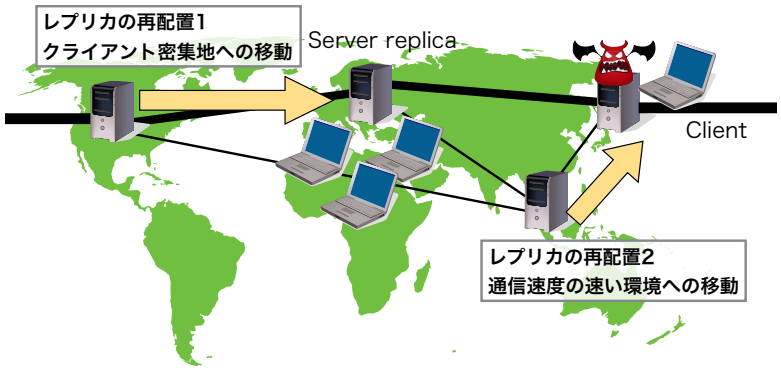

図 2. 地理的分散 BFT レプリケーション

2．研究の目的

本研究課題では，地理的分散 BFT レプリケーションの実用性向上を目的として，次の二点を明らかにする．

- 地理的分散 BFT レプリケーションとレプリカ配置の特性
- レプリカ動的再配置によるレプリケーション処理性能の向上手法

地理的に分散した BFT レプリケーションを対象とした既存研究としては，Steward [2]や，WHEAT [3]などがある．これらの既存研究は，BFT レプリケーションの実行中はレプリカの配置が変化しない，静的なレプリケーションを仮定している．一方本研究では，BFT レプリケーション実行中にレプリカ配置を積極的に変更する点に学術的独自性及び創造性がある．レプリカの動的な配置変更は，パブリッククラウドサービスの普及によって初めて可能になったもので，類似の先行研究は筆者の知る限り存在しない．

BFT レプリケーション実行中には，レプリカ間の通信帯域やクライアントの活動状況は随時変化する．図 2 では，レプリカの再配置によってレプリケーション性能が向上すると期待される二つの例を示している．一つ目（「レプリカの再配置 1」）は，クライアントが多く存在する地域にレプリカを移動させることで，クライアントとレプリカの間の通信遅延を短くする例である．二つ目（「レプリカの再配置 2」）は，通信速度が遅い地域から，通信速度が速い地域にレプリカを移動させることで，レプリカ間の通信遅延を短くする例である．このようにレプリカの配置を変更することでレプリケーション環境の変化に対応し，レプリケーション全体の性能を保ち続けることが可能になると期待される．

3．研究の方法

(1) パラメータがレプリケーション性能に与える影響

まず，パブリッククラウド上に地理的分散 BFT レプリケーションを構築し，3 つのパラメータ（レプリカ配置，クライアント配置，日中・夜間による通信帯域の変化）に着目して性能評価を実施する．これにより，本アプローチの有効性・妥当性を明らかにする．

(2) 最適なレプリカ配置の決定方法

次に，得られたレプリカ配置とクライアント配置の特性から，与えられた条件下で最適なレプリ

カ配置を決定する手法を考案する．レプリケーション環境が変化する度にこの手法によって求めたレプリカ配置に再配置することで，レプリケーション性能を維持する．

(3)　サービス状態の転送手法

最後に，レプリカをある地点から別の地点に移動させる際に必要となる，状態転送手法を設計する．地理的分散 BFT レプリケーションの特性を活用することで，レプリカの再配置にかかる時間を短縮する．

４．研究成果

(1)　パラメータがレプリケーション性能に与える影響

パブリッククラウドサービス Amazon Web Services にて，可能なすべてのレプリカ配置 (5460 種）のレイテンシを計測した結果を図 3 に示す．計測は，2018 年 3 月（Term A）と 2019 年 4 月（Term C）という二つの異なる時期に実施した．図からわかるように，13 ヶ月の期間が経つことで各レプリカ配置のレイテンシは大きく変化しており，定期的なレプリカ再配置の必要性が確認できる．

