

コーディネータにおける動的部分再構成を用いた 匿名化ハードウェアの提案

澤口 聡太[†] 西 宏章[†]

[†]慶應義塾大学理工学部システムデザイン工学科 〒223-8522 神奈川県横浜市港区日吉 3-14-1

E-mail: [†] {sawaguchi, west}@west.sd.keio.ac.jp

あらまし 近年、クラウドやエッジ、フォグといったコンピューティング環境が導入・提案されている一方で、ネットワークにおけるアプリケーションが多様化している。アプリケーションによって許容遅延や要求セキュリティレベル、地域性などが大きく異なるため、その都度適切なコンピューティング環境の選択が重要となる。本研究では匿名化処理に焦点を当て、動的部分再構成による複数の匿名化処理に対応可能な中間処理ノードの実装と提案を行い、回路規模削減、低消費電力化、高スループット化の可能性を示す。

キーワード 動的部分再構成, 匿名化, コーディネータ, 中間処理ノード

Hardware Accelerator for Data Anonymization Using Dynamic Partial Reconfiguration of Coordinator

Sota SAWAGUCHI[†] and Hiroaki NISHI[†]

[†] Department of System Design Engineering 3-14-1 Hiyoshi, Yokohama-shi Kanagawa, 223-8522 Japan

E-mail: [†] {sawaguchi, west}@west.sd.keio.ac.jp

Abstract Recently, cloud, edge, and fog computing have been introduced or proposed, whilst several different kinds of applications exist in the network. All the parameters such as delay envelope, security requirement, locality and so on vary from application to application, so that it is of the essence to choose a proper computing resource every time each application needs to be dealt with. In this research, we have focused on anonymization, implemented an intermediate node adapting itself to different kinds of anonymization methods using dynamic partial reconfiguration, and attempted to prove the possibility that less area usage, less energy consumption, and high throughput could be achieved.

Keywords Dynamic Partial Reconfiguration, Anonymization, Coordinator, Intermediate node

1. はじめに

ネットワーク上には多様なサービスとそれに伴う多様なアプリケーションが存在する。機械学習や電力モニタリングに始まるデータの解析、データベース処理などのデータの保存、匿名化や暗号化といったデータの加工、そしてスマートハウスにおける家電機器の制御などに見られるデバイスへの指令がアプリケーション例として挙げられる。これらのアプリケーションは、適用対象となるサービスによって、その許容遅延や要求セキュリティレベル、地域性が大きく異なる。一方でクラウドやエッジ、フォグやセンサネットワークといった様々なコンピューティング環境が導入・提案されており、それぞれにおける処理・通信遅延、セキュリティレベル、対象地域などは互いに異なる(図

1)。そのため、アプリケーション毎に適したコンピューティング環境を選択することが重要となる。

Internet of Things(IoT)が近年注目されており、経済産業省によれば、2020年までに250億もの端末がインターネットに接続されると推計されている[1]。また、エッジコンピューティングやスマートシティ、M2Mといった流れが存在し、多くの中間処理ノードが配置されると考えられる。こうしたIoT、エッジコンピューティング、スマートシティといった取り組みが広がるにつれ、センサノードや、その集めた情報を処理するノードが増加し、これらの機器による消費電力増加が予想される。今後サービスの多様性に対応するにあたり、その多様性に見合う様々なセンサや処理ノードが必要となり、機器導入数の削減は望みにくい。

さらに、IoTの促進により、今後より多くの種類のデータが取得でき、データ量が増大すると考えられる。すべてのアプリケーションの要求に対応するためには、これらすべてのデータ処理を許容遅延以内に収める必要がある。そのため、各処理ノードにおける高スループット化も要求される。

	対象地域	許容遅延	処理粒度	情報量	情報粒度	匿名性	スループット
上位階層(クラウド等)	広い	大きい	粗い	多い	大きい	高い	10Gbps~
中間階層(中継器等)							1Gbps~10Gbps
下位階層(センサネットワーク等)	狭い	小さい	細かい	少ない	小さい	低い	数Mbps~Gbps

図 1 階層深さによる特性の違い

そこで本研究では、低消費電力化を実現するために Field Programmable Gate Array (FPGA) を使用し、動的再構成（以下再構成）による論理回路の時分割多重化を用いることで、複数のアプリケーションに対応可能なコーデネータを提案する。再構成により複数の処理機構を実装できるため、多能性を有しつつ回路面積削減を図り、低消費電力化も期待できる。さらに、専用ハードウェアによる高スループット化も期待できる。

