A publication of ISCA*:
International Society for Computers
and Their Applications

INTERNATIONAL JOURNAL OF COMPUTERS AND THEIR APPLICATIONS

TABLE OF CONTENTS

***"International Journal of Computers and Their Applications is Peer Reviewed".**

Volume 30, No. 2, June 2023 **ISSN 1076-5204**

International Journal of Computers and Their Applications

A publication of the International Society for Computers and Their Applications

ISCA Headquaters • 278 Mankato Ave, #220, Winona, Minnesota 55987, USA
Phone 507-458-4517 • Email isca@isca-hq.org • URL: http://www.isca-hq.org

Editorial

It is my distinct honor, pleasure, and privilege to serve as the new Editor-in-Chief of the International Journal of Computers and Their Applications (IJCA) for the second year. I have a special passion for the International Society for Computers and their Applications. I have been a member of our society since 2014 and have served in various capacities. These have ranged from being on program committees of our conferences to being Program Chair of CATA 2021, CATA2022 and CATA 2023 and currently serving as one of the Ex-Officio Board Members. I am very grateful to the ISCA Board of Directors for giving me this opportunity to serve society and the journal in this role.

I would also like to thank all the editorial board, editorial staff, and the authors for their valuable contributions to the journal. Without everyone's help, the success of the journal would be impossible. I look forward to working with everyone in the coming years to maintain and further improve the journal's quality. I want to invite you to submit your quality work to the journal for consideration of publication. I also welcome proposals for special issues of the journal. If you have any suggestions to improve the journal, please feel free to contact me.

Dr. Ajay Bandi
School of Computer Science and Information Systems
Northwest Missouri State University
Maryville, MO 64468
Email: AJAY@nwmissouri.edu

In 2023, we have four issues planned (March, June, September, and December). The March issue includes the selected papers from the CAINE 2022 and open submissions. The September issue will contain the best papers from CATA 2022 and open submissions. The last issue is taking shape with a collection of submitted papers.

I would also like to announce that I will begin searching for a few reviewers to add to our team. There are a few areas in which we would like to strengthen our board. If you would like to be considered, please contact me via email with a cover letter and a copy of your CV.

Ajay Bandi, Editor-in-Chief
Email: AJAY@nwmissouri.edu

Guest Editorial
June 2023

This issue of the International Journal of Computers and their Applications (IJCA) is split into two parts. The first part is a collection of five refereed papers selected from CAINE 2022, the second part is IJCA contributed papers which have gone through the normal review process. The papers in this issue cover a broad range of research interests in the community of computers and their applications.

ISCA Fall 2022 CAINE Conference: CAINE 2022 - The 35th International Conference on Computer Applications in Industry and Engineering, was held October 17-18, 2022. Due to the pandemic, it was held virtually. Each paper submitted to the CAINE 2022 conference was reviewed by at least two members of the international program committee, as well as by additional reviewers, judging for originality, technical contribution, significance, and quality of presentation. The proceedings for this conference can be found online at https://easychair.org/publications/volume/CAINE_2022

After the conference, the five best papers were recommended by the program committee members to be considered for publication in this special issue of IJCA. The authors were invited to submit a revised version of their papers. After extensive revisions and a second round of review, these papers were accepted for publication in this issue of the journal. The topics and main contributions of the papers are briefly summarized below:

YUE MA from Northern Illinois University, ZIPING LIU and CHARLES D. MCALLISTER from Southeast Missouri State University, cape Girardeau presents their work "Deep Reinforcement Learning for Portfolio Management." In this paper they propose a method to build deep reinforcement learning (DRL) agents to optimize asset allocation for maximum returns in financial portfolios. The agents are built using a combination of policy gradient methods from reinforcement learning and neural networks (CNN/RNN/CNN-RNN). Three types of portfolios are tested, including stocks, stocks affected by Covid-19, and a randomly selected combination of stocks and cryptocurrency. The performance of DRL agents is compared to equal-weighted agents and one-stock agents, and the results show that DRL agents outperform other agents in the randomly selected portfolio. A linear regression model is also tested and shows poor results compared to other agents.