(2)　最適なレプリカ配置の決定方法

提案した最適レプリカ配置決定手法によって，可能な全てのレプリカ配置のレイテンシを見積もった結果を，図 4 に示す．レイテンシは SMR プロトコルのメッセージパターンに基づいて見積もっており，比較的高い精度を達成できていることが確認できる．

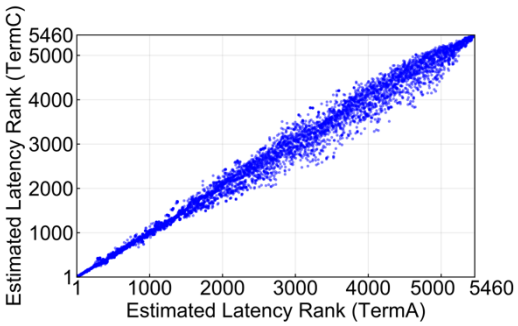

図 3．時間変化がレイテンシに与える影響

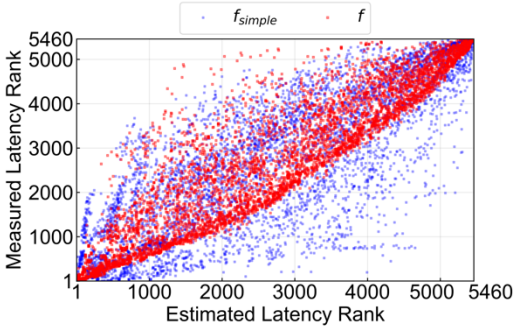

図 4．見積もりレイテンシ

(3)　サービス状態の転送手法

図 5 に提案した状態転送手法の概要を示す．サービス状態を複数のチャンクに分割し，各レプリカが並行してチャンクを転送することで，状態転送にかかる時間を短縮する．

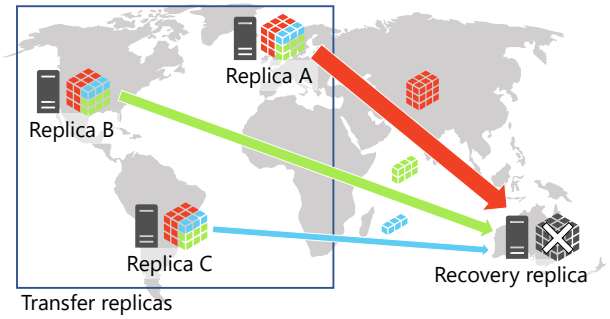

図 5．提案状態転送手法の概要

図 6 に，既存手法と提案手法の状態転送時間を示す．Amazon Web Services 上に設置されたレ

プリカ 4 台（シドニー，サンパウロ，バージニア北部，アイルランド）という条件において，転送時間を最大で 47%短縮できることを確認した．

レプリケーション中にレプリカ 1 台を移動して再配置した際のレイテンシの変化を，図 7 に示す．レプリカ移動前のレイテンシは約 340 ms だったが，移動後は約 210 ms に改善した．1GB のサービス状態の転送が 30 秒程度で完了しており，転送中もレイテンシは悪化していない．以上から，レプリカ再配置における提案手法の有効性を確認できる．