アプリケーションの一例として、本研究では匿名化処理を実装した。今後 IoT や企業による事業のクラウド化などの促進に伴い、非常に多くの種類・量のデータがクラウドやエッジ、フォグ等で処理されることが予想される。こうしたデータには企業の機密情報や個人情報が多く含まれているため、データの二次利用や第三者機関へのデータ公開においてプライバシー保護を考慮する必要がある。こうしたプライバシー保護を行う際に用いられる技術の1つが、匿名化技術である。匿名化技術は処理対象のデータの特徴によって適した処理手法が異なる。例えば、IP アドレスには大域的再符号化や局所再符号化などの一般化処理、温度データであれば、マイクロ・アグリゲーションやノイズ付加などが適していると言える。そのため、サービスやアプリケーション、データの種類の匿名化処理を適宜選択することが求められる。以上の理由から、本研究の提案機構では複数存在する匿名化処理モジュールを、再構成を用いて実装し、提案する。

本研究報告の構成は次のとおりである。2章では、本研究の関連研究を述べ、3章では、本研究で実装した提案機構について述べる。具体的には、提案機構延滞の概念と再構成領域における再構成モジュールの動作について説明する。4章では、提案機構における再構成にかかる時間と消費電力、回路規模、スループットそれぞれについて、Raspberry Pi 2 Model B のソフトウェア処理との比較と評価を行った。5章では結論、6章では今後の課題を述べる。

2. 関連研究

再構成を用いた低消費電力かつ高スループットであるシステムの構築に関して多くの研究がなされている。例えば、論文 [2] では、低エネルギー指向型 SQL クエリ処理の高速化を行うため、FPGA の専用ハードウェアを利用しつつ、再構成を用いて複数のクエリ処理にオンザフライで対応するアーキテクチャの実装と提案を行っている。データセンタでの利用を想定し、サーバのソフトウェア処理と比較した結果、電力効率が提案機構により改善されている。論文 [3] では、SPREAD というストリーミングベースの再構成アーキテクチャとソフトウェア・ハードウェアのマルチスレ

ッドプログラミングモデルを構築することにより、ストリーミングアプリケーションにおけるハードウェア効率を向上し、低消費電力化・高スループット化を図っている。アプリケーション例として AES, DES, 3DES の3つの暗号化処理を挙げており、従来の GPU と比較して 1.61-4.59 倍の電力効率を示している。しかし、スループットの観点で見ると、GPU と比較して少なくとも 7 倍低く数百 Mbps 程度のスループットであり、使用される環境が限定されることが考えられる。本研究報告では、想定する使用環境と算出したスループットから、スループットが必要充分であるかを定量的に評価する。また論文 [4] では、1920x1080 の静止画処理におけるパイプライン処理機構を、再構成を用いて実装することで、回路コストの削減と高スループット化を提案している。また、再構成にはコンフィギュレーションエンジンを構築し、Zynq に搭載されているコンフィギュレーションポートと比較して、再構成の高速化を実現している。これにより高スループット化が実現されている上、再構成により回路規模が削減されている。以上のように、再構成を用いたリコンフィギュラブルなシステムを構築することにより、回路規模の削減による低消費電力化を図りつつ、高スループットを達成することが可能であると言える。

匿名化に関する研究も数多く行われている。例えば Ubik らは論文 [5] において、FPGA を用いて、ネットワークを流れる実際のパケットデータに含まれる IP アドレスのツリー構造を生成し、そのノードを入れ替えるスワッピングによって、匿名化処理高速化を提案している。この手法ではツリー構造を事前処理で生成する必要があり、ネットワーク上での運用では柔軟性が失われる。一方、事前処理が不要で FPGA に実装可能なスワッピングによる IP アドレスの匿名化機構が Blake らにより提案されている [6]。しかし、この提案手法では IP アドレスの匿名化のみで汎用性が無く、様々なデータの匿名化が求められる環境には適さない。澤田らは、TCAM とキャッシュ機構を用いたマスキングによる匿名化機構を提案し、匿名化の高速化と情報損失率の低減を図っている [7]。また山口らは、RAM ベースの機構に依る回路規模削減とそれに伴う高速なマスキング処理に基づく匿名化処理機構を提案している [8]。両手法ともに情報損失率に加え、k-匿名性と l-多様性を満たした手法であり、後者はスループットが 10Gbps を超える。しかし、匿名化処理手法としてはマスキング処理のみで、処理の多様性に欠ける。

