A Performance Study of Secure Data Mining on the Cell Processor

STPG 2008
22 May 2008

Hong Wang, Hiroyuki Takizawa and Hiroaki Kobayashi.
GSIS, Tohoku University, Japan

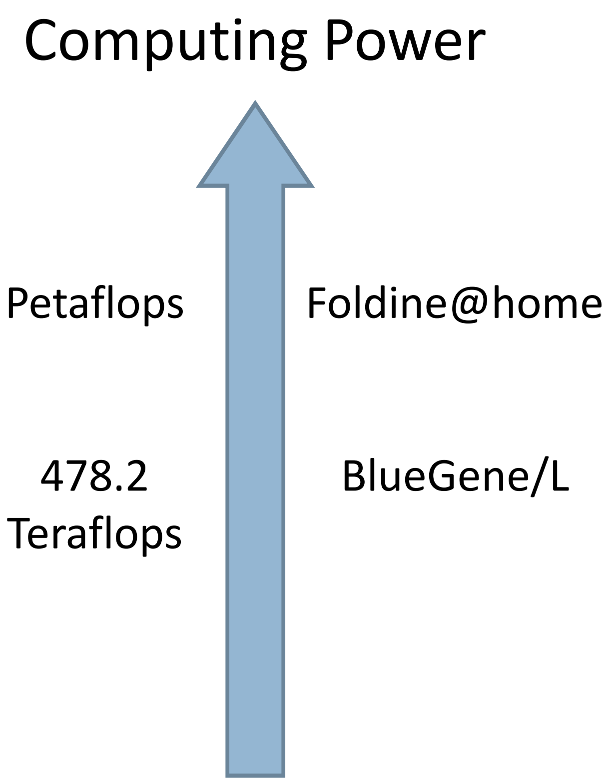
Outline

- Introduction
- The Cell Processor
- Secure K-Means Clustering for Volunteer Computing
- Performance Evaluation and Discussion
- Conclusion and Future Work

Introduction

Background, Motivation, and Related Work

Introduction



- Latest generation game consoles are equipped with high performance processors.
  - Ideal new volunteer computing peers.

- ***Limitation?***
  - Less of security features
  - Cannot be applied to sensitive data processing.

Introduction

- The Cell processor inside PlayStation 3 comes with hardware security features. These features can be utilized to address the security concerns for sensitive data processing on the volunteer computing platforms.
- Data mining on increasing volume of data requires more computing power. It also requires security features to process sensitive data.
- **A secure volunteer data processing method** is designed, and applied to K-Means clustering.

Related Work

- Privacy preserving data mining: modify original data to guarantee the privacy of data and knowledge.
  - Heuristic-based: selective data modification or sanitization; *has side effects*.
  - Reconstruction-based: perturbs data and reconstructs the distribution at an aggregate level; *has side effects*.
  - Cryptography-based: conducts data mining on private data from multiple parties; *not suitable for the volunteer computing model*.

Objective

- Volunteer computing requires security features to guarantee that even volunteer peer’s owner cannot access sensitive data.
- Cell processor’s hardware features make it possible to achieve this, while no side effect is introduced.
- This paper explores the potential of the cell processor for sensitive data processing with an data mining application.