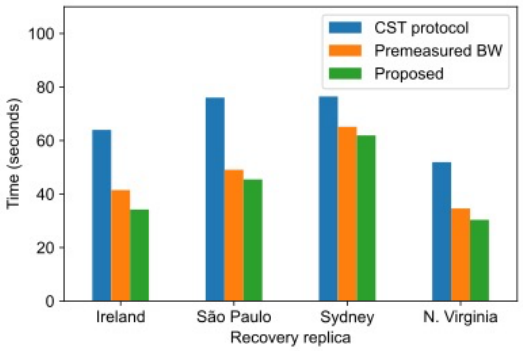

図 6．提案手法の状態転送時間

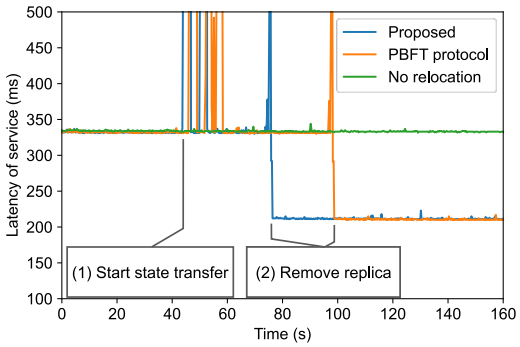

図 7．レプリカ再配置の効果

参考文献

1. M. Castro and B. Liskov. Practical byzantine fault tolerance and proactive recovery. ACMTransactions on Computer Systems, 20(4):398–461, 2002.
2. Y. Amir, et al. Steward: Scaling byzantine fault-tolerant replication to wide area networks. IEEE Transactions on Dependable and Secure Computing, 7(1):80–93, 2010.
3. J. Sousa and A. Bessani. Separating the wheat from the chaff: An empirical design for geo-replicated state machines. In Proceedings of the 2015 IEEE 34th Symposium on Reliable Distributed Systems, SRDS ’15, pages 146–155, 2015.

５．主な発表論文等

〔雑誌論文〕　計5件（うち査読付論文　5件／うち国際共著　0件／うちオープンアクセス　4件）

１．著者名 Kim Yonghwan, Shibata Masahiro, Sudo Yuichi, Nakamura Junya, Katayama Yoshiaki, Masuzawa Toshimitsu	４．巻 874
２．論文標題 A self-stabilizing algorithm for constructing a minimal reachable directed acyclic graph with two senders and two targets	５．発行年 2021年
３．雑誌名 Theoretical Computer Science	６．最初と最後の頁 1-14
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.tcs.2021.05.005	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 Hirose Jion, Nakamura Junya, Ooshita Fukuhito, Inoue Michiko	４．巻 E105.D
２．論文標題 Weakly Byzantine Gathering with a Strong Team	５．発行年 2022年
３．雑誌名 IEICE Transactions on Information and Systems	６．最初と最後の頁 541-555
掲載論文のDOI（デジタルオブジェクト識別子） 10.1587/transinf.2021FCP0011	査読の有無 有
オープンアクセス 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 －

１．著者名 Masahiro Shibata, Masaki Oyabu, Yuichi Sudo, Junya Nakamura, Yonghwan Kim, Yoshiaki Katayama	４．巻 12(1)
２．論文標題 Visibility-optimal gathering of seven autonomous mobile robots on triangular grids	５．発行年 2022年
３．雑誌名 International Journal of Networking and Computing	６．最初と最後の頁 2-25
掲載論文のDOI（デジタルオブジェクト識別子） 10.15803/ijnc.12.1_2	査読の有無 有
オープンアクセス 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 －

１．著者名 Nakamura Junya, Kim Yonghwan, Katayama Yoshiaki, Masuzawa Toshimitsu	４．巻 33
２．論文標題 A cooperative partial snapshot algorithm for checkpoint‐rollback recovery of large‐scale and dynamic distributed systems and experimental evaluations	５．発行年 2020年
３．雑誌名 Concurrency and Computation: Practice and Experience	６．最初と最後の頁 －
掲載論文のDOI（デジタルオブジェクト識別子） 10.1002/cpe.5647	査読の有無 有
オープンアクセス 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 －

１．著者名 Chiba Tairi, Ohmura Ren, Nakamura Junya	４．巻 －
２．論文標題 Network bandwidth variation‐adapted state transfer for geo‐replicated state machines and its application to dynamic replica replacement	５．発行年 2022年
３．雑誌名 Concurrency and Computation: Practice and Experience	６．最初と最後の頁 －
掲載論文のDOI（デジタルオブジェクト識別子） 10.1002/cpe.7408	査読の有無 有
オープンアクセス 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 －