ネットワーク上では様々なデータ通信が利用され、時間帯によりデータの種類や量が変化する。そこを流れるデータの匿名化処理を行う場合、事前処理を行うことは困難であり、また、データはそれぞれ異なる特

徴を持つため、それぞれのデータの特徴に合わせた匿名化を行う必要がある。つまり、データの特徴に合わせて適切な匿名化処理を動的に選択することが重要だと言える。しかし、これまで述べたように、複数の匿名化処理を動的に選択して実行するアーキテクチャを再構成により実装した研究例は筆者の調べた限り存在しない。そこで本研究では、複数の匿名化処理を、再構成を用いて実装し、提案する。

匿名化処理の種類であるが、数値データに適した攪乱的手法と、文字列データに適した非攪乱的手法が存在する。前者では、ノイズ付加、スワッピング、マイクロ・アグリゲーション、端数処理、リサンプリング、後者では、サンプリング、大域的再符号化や局所再符号化(マスキング処理など)が挙げられる。本研究では、ノイズ付加、スワッピング、マイクロ・アグリゲーション、大域的再符号化の4つの匿名化処理を再構成モジュールとして、再構成により実装した。

3. 提案機構

本研究での提案機構を図2に示す。実装環境として、Zynq ZC702 評価ボードを用いている。再構成モジュールとして、マスキング、ノイズ付加、マイクロ・アグリゲーション、スワッピングの4つの匿名化処理モジュールを実装した。これにより、動作中にそれぞれの匿名化処理から必要な処理モジュールを選択し、再構成を行うことが可能となる。今回の処理フローは次の通りである。まず、DDR から AXI DMA に 32bit データを 32 個バースト転送し、それを Reconfigurable Anonymization Module にストリーミングする。処理結果は DMA を介して、再び DDR に書き込まれる。

4つの処理モジュールの動作について説明する。マスキング処理では、処理機構としては、指定された値だけ入力データの下位ビットを0によりAND演算し、マスキング処理を行うことができる。しかし、本研究では下位8ビットに固定してマスキングを行っている。1つのデータのマスキングには、1クロックを要する。

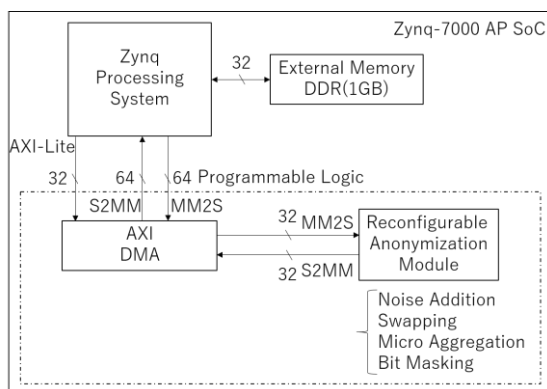


図2 提案機構

表1 ノイズ付加における各値の例

元データ	00100111
マスク値	00011111
乱数	10110110
ノイズ値(乱数&マスク値)	00010110

ノイズ付加では、入力データに対して、入力データ以下の任意の値を付加する処理を実装した。まず32個のデータエントリをバッファに格納したのち、当該データを2進数としてとらえ、最上位ビットから見て初めて1が出現する位置をバレルシフトにより算出する。この処理を32個並列で行い、算出された位置情報をもとに32個のマスク値を作成し乱数とのAND演算を行うことで、入力データ以下のノイズ値を生成する。表1に8ビットのノイズ値の生成の例を示す。マイクロ・アグリゲーションに関して、ストリーミングで入力される32個のデータのうち、値の近い8個のデータの累積和を算出し、3ビット右シフトにより平均値を返す処理とした。値の近いデータをストリームで得るため、FIFOベースのマージソートを用いて昇順に整列した。

最後にスワッピング処理に関して、論文[5]のスワッピング手法を実装した。図3は3ビットで表現できる8つの連続する値のスワッピングの概念図である。8つの連続する値を木構造の末端に対応させ、スワッピングを行う木のノードを1、行わないノードを0で指定する。図3では、前者を黒色、後者を白色に対応させている。木構造の最上位ノードから末端のそれぞれの値へのパスを通る際の各ノードの値を組み合わせた3桁の値を、その末端の値とXOR演算することで、スワッピングが完了する。このアルゴリズムを利用して、32個のデータをランダムに入れ替える処理を実装した。32個のアドレス値をスワッピングする際、木構造を組んだ時の31個のノードにおける値(0もしくは1)を決定する必要があるが、本研究ではXorshiftにより生成した32ビット乱数の下位31ビットをそれぞれのノードに対応させた。このスワッピング処理は、データが入力される直前に1クロックでアドレスをスワッピングできるため、1クロックでのデータのスワッピングが可能となる。