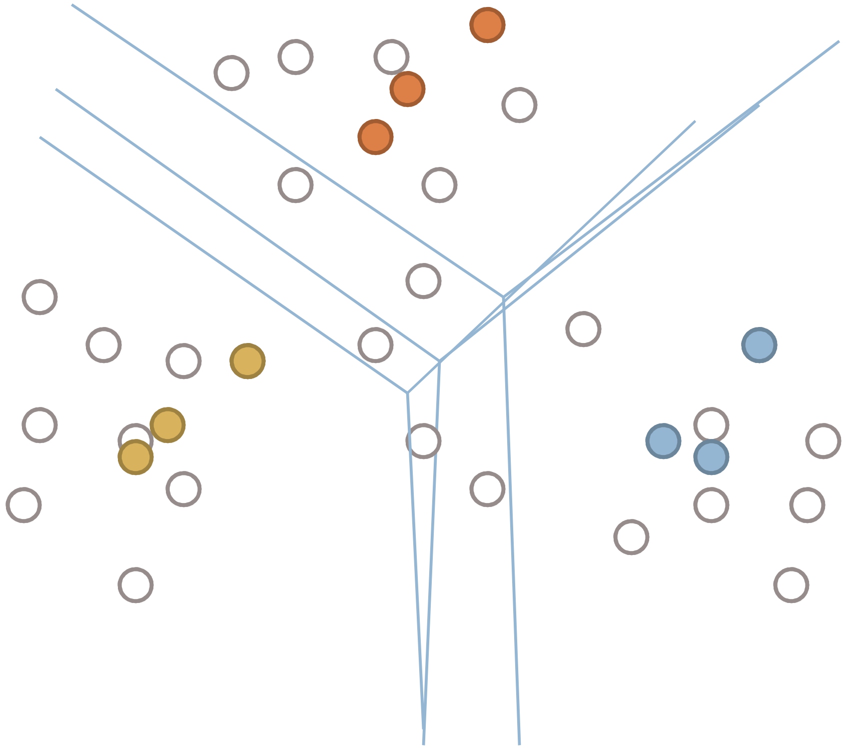
K-Means Clustering Algorithm

- Data clustering: cluster a dataset into a certain number of subsets.

- Application area: machine learning, data mining, pattern recognition, image analysis and bioinformatics.

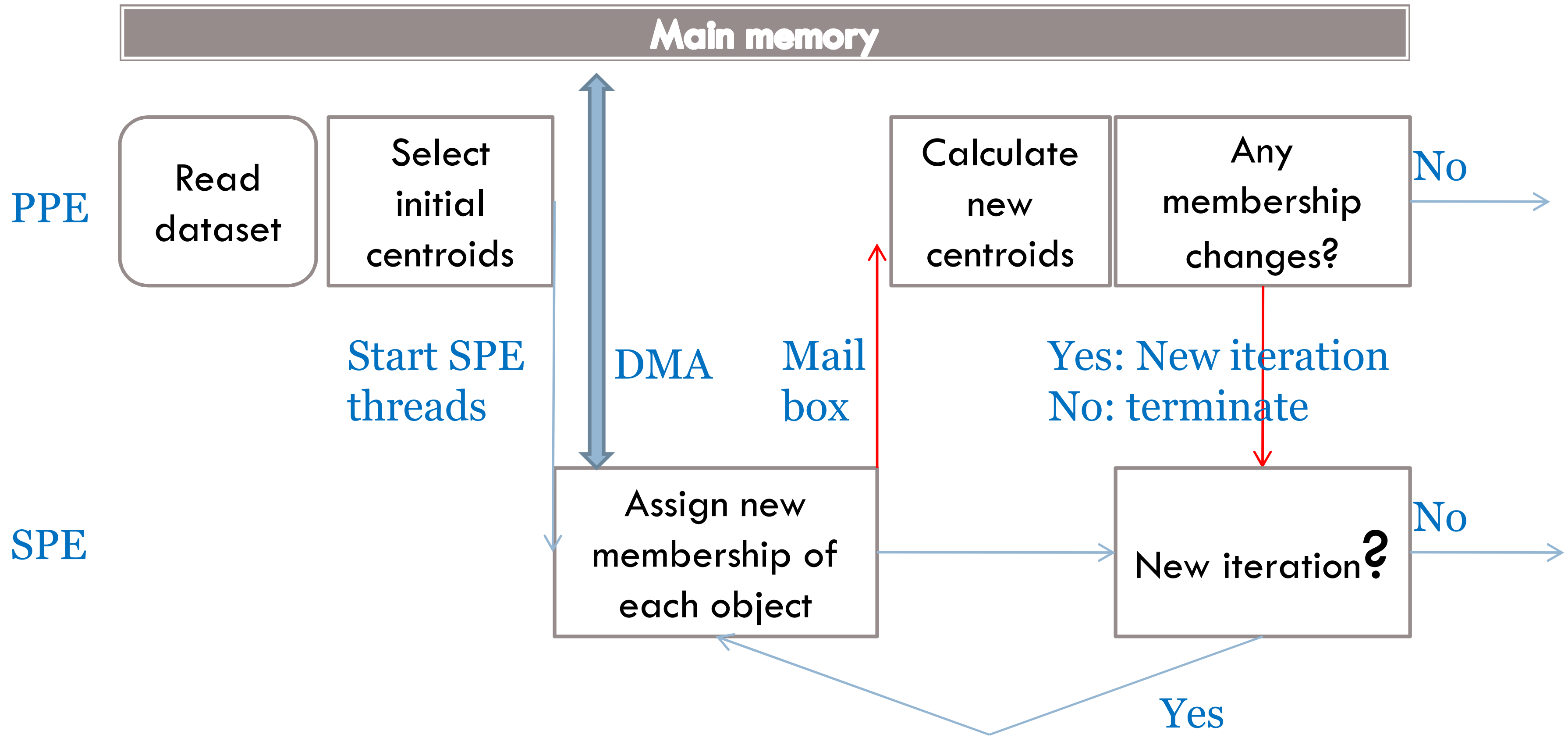
- Minimize the objective function: $\sum_{i=1}^{k} \sum_{n \in = S_i} |x_n - c_i|^2$