RUI YANG, AZAD AZADMANESH, and HASSAN FARHAT of University of Nebraska, Omaha presents their work "Polygon Formation in Distributed Multi-Agent Systems." The authors here propose a two-phase procedure to address the challenge of forming a geometric pattern for a fleet of randomly distributed agents on the boundary of a polygon formation. The first phase involves forming an enclosing circle around the formation to prevent location-conflicts on the polygon. The second phase involves establishing projected points and moving toward locations on the polygon while avoiding collisions. The circle formation is used as a regrouping feature before the agents reconfigure themselves into a different polygon formation. The formation control laws have been verified through simulations for circle formation, convex polygons, and some categories of concave polygons, including cases for rotation, translation, and scaling of polygons.

SUSUMU MATSUMAE of Saga University, Japan presents his work entitled paper “Accelerating dynamic programming by *p*-fold pipeline implementation on GPU.” In this work he discusses the use of pipeline implementations of Dynamic Programming (DP) on Graphics Processing Unit (GPU) for a simplified DP problem. The approach involves using GPU cores to partially compute several elements of the solution table at one time. A *p*-fold pipeline technique is proposed to enable larger parallelism beyond the number of pipeline-stages, to accelerate the pipeline implementation. The effectiveness of the approach is shown through experimentation.

JONAS KALLISCH and CHRISTOPH WUNCK from Emden/Leer University of Applied Science, Germany presents their work “Development of a Prototype for a Process Support and Analysis Platform for Small and Medium-sized Enterprises.” In this paper the authors present a prototype for an information exchange system that enables companies to exchange information without sharing data. The need for such a platform is explained, and the value for supply chains is described. The paper presents a literature review of existing concepts and techniques that contribute to the development of architecture. The information exchange concept and the prototype implementation are explained in detail.

ANJILA NEUPANE, RESHMI MITRA, INDRANIL ROY MCALLISTER from Southeast Missouri State University, cape Girardeau, BIDYUT GUPTA from Southern Illinois University, Carbondale and NARAYAN DEBNATH from Eastern International University, Vietnam present their work “Efficient and Secured Data Lookup Protocol using Public-Key and Digital Signature Authentication in RC-Based Hierarchical Structured P2P Network.” This paper proposes a public key cryptography and digital signature-based data lookup protocol that is efficient and secure for a Residue Class (RC)-based hierarchical structured Peer-to-Peer (P2P) network. RC allows for minimal latency in intra and inter-group communications, making it an ideal design choice. The paper provides efficient schemes for public-key cryptographic security and digital signature-based authentication for existing communication protocols. P2P networks are often vulnerable to network security attacks due to their limited resources, and this protocol aims to address these issues.

IJCA Contributed Papers: As was mentioned earlier, the second part of this issue is made up of papers that were contributed to the International Journal of Computers and their Applications (IJCA). The topics and main contributions of the papers are briefly summarized below:

TASNEEM YOUSEF and SAID GHOUL from Philadelphia University, Amman present their work “Toward Automated Feature Model Generation from UML Use Case and Class Diagrams.” This paper discusses how the Feature Model is important in Software Product Line development as it represents commonalities and variabilities between software products. However, generated feature models do not include data and actor aspects. This paper proposes an approach that defines a methodology for automated feature model generation from requirement documents limited to UML use case and class diagrams. The target generated feature model is enhanced by the introduction of data and actor concepts. An evaluating real case study (quality assurance at Philadelphia University) was used to evaluate the feasibility of the proposed approach, and it showed practical benefits.

AHLEM YOUSEF, SAID GHOUL and MOHAMMAD TAYAE from Philadelphia University, Amman present their work “Towards Automated Goal Model Generation from UML Use Case and Swimlane Diagrams.” This paper discusses how the Goal Model is important in goal-based requirements engineering and specifies software goals and their relationships. Previous research has shown that merging behavior and soft goals into a single model unit result in complex and costly maintenance of the generated Goal Model. This paper proposes a semi-automated approach to split the generated Goal Model into three separate aspects models (behavior, soft, and constraints) to facilitate maintenance and evolution. UML use case and Swimlane diagrams are inputs, and the output is a separate aspects model GM. The feasibility of the approach was validated on a concrete business case, and the approach was found to be valuable in any goal-oriented requirements engineering application.