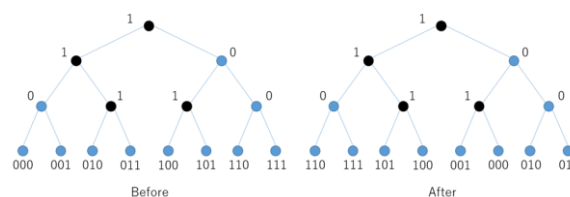


図3 スワッピング処理の概念図

表 2 スワッピングにおける元の値に対するパスの値とスワッピング結果

元の値	パスの値	スワップ結果
000	110	110
001	110	111
010	111	101
011	111	100
100	101	001
101	101	000
110	100	010

4. 評価

図 2 に示した機構は, Vivado 2015.3 を用いて実装した. 再構成は, Processor Configuration Access Port(PCAP)を介して行うことができ, 再構成にかかる時間 Reconf. Time は以下の式(1)で表すことができる.

$$\text{Reconf. Time(s)} = \frac{\text{The size of a partial bitstream(bit)}}{\text{The Bandwidth of PCAP(bps)}} \quad (1)$$

PCAP のバンド幅はノンセキュアモードで 400MB/s であり, 本研究で実装した再構成モジュールのパーソナルビットストリームのデータサイズは 323KB であったため, 上式(1)より再構成にかかる時間は, $\frac{323\text{KB}}{400\text{MB/s}} =$

0.8075ms と求まる.

本研究で実装した再構成領域 pblock と再構成モジュールそれぞれの回路規模を表 3 に示した. pblock は, Zynq ZC702 に含まれる Artix-7 の回路規模の約 11.2% の使用率であった. pblock におけるそれぞれの再構成モジュールの回路規模について, マスキング, マイクロ・アグリゲーション, ノイズ付加, スワッピングの順に, LUT は 662 個, 3,165 個, 1,590 個, 531 個で, その合計値は 5,948 個となり, レジスタは 1,107 個, 5,840 個, 2,402 個, 1,387 個で, 合計値は 10,736 個となった. したがって, 再構成を用いた提案機構では,

表 3 再構成領域における回路規模

	Pblock/11.2%	Masking	Micro_agg	Noise	Swapping
Slice as LUTs	6,000	662/11.03%	3,165/52.75%	1,590/26.50%	531/8.85%
Slice Registers	12,000	1,107/9.23%	5,840/48.67%	2,402/20.02%	1,387/11.56%

表 4 消費電力

	消費電力	
	Raspi 2	Zynq ZC702
Masking	1.0W (200mA, 5V)	1.7W
Noise		1.7W
Swapping		1.7W
Micro_agg		1.8W

再構成を用いない場合と比較して, 回路規模が増加したと言える. Xilinx によれば, 再構成を用いると, 配置配線の影響で 20%程度の回路規模のオーバーヘッドを考慮しなければならない [9]が, 表 3 より提案機構における再構成領域の回路規模のオーバーヘッドは約 50%であることが読み取れる. 以上より, 再構成を用いた回路規模の削減方法として, より多くの処理モジュールを実装すること, もしくは, 回路規模の類似する処理モジュールを同一再構成領域に実装することなどが考えられる.