K-Means Clustering Algorithm



Parallelization Scheme

Previous work

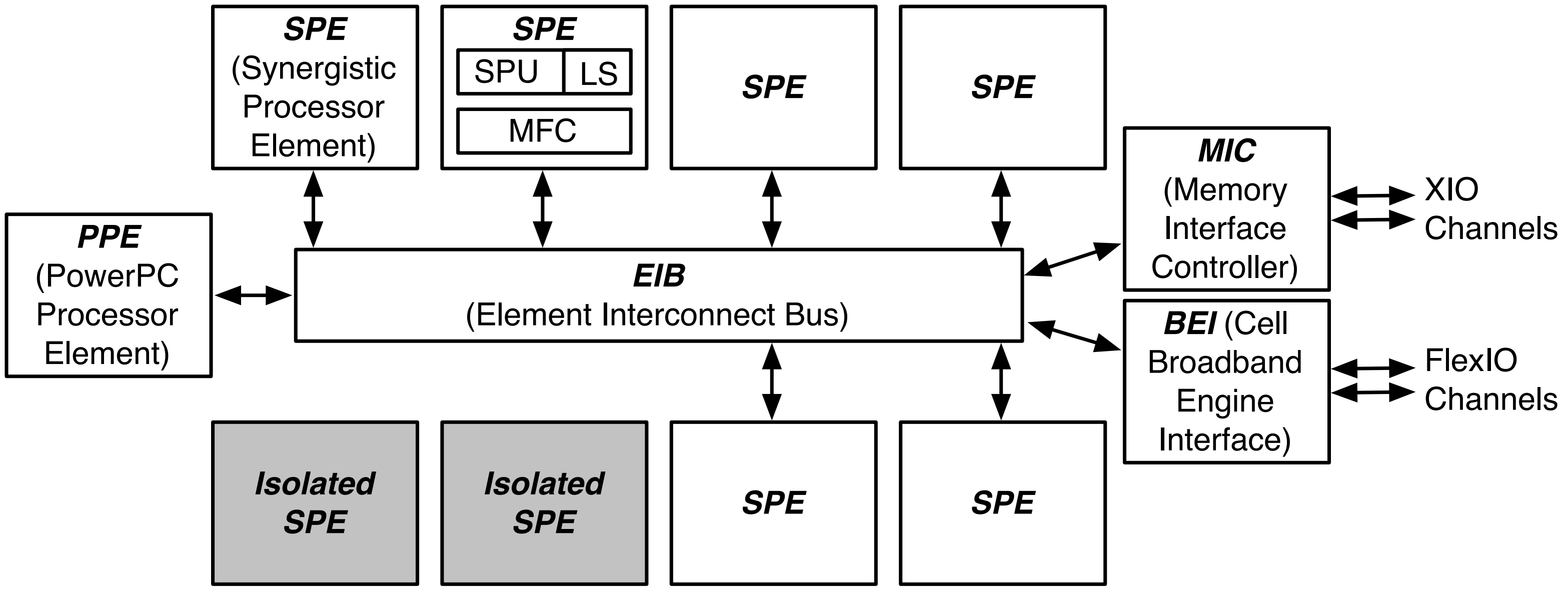


The Cell Processor

Cell Architecture Overview

Cell Security Features