MOUNIR ELKHATIB, KHADEEJ A. ALNAQBI, SHAIMA ALHARMOODI and WESAL ALNAQBI from Hamdan Bin Mohamad Smart University, Dubai, RAED ABU ZITAR from Sorbonne University, Abu Dhabi, and ALI BAYDOUN from St. George’s University, Grenada presents their work “Effect of Big Data and Analytics on Managing Projects.” This paper discusses the importance of embracing technological advancements, specifically Big Data analytics, in project management to stay relevant and adapt to the rapidly evolving field. The potential benefits of Big Data analytics, including reduced project complexities, cost, and enhanced project risk management, are motivating businesses to invest in this technology. The paper presents a literature review followed by primary and secondary data analysis through interviews and surveys to examine the impact of Big Data and business analysis on project management. Qualitative and quantitative data are used to explore the impact of Big Data analytics on project management.

KALIM QURESHI and NOOR ALSAEED from Kuwait University, Kuwait present their work “A Simulation Study of Auction Based Pricing Strategies in Grid Computing.” This paper investigates different auction strategies, such as Continuous Double Auction (CDA), Stable Continuous Double Auction (SCDA), Threshold Price Double Auction (TPDA), and Preston McAfee Double Auction (PMDA), for optimal utilization of cloud/grid resources. The study evaluates these strategies based on their suitability for end-users and their performance in terms of prospective resources, deadline consumption, budget spending, resource-derived profits, and immediate resource allocation. The results indicate that SCDA is effective for budget spending, CDA is good for resource allocation, and TPDA and PMDA have the highest performance in terms of deadline consumption.

MOUNIR EL KHATIB, AHMED ALNAQBI, ABDULLA ALNAQBI, HAMAD ALSUWAIDI, MEERA AL MARRI and AHMED ANKIT from Dubai Academic City, Dubai, RAED ABU ZITAR from Sorbonne University, Abu Dhabi present their work “Implementing IOT in Effective Project Management.” This paper discusses how the Internet of Things (IoT) can enhance project management by automating tasks, improving communication, and sharing data among team members. The research highlights the challenges associated with IoT, such as security, interoperability, and managing complexities, and proposes solutions and best practices to overcome them. The study involved semi-structured interviews and questionnaires. The research indicates that IoT can significantly reduce costs, increase efficiency, and improve communication and coordination in project management. However, careful planning, design, and management are required to address the challenges and ensure a successful implementation. Key takeaways include considering the user experience, managing data and devices, and proper testing of the system.

As guest editors, we would like to express our deepest appreciation to the authors and the reviewers. We hope you will enjoy this issue of the IJCA. More information about ISCA society can be found at http://www.isca-hq.org.

Guest Editors:
Yan Shi, CAINE 2022 Conference Chair
Gongzhu Hu, CAINE 2022 Program Co-Chair
Krishna Khambhampaty, CAINE 2022 Program Co-Chair
Takaaki Goto, SEDE 2022 Program Co-Chair
Ajay Bandi, Northwest Missouri State University, USA

June 2023

Deep Reinforcement Learning for Portfolio Management

Yue Ma*
Northern Illinois University, DeKalb, IL

Ziping Liu † and Charles D. McAllister †
Southeast Missouri State University, Cape Girardeau, MO

Abstract

This paper discusses how to build deep reinforcement learning (DRL) agents to determine the allocation of money for assets in a portfolio so that the maximum return can be gained. The policy gradient method from reinforcement learning and convolutional neural network/recurrent neural network/convolutional neural network concatenated with the recurrent neural network from deep learning are combined to build the agents. With the proposed models, three types of portfolios are tested: stocks portfolio which has a positive influence due to the Covid-19; stocks portfolio which has a negative influence due to the Covid-19; and stocks, cryptocurrency combined portfolio which are randomly selected. The performance of our DRL agents is compared with that of equal-weighted agent and all the money fully invested in one-stock agents. All of our DRL agents showed the best performance on the randomly selected portfolio, which has an overall stable increasing trend. In addition, the performance of a linear regression model is also tested with the random selected portfolio, and it shows a poor result compared to other agents.

Key Words: Deep reinforcement learning; financial portfolio management; machine learning; neural networks; stock.

1 Introduction

Deep reinforcement learning shows its advantages in various fields, such as gaming [17, 7, 5], transportation [18, 11, 8] and the medical field [1, 24, 20]. These successful applications have inspired more and more researchers to apply this framework to other fields. For example, use in the financial market can expect to achieve the same success.