次に提案機構の匿名化処理のスループットと消費電力量に関する評価を述べる. 本研究ではコーディネータを想定しており, 比較対象の一例として, Raspberry Pi 2 Model B(以下 Raspi 2)が挙げられるため [10] [11], 本研究報告では, Zynq ZC702 の Artix-7 における専用ハードウェアによる処理(HW 処理)と Raspi 2 の Cortex-A7 におけるソフトウェア処理(SW 処理)によるスループットと消費電力量の比較・評価を行った. なお, Zynq 上の FPGA の動作周波数は 100MHz, Raspi 2 の動作周波数は 900MHz であり, クアッドコアである [12]. 表 4 は Raspi 2 と Zynq ZC702 での各匿名化処理での消費電力を表す. Adafruit によればアイドル時の電流は 200mA で供給電源は 5V より Raspi 2 の消費電力は 1.0W と求められ [13], Zynq ZC702 に関してはそれぞれの処理モジュールにおいて Vivado で解析した. また以下の表 5 は, 各匿名化処理において, 32 ビットデータ 32 個のうち, はじめのデータが処理モジュールに入力されてから, 最後のデータが出力されるまでのクロック数とそのクロック数より算出されたスループットが示されている. 表 6 と表 7 はそれぞれ, 各匿名化処理の SW 処理, HW 処理による処理時間, スループット, 消費電力量を示している. ここで処理時間とは, 1 つの 32 ビットデータを処理するのにかかる時間を指す. 図 1 を考慮すると, マスキングとスワッピング処理では, 1Gbps を超えており, 中継器等の中間階層での利用を視野に入れることができ, 無線センサネットワークでの利用には十分なスループットであることがわかる. また, ノイズ付加とマイクロ・アグリゲーションに関しては, 1Gbps を下回ってはいるものの, SW 処理と比較して, より高スループットであり, 他のアプリケーションを処理する余地ができたと考えられる.

表 8 は, 表 6 と表 7 に示す SW 処理に対する HW 処理のスループット・消費電力量の改善率を示している. 改善率はスワッピングで最も大きく, スループットが約 328 倍, 消費電力量は約 $\frac{1}{200}$ 倍となった. これは, ハードウェア実装に適した XOR 演算を多用している

表 5 クロック数とスループット

	クロック数/clock	スループット/Gbps
Masking	65	1.58
Noise Addition	112	0.91
Swapping	65	1.58
Micro Aggregation	132	0.78

表 6 各匿名化処理の SW 処理による処理時間・スループット・消費電力量

	Raspi 2		
	処理時間	スループット	消費電力量
Masking	4.49 μ s	7.13Mbps	4.49W μ s
Noise	4.21 $\times 10^{-1}$ μ s	7.60 $\times 10$ Mbps	4.21 $\times 10^{-1}$ W μ s
Swapping	6.65 μ s	4.81Mbps	6.65W μ s
Micro_agg	5.53 $\times 10^{-1}$ μ s	5.78 $\times 10$ Mbps	5.53 $\times 10^{-1}$ W μ s

表 7 各匿名化処理の HW 処理による処理時間・スループット・消費電力量

	Zynq ZC702		
	処理時間	スループット	消費電力量
Masking	20.3ns	1.58Gbps	34.51Wns
Noise	35.0ns	0.91Gbps	59.50Wns
Swapping	20.3ns	1.58Gbps	34.51Wns
Micro_agg	41.3ns	0.78Gbps	74.34Wns

表 8 各匿名化処理の SW 処理に対する HW 処理による改善率

	スループット	消費電力量
Masking	222倍	$\frac{1}{125}$ 倍
Noise addition	12倍	$\frac{1}{6.67}$ 倍
Swapping	328倍	$\frac{1}{200}$ 倍
Micro_agg	13倍	$\frac{1}{7.14}$ 倍

ことに起因すると考えられる。一方、ノイズ付加の改善率が最も小さく、スループットが約 12 倍、消費電力量が約 $\frac{1}{6.67}$ 倍であった。改善率が低い理由として、ノイズを付加する処理の並列処理化を実装していないことが挙げられる。表 3 の回路規模の結果を考慮すれば、ノイズ付加処理を並列化することで、スループットの向上と再構成領域の回路規模削減が期待できる。また、マイクロ・アグリゲーションにおける改善率もノイズ付加と同様に低いが、これは SW 処理で利用している qsort によるソート処理が十分高速であることが考えられる。

5. 結論

本研究では、マスキング、ノイズ付加、マイクロ・アグリゲーション、スワッピングの 4 つの匿名化処理モジュールを、再構成を用いて Zynq ZC702 評価ボード上に実装し、動作を確認した。再構成による回路規模の削減を試みたが、再構成する処理モジュールの回路規模のバラつきが大きく、結果として、回路規模の削減を達成できなかった。しかし、同一再構成領域での処理モジュールを増やす、もしくは類似する回路規模を有する処理モジュールを複数実装することなどにより、回路規模削減が期待できると考えられる。スループットに関しては、Raspi 2 の SW 処理と比較して、その改善率に差はあるものの、基本的に HW 処理により高スループット化が図れ、また消費電力量に関しても、FPGA を用いたシステム設計により低消費電力化が図れることが示された。無線センサネットワークのコーディネータにおいて、スループット・電力効率とともに優れたシステムを設計できる可能性が示されたと考えられる。