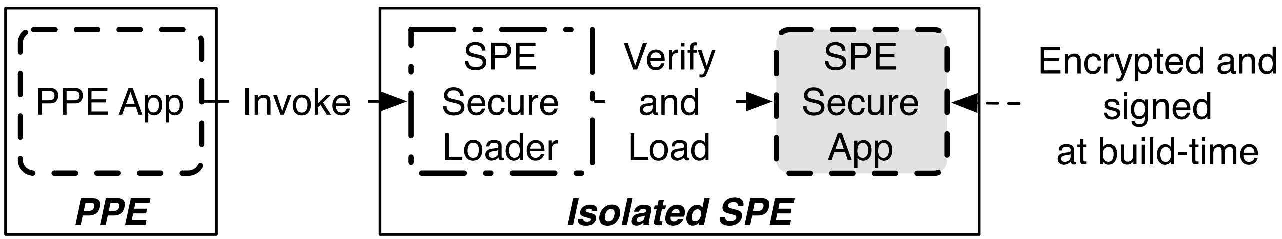
Cell Architecture Overview



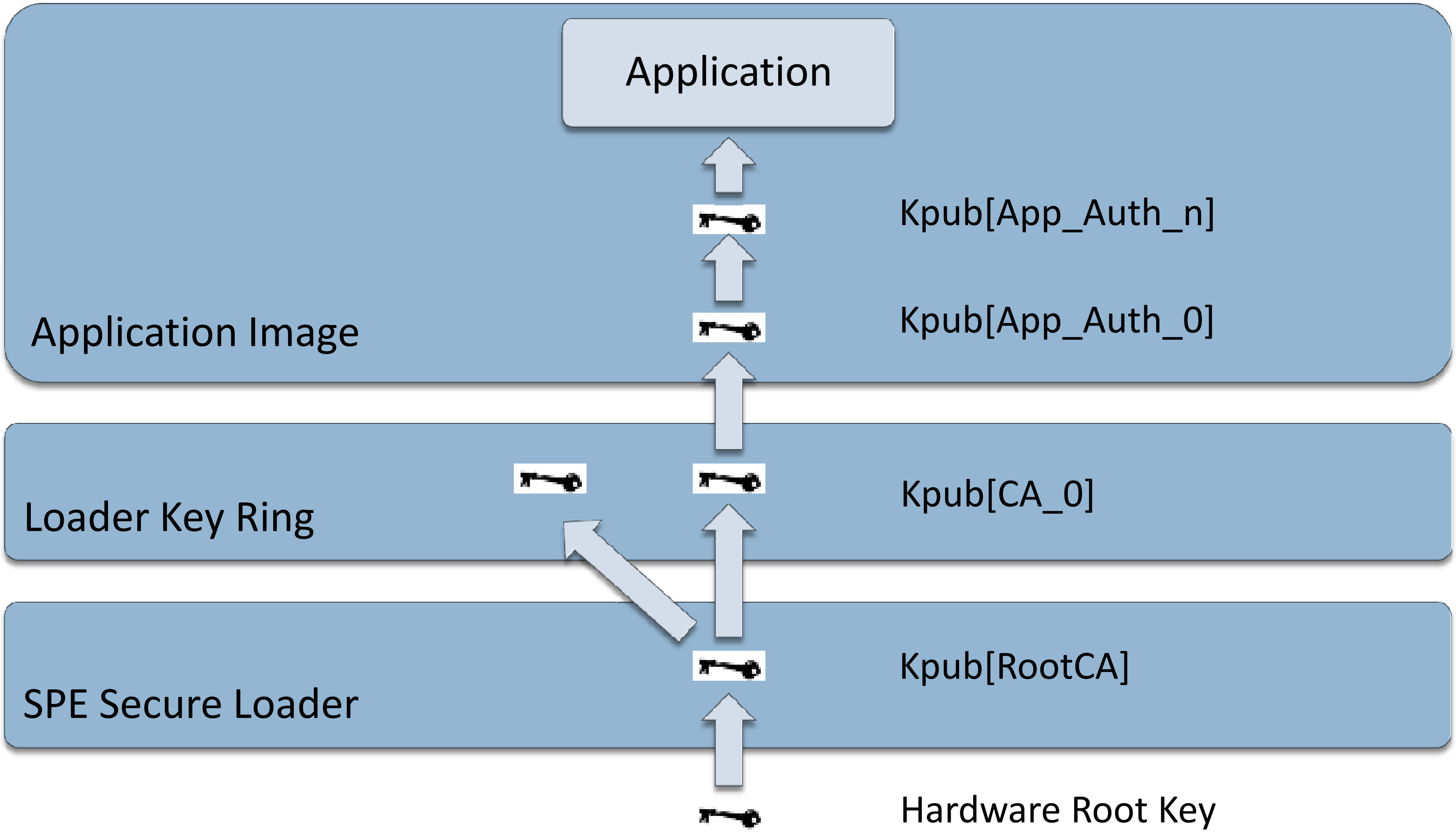
- PPE: PowerPC processor, responsible for overall control.
- SPE: 128-bit SIMD processors.