Investment in the financial market plays an important role in a global economy, especially the stock market investment [3]. However, the financial market is not only large and volatile [16], but also can be heavily impacted by all sorts of unpredictable natural [9], social [6], and political factors [13]. Hence, the impacts of actions taken on the financial market will not be measured easily and quickly, which adds difficulties on the building of effective financial models. In deep reinforcement learning, the goal is to maximize the reward. Because of this special feature, deep reinforcement learning has gained a lot of attention in the field of gaming. Bearing the goal to maximize the overall return in financial portfolio management, which is similar to gaming, deep reinforcement learning should be applicable to manage financial portfolios. Considering the impact of the volatility of Covid-19 on the financial market, three different types of portfolios were selected to test the hypothesis in this paper.

This paper investigates the applicability of deploying deep reinforcement learning agents in the design of portfolio management agents such that weights in the portfolio can be allocated automatically to maximize the overall return. Convolutional neural network and recurrent neural network are investigated in the construction of the policy network, and the performances between agents with different network structures are compared.

2 Literature Review

Financial portfolio management is about the allocation of financial assets to meet the financial goals of investors. Generally, most investors have the same goal: maximize the expected return. In many different fields, such as medical diagnosis, speech recognition, and market clustering analysis, machine learning methods have proved to be powerful analytical tools. The above successful cases have motivated researchers to design machine learning models for portfolio management.

In [4], Gah-Yi Ban et al. introduced performance-based regularization (PBR) and performance-based cross-validation for the portfolio optimization problem and investigated them in detail. Tianello [2] used regression analysis to predict how each type of asset performs on a risk adjusted basis, and then allocates funds to those assets proportionately. Compared to classic machine learning, deep reinforcement learning prefers to achieve a pursued long-term result and portfolio management has a similar goal to maximize the rewards. The aforementioned similar goal has driven many researchers to develop and implement different deep reinforcement learning methods to manage financial portfolios. Liang et al. in [23] used deep reinforcement learning algorithms with continuous action space for the asset allocation. The agents were tested with risk adjusted accumulative portfolio value as objective function and the inputs took the combinations of different features. In [22], Liang et al. implemented three state-of-art continuous reinforcement learning algorithms, Deep Deterministic Policy Gradient (DDPG), Proximal Policy Optimization (PPO) and Policy Gradient (PG) in portfolio management and proposed

*e-mail: yuema@niu.edu
†e-mail: zliu,cdmcallister@semo.edu

an adversarial training method and showed that it can greatly improve the training efficiency and significantly raise average daily return and Sharpe ratio in backtests. Zhang et al. proposed a cost-sensitive portfolio selection method with deep reinforcement learning in [21], where a novel two-stream portfolio policy network was devised to extract both price series patterns and asset correlations. The proposed method also includes a new cost-sensitive reward function to maximize the accumulated return as well as to constrain both costs via reinforcement learning. Kanwar in [10], applied model free reinforcement learning algorithm which is a computationally efficient algorithm that directly estimates the optimal policy or value function through algorithms such as policy iteration or value iteration, with emphasis on policy gradient and Actor Critic Methods to build the trading agent in order to maximize its overall return. Opposite to [14], Yu et al. used a model-based approach which allows agents to plan ahead a range of possible choices to build the deep reinforcement learning agents. Jiang et al. also presented a model-free convolutional neural network with historic prices on a set of financial assets as its input, which outputs portfolio weights of the set in [26]. In [25], Jiang et al. proposed portfolio vector machine (PVM) which is an online stochastic batch learning scheme. In the proposed scheme, PVM works together with the policy network via convolutional neural network (CNN), recurrent neural network (RNN) and the Long Short-Term Memory (LSTM) to re-balance the cryptocurrency portfolio every 30 minutes. Selim Amrouni et al. in [15] extended the work of Jiang et al., applied the PVM to trade the daily stock and cryptocurrency data with a daily rebalance.

This research was motivated by both Jiang et al. [25] and Amrouni et al. [15], and used their work as reference for the basic policy network architecture of a convolutional neural network but implemented with a recurrent neural network and a mixed neural network (RNN+CNN) to build agents and used various financial asset combinations for testing. Since Covid-19, the stock market has experienced the most volatile ups and downs. Meanwhile, there is less related research taking Covid-19 into consideration regarding deep reinforcement learning on portfolio management. In order to gain insights, this research selected assets with different combinations to build the financial portfolio on testing the performance of deep reinforcement learning agents. To compare the performance of deep reinforcement learning with classic machine leaning approaches in portfolio management, linear regression is also tested for weights allocation.