6. 今後の課題

現在のシステムでは、データに対して匿名化処理を行うのみであり、実行した匿名化処理の種類や匿名化の程度などの情報を結果として返すシステムを構築する必要がある。また、匿名化処理には明確な匿名化基準として k-匿名性や l-多様性が存在し、データの 2 次利用や情報公開のためにこうした基準を満たす必要がある。そのため、k-匿名性や l-多様性の処理を Zynq PS 上のプロセッサで処理を行うなどが課題として挙げられる。さらに、匿名化処理後のデータが、元データと比較してどれだけ情報が損失されたかを示す情報損失率の評価も同時に行う。

再構成の観点では、回路規模削減とスループット向上のため、処理モジュールの並列性などの改良が必要と考えられる。回路規模の削減において、回路規模の大きさにバラつきがある場合、同一の再構成領域に実装せず、それぞれを独立に実装して回路規模削減を図ることも課題として挙げる。

謝 辞

本研究は、公益財団法人セコム科学技術振興財団研究助成、文部科学省科学技術研究費補助金基盤研究 B 「機能維持性を高める建物・複数機器の協調制御」(24360230)ならびに「コンテンツベース・スマートコミュニティインフラの構築と展開」(25280033)、国土交通省住宅・建築物技術高度化事業の一環としてなされた。

文 献

- [1] 経済産業省商務情報政策局, “IoT 時代に対応したデータ経営 2.0 の促進のための論点について (討議用資料),” 2 2015. [オンライン]. Available:

- http://www.meti.go.jp/committee/sankoushin/shojo/johoikeizai/pdf/002_07_00.pdf. [アクセス日: 13 1 2016].
- [2] B. Andreas, B. Florian, Z. Daniel , J. Teich, “Energy-Aware SQL Query Acceleration through FPGA-Based Dynamic Partial Reconfiguration”
- [3] W. Ying, Z. Xuegong, W. Lingli, Y. Jian, L. Wayne, P. Chenglian , T. Jiarong, “ SPREAD: A Streaming-Based Partially Reconfigurable Architecture and Programming Model,” IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS, VOL. 21, NO. 12, DECEMBER 2013, 2013.
- [4] K. Jalal, E. Ali, A. Adewale , A. Tughrul, “A Dynamic Partial Reconfiguration Design for Camera systems,” Adaptive Hardware and Systems (AHS), 2015 NASA/ESA Conference on , 2015.
- [5] S. Ubik, P. Zejdl and J. Halák, "Real-time anonymization in passive network monitoring," *Third International Conference on Networking and Services*, 2007.
- [6] A. Blake and R. Nelson, "Scalable Architecture for Prefix Preserving Anonymization of IP Addresses," *8th International Workshop SAMOS*, pp. 33 - 42, 2008.
- [7] J. Sawada and H. Nishi, "Hardware acceleration and data-utility improvement for low-latency privacy preserving mechanism," *22nd International Conference on Field Programmable Logic and Applications*, pp. 499 - 502, 2012.
- [8] Y. Fumito, M. Kanae , N. Hiroaki, “RAM-based Hardware Accelerator for Network Data Anonymization,” Field Programmable Logic and Applications (FPL), 2014 24th International Conference, 2014.
- [9] C. Kohn, “Partial Reconfiguration of a Hardware Accelerator on Zynq-7000 All Programmable SoC Devices,” 2013.
- [10] F. Mio, I. Minako, Y. Fumito , N. Hiroaki, “Construction of HEMS in Japanese Cold District for Reduction of Carbon Dioxide Emissions,” Industrial Electronics Society, IECON 2014 - 40th Annual Conference of the IEEE, 2014.
- [11] X. Patrick, F. George, O. Seán, K. Niall, E.-M. Fiona, L. Paul , P. Emanuel, “Sensing wind for environmental and energy applications,” Irish Signals & Systems Conference 2014 and 2014 China-Ireland International Conference on Information and Communications Technologies (ISSC 2014/CIICT 2014). 25th IET , 2014.
- [12] Raspberry Pi Foundation, [オンライン]. Available: <https://www.raspberrypi.org/products/raspberrypi-2-model-b/>. [アクセス日: 21 2 2016].
- [13] Adafruit, [オンライン]. Available: <https://learn.adafruit.com/introducing-the-raspberrypi-2-model-b?view=all>. [アクセス日: 21 2 2016].