Cell Security Features

- Isolation mode:
  - Lock a SPE's local store for its own use.
  - External execution path control of the SPE is disabled.
  - The only external action for an isolated SPE is "cancel." The data in the LS is erased.
  - "decrypt in" and "encrypt out" functions.
  - The encryption/decryption key: 128-bit application key protected through a key hierarchy.



The Key Hierarchy



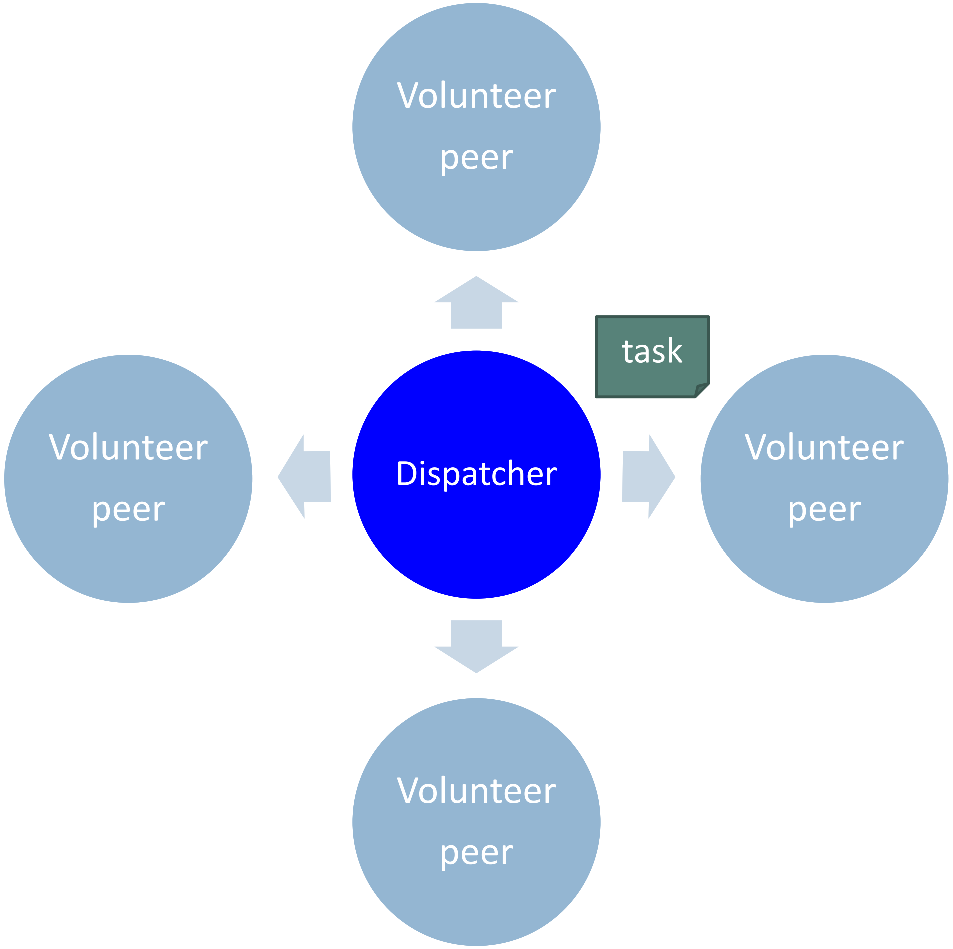
Secure K-Means Clustering for Volunteer Computing onthe Cell Processor