3 Methodology

3.1 Deep Learning

Deep learning is one branch among the many fields of machine learning, and it is based on artificial neural networks. Convolutional neural networks are designed to work with grid-structured inputs, which have strong spatial dependencies in local regions of the grid. The most obvious example of grid-structured data is a 2-dimensional image, while other forms of sequential data like text, time-series, and sequences can also be considered special cases of grid-structured data with various types of relationships among adjacent items. The core concept of convolutional neural network is to simplify complicated questions. It reduced the dimensionality of a large number of parameters into a small number of parameters before processing.

Certain data types such as time-series, text, and biological data contain sequential dependencies among the attributes. Taking time-series data set as example, the values on successive timestamps are closely related to one another. If one uses the values of these timestamps as independent features, then key information about the relationships among the values of these timestamps is lost. For example, the value of a time-series at time t is closely related to its values in the previous window. However, this information is lost when the values at individual timestamps are treated independently of one another. A recurrent neural network, also known as RNN, can solve this problem. RNN is a class of neural networks which allow previous outputs to be used as inputs while having hidden states. The main advantage of recurrent neural network is that it can model sequence data, which is also the reason why RNN models are mostly being used by processing natural language.

3.2 Deep Reinforcement Learning

Learning can be divided into supervised learning, unsupervised learning and reinforcement learning. Supervised learning models receive features (X) and labels (y) as training data, using them to get predicted values (y); while unsupervised learning models receive only features (X) and to do further analysis; but reinforcement learning can be different with these learning types. Reinforcement learning (RL) [12] is the study of how an agent can interact with its environment to learn a policy which maximizes expected cumulative rewards for a task. Environment in DRL represents the outside world the agent interacts with. Agent is the individual which makes decisions to decide what actions to take in the environment. State is the current situation of agent, what the agent observed in the environment. Action is what the agent can do in the environment, the input provides to the environment. Reward is a feedback signal from the environment. Policy is used by an agent as a strategy to decide what actions should be taken given state s. Policy had deterministic policy and stochastic policy.

3.3 Financial Portfolio Management

Markowitz portfolio theory, which is also called Modern portfolio theory (MPT), is a theory on how risk-averse investors can construct portfolios to maximize expected return based on a given level of market risk. The investor can decide the risk or the expected rule, but there is a rule to follow, which is: More risk will be taken, when more money is expected in return. How we compute the expected rate of return for an individual investment as shown in the formula below:

$$E(R_{port}) = \sum_{i=1}^{n} w_i R_i$$

where:

w_i = the weight of an individual asset in the portfolio, or the percent of the portfolio in Asset i

R_i = the expected rate of return for Asset i

The expected rate of return for a portfolio of investments is simply the weighted average of the expected rates of return for the individual investments in the portfolio.

4 Research Methodology

4.1 Overall Architecture

The concept of using efficient portfolio management agents to maximize the expected return and to avoid risk follows Markowitz portfolio theory. In this paper, a model with deep reinforcement learning (DRL) agents is studied to determine the allocation of money for assets in a portfolio so that the maximum return can be gained. The policy gradient method from reinforcement learning and convolutional neural network/ recurrent neural network/ recurrent neural network concatenated with the convolutional neural network from deep learning are combined together to build the agents. Figure 1 shows the overall architecture of the deep reinforcement learning agent. Data should be pre-processed based on different choices of neutral networks, and then the pre-processed data is passed to the neural networks. Softmax is applied in order to get the weight distribution of the assets. Input data, neural networks and the weight distribution formed the policy network, and reward is calculated based on the calculated weights. Reward is next fed into an objective function to update the policy network.

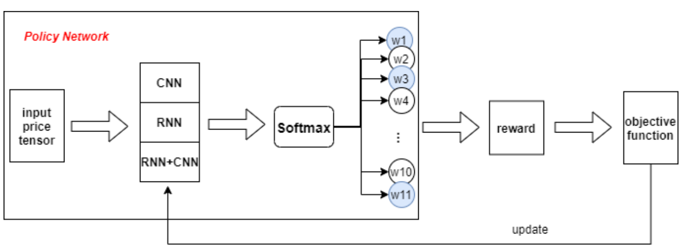


Figure 1: Overall deep reinforcement learning architecture

4.2 Data Collection

Considering the impact of the volatility of Covid-19 on the financial market, in this research, various combinations of stocks and cryptocurrency are selected to be tested, and all the data were collected from Yahoo! Finance [19].