〔学会発表〕　計23件（うち招待講演　0件／うち国際学会　10件）

１．発表者名 Tairi Chiba, Ren Ohmura, Junya Nakamura
２．発表標題 A State Transfer Method That Adapts to Network Bandwidth Variations in Geographic State Machine Replication
３．学会等名 The 9th International Symposium on Computing and Networking (CANDAR)（国際学会）
４．発表年 2021年

１．発表者名 Jion Hirose, Junya Nakamura, Fukuhito Ooshita, Michiko Inoue
２．発表標題 Gathering with a strong team in weakly Byzantine environments
３．学会等名 The 22nd International Conference on Distributed Computing and Networking (ICDCN 2021)（国際学会）
４．発表年 2021年

１．発表者名 Masahiro Shibata, Yuichi Sudo, Junya Nakamura, Yonghwan Kim
２．発表標題 Partial gathering of mobile agents in dynamic rings
３．学会等名 The 23rd international conference on Stabilization, Safety, and Security of distributed systems (SSS)（国際学会）
４．発表年 2021年

１．発表者名
千葉 泰理，大村 廉，中村 純哉

２．発表標題
広域State Machine Replicationの通信帯域変化に適応する状態転送手法

３．学会等名
第17回情報科学ワークショップ

４．発表年
2021年

１．発表者名
廣瀬 慈恩，中村 純哉，大下 福仁，井上 美智子

２．発表標題
故障数が線形な弱ビザンチン環境におけるモバイルエージェント集合アルゴリズム

３．学会等名
第17回情報科学ワークショップ

４．発表年
2021年

１．発表者名
千葉 泰理，大村 廉，中村 純哉

２．発表標題
広域State Machine Replicationの通信帯域変化に適応する状態転送手法

３．学会等名
第2回東海ユビキタスコンピューティング研究室合同研究発表会

４．発表年
2021年

１．発表者名
千葉 泰理，中村 純哉，大村 廉

２．発表標題
通信帯域に基づく状態分割を用いた広域State Machine Replicationにおける状態転送手法

３．学会等名
マルチメディア，分散協調とモバイルシンポジウム2020

４．発表年
2020年

１．発表者名
斎田 誠宏，金 鎔煥，中村 純哉，片山 喜章

２．発表標題
論理時計を用いた通信効率の良いCheckpoint-Rollbackアルゴリズムに関する考察

３．学会等名
第16回情報科学ワークショップ 予稿集

４．発表年
2020年

１．発表者名
沼倉 正太，中村 純哉，大村 廉

２．発表標題
RTTとState Machine Replicationの通信パターンに基づく広域SMRの応答時間最適なレプリカ配置決定手法とその応用

３．学会等名
情報処理学会研究報告 2020-DPS-182(18) 1 - 8

４．発表年
2020年

１．発表者名
千葉 泰理，中村 純哉，大村 廉

２．発表標題
広域State Machine Replicationに適した通信帯域に基づく状態転送手法

３．学会等名
情報処理学会第82回全国大会講演論文集 2020(3) 117 - 118

４．発表年
2020年

１．発表者名
沼倉 正太，中村 純哉，大村 廉

２．発表標題
広域State Machine Replicationにおけるレプリカ配置の評価とランキング

３．学会等名
第15回情報科学ワークショップ 予稿集 201 - 206

４．発表年
2019年

１．発表者名
沼倉 正太，中村 純哉，大村 廉

２．発表標題
パブリッククラウド上の広域State Machine Replicationのためのリージョン選択手法の提案

３．学会等名
信学技報 119(148) 241 - 246

４．発表年
2019年

１．発表者名
沼倉 正太，中村 純哉，大村 廉

２．発表標題
パブリッククラウド上の広域State Machine Replicationのためのリージョン選択手法の提案

３．学会等名
信学技報，vol．118，no．166，DC2018-14，pp．7-12

４．発表年
2018年

１．発表者名
沼倉 正太，中村 純哉，大村 廉

２．発表標題
RTTと通信パターンに基づく広域SMRにおけるレプリカ配置の決定手法の提案

３．学会等名
第14回情報科学ワークショップ 予稿集，pp．252-257

４．発表年
2018年

１．発表者名
沼倉 正太，中村 純哉，大村 廉

２．発表標題
RTTとSMRの通信パターンに基づく広域SMRにおけるレプリカ配置の決定手法の提案

３．学会等名
ユビキタス・ウェアラブルワークショップ2018予稿集，p．14

４．発表年
2018年

１．発表者名
Nakamura Junya, Shibata Masahiro, Sudo Yuichi, Kim Yonghwan