The problem

The security method

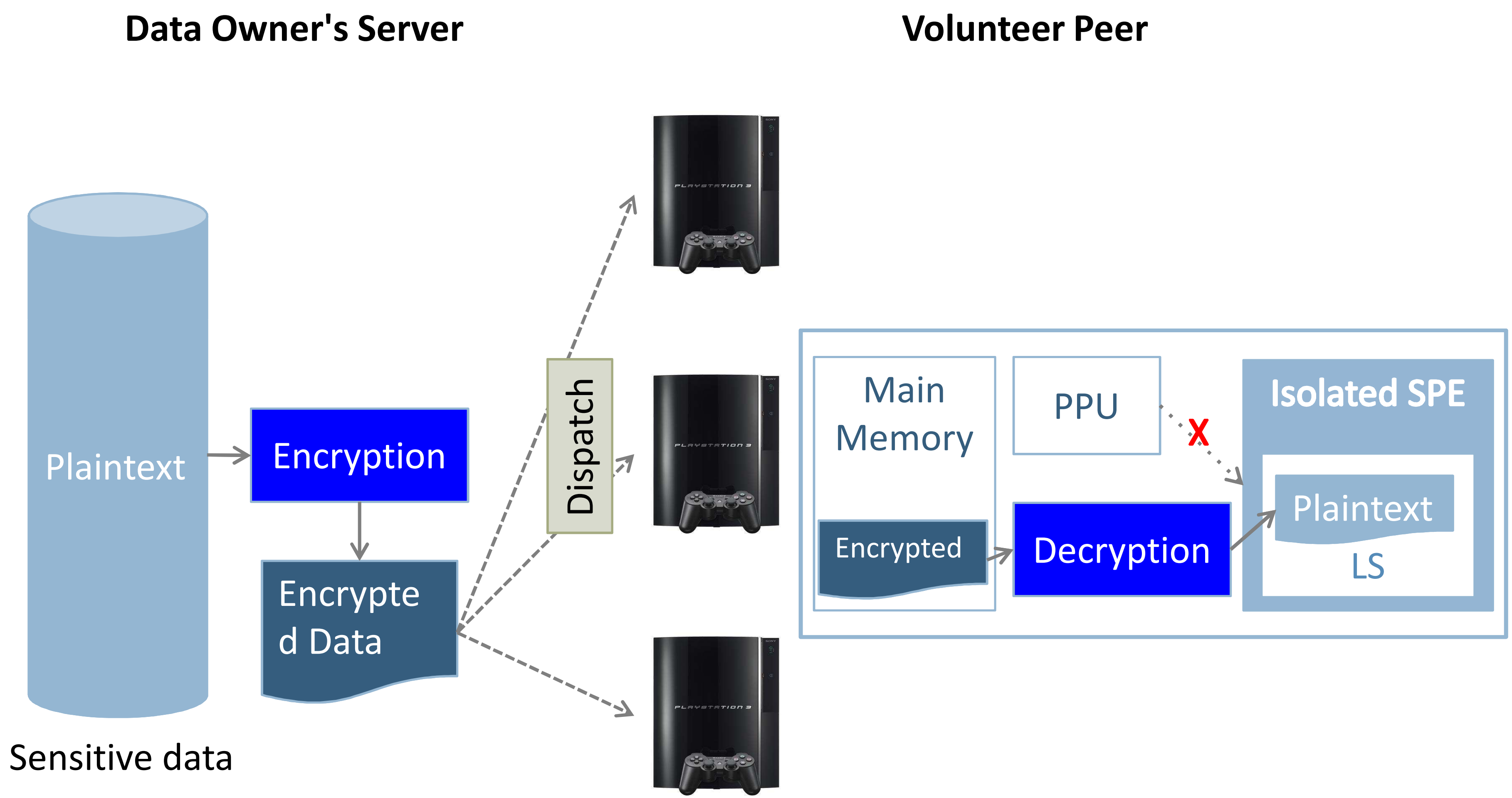
The secure K-Means clustering on Cell

Problem Definition



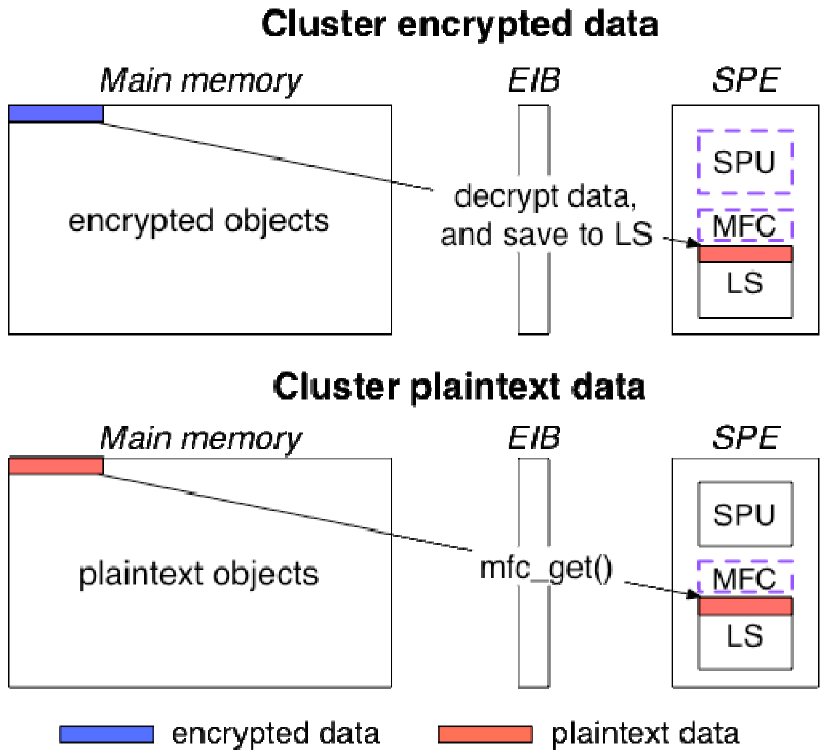
- Data inside the task is not owned by the volunteer peer.
- Thus, unauthorized access to the plaintext data should be disabled on the volunteer peer.
- ***How?***

The Cryptography Based Method



Secure K-Means Clustering on Cell

- Apply the security method to the parallel K-Means clustering algorithm for the Cell processor.
- Add extra data decryption while loading data from main memory.



Data transfer for encrypted data occupies both SPU and MFC. Thus double buffering optimization cannot be applied.

Performance Evaluation and Discussion

The advantage over non-secure algorithm on commodity processors

The overhead for the security features

Evaluation Environment

- OS: Fedora Core 5 for PPC.
- SDK: IBM Cell BE Security SDK 2.1.
- Simulated Cell Processor: 3.2GHz with 8 SPEs and 25.6GB/s memory bandwidth.

- SPE performance statistics of both the secure K-Means clustering and the non-secure K-Means clustering are gathered and compared.

Compare With Commodity Processors

Advantage over non-secure K-Means clustering on commodity processors	Athlon 64 3400+	PowerPC G4 1.67GHz
Single Precision	3.07x	8.29x
Double Precision	1.83x	5.57x

- The secure K-Means clustering algorithm running on the Cell processor *greatly outperforms* the non-secure K-Means clustering algorithm running on the commodity processors.

Performance Statistics

Single Precision	Secure	Non-Secure	
Total process cycles (20 iterations)	342,559,460	35,644,760	10.4%