２．発表標題
Self-Stabilizing Construction of a Minimal Weakly ST-Reachable Directed Acyclic Graph

３．学会等名
Proceedings of the 39th International Symposium on Reliable Distributed Systems (SRDS)（国際学会）

４．発表年
2020年

１．発表者名
Thazin Nwe, Tin Tin Yee, Ei Chaw Htoon, Junya Nakamura

２．発表標題
A Consistent Replica Selection Approach for Distributed Key-Value Storage System

３．学会等名
Proceedings of the 3rd International Conference on Advanced Information Technologies (ICAIT)（国際学会）

４．発表年
2019年

１．発表者名
Shota Numakura, Junya Nakamura, Ren Ohmura

２．発表標題
Evaluation and Ranking of Replica Deployments in Geographic State Machine Replication

３．学会等名
Proceedings of the 38th International Symposium on Reliable Distributed Systems Workshops (SRDSW)（国際学会）

４．発表年
2019年

１．発表者名
Junya Nakamura, Masahiro Shibata, Yuichi Sudo, Yonghwan Kim

２．発表標題
Brief Announcement: Self-stabilizing Construction of a Minimal Weakly ST-Reachable Directed Acyclic Graph

３．学会等名
Proceedings of the 21st international conference on Stabilization, Safety, and Security of distributed systems (SSS)（国際学会）

４．発表年
2019年

１．発表者名
Yonghwan Kim, Masahiro Shibata, Yuichi Sudo, Junya Nakamura, Yoshiaki Katayama, Toshimitsu Masuzawa

２．発表標題
Improved-Zigzag: An Improved Local-Information-Based Self-optimizing Routing Algorithm in Virtual Grid Networks

３．学会等名
Proceedings of the 21st international conference on Stabilization, Safety, and Security of distributed systems (SSS)（国際学会）

４．発表年
2019年

１．発表者名
Yonghwan Kim, Masahiro Shibata, Yuichi Sudo, Junya Nakamura, Yoshiaki Katayama, Toshimitsu Masuzawa

２．発表標題
A Self-Stabilizing Algorithm for Constructing an ST-Reachable Directed Acyclic Graph When |S|<=2 and |T|<=2

３．学会等名
Proceedings of the 39th IEEE International Conference on Distributed Computing Systems (ICDCS)（国際学会）

４．発表年
2019年

１．発表者名
Kim Yonghwan, Nakamura Junya, Katayama Yoshiaki, Masuzawa Toshimitsu

２．発表標題
A Cooperative Partial Snapshot Algorithm for Checkpoint-Rollback Recovery of Large-Scale and Dynamic Distributed Systems

３．学会等名
Proceedings of the 6th International Symposium on Computing and Networking Workshops (CANDARW)（国際学会）

４．発表年
2018年

１．発表者名
塩崎 功也, 中村 純哉

２．発表標題
通信パターンに基づく応答時間最適なGeographical SMRプロトコルとレプリカ配置の選択手法

３．学会等名
情報処理学会 第85回全国大会

４．発表年
2023年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

―

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考

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			
ミャンマー	University of Information Technology			