Cycles for buffer transfer	287,501	9,364	
Cycles for buffer clustering	51,759	51,600	
Double Precision	**Secure**	**Non-Secure**	
Total process cycles (20 iterations)	744,896,330	242,251,574	32.5%
Cycles for buffer transfer	287,965	9,373	
Cycles for buffer clustering	172,806	172,693	

- Buffer transfer cycles for secure clustering is 30.7x of the value for non-secure clustering.
- All the data needs to be decrypted each iteration.
- Double buffering provides 17% (SP) and 10% (DP) improvement for the non-secure clustering.

Conclusion and Future Work

Conclusion

- Designed a security method on top of Cell processor's hardware security features to address the security issue with volunteer computing.
- Applied this method to solve a classic data mining algorithm – K-Means clustering.
- Evaluate the performance of the secure K-Means clustering; compared with non-secure algorithm on the Cell processor and commodity processors.
- Huge performance degradation is introduced by the security method; but still outperformance commodity processors.

Future Work

- Performance optimization:
  - Design and evaluate a data compression method to reduce the computation cost for data decryption.
- General secure volunteer data processing framework:
  - On top of the Cell security SDK and the BOINC platform.
  - Solve more applications with the general framework.

Thank You! Question?

This research was partially supported by Grant-in-Aid for Scientific Research on Priority Areas #18049003 and Strategic Information and Communications R&D Promotion Program (SCOPE-S) #061102002.