4.2.1 Negatively Influenced Portfolio. The stocks which are heavily impacted by the Covid-19 are selected as the negatively influenced portfolio. These stocks are from the following companies: American Airlines Group (AAL), Delta Airlines (DAL),Southeast Airlines (LUV), United Airlines Holdings (UAL), Spirit Airlines (SAVE), Hilton Worldwide Holdings (HLT), Marriott International(MAR), Choice Hotels International (CHH), InterContinental Hotels Group PLC (IHG), and International Consolidated Airlines Group (ICAGY). Figure 2 shows the daily closing price evolution for the negatively influenced portfolio from January 25, 2016 to January 25, 2021. X-axis represents the days, and y-axis represents close price(US dollars).

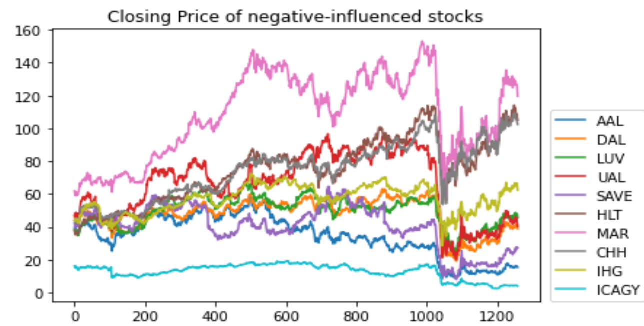


Figure 2: Negatively influenced stocks

4.2.2 Positively Influenced Portfolio. The stocks which have a positive influence because of the Covid-19 are selected as the positive influenced portfolio. These stocks are from the following companies: Alibaba Group Holding Limited (BABA), Walmart (WMT), eBay (EBAY), Best Buy (BBY), Allied Healthcare Products (AHPI), Amedisys (AMED), Viridian Therapeutics (MGEN), Meridian Bioscience (VIVO), NanoViricides (NNVC), Netflix (NFLX). Figure 3 shows the daily closing price evolution for the positive influenced portfolio from January 25, 2016 to January 25, 2021. X-axis represents the days, and y-axis represents close price (US dollars).

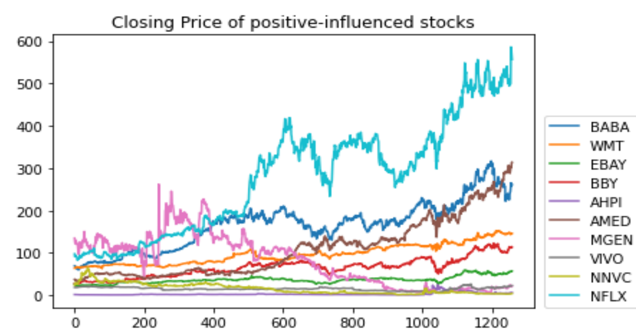


Figure 3: Positively influenced stocks

4.2.3 Randomly Selected Portfolio. To consider neither the positive impact nor the negative impact of Covid-19, seven random US stocks and three cryptocurrencies are selected from the following companies: American Airlines Group (AAL), Facebook (FB), Bank of America (BAC), Best Buy (BBY), Alibaba Group (BABA), Axon Enterprise (AAXN), BlackRock (BLK) , and cryptocurrency of Ethereum (ETH), Ripple (XRP), Litecoin (LTC). Figure 4 shows the daily closing price evolution for the positive influenced portfolio from January 25, 2016 to January 25, 2021. X-axis represents the days, and y-axis represents close price (US dollars).

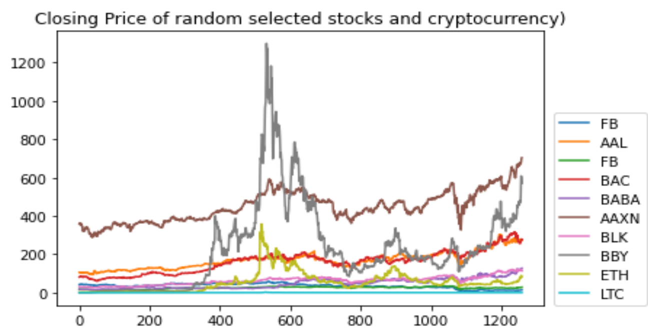


Figure 4: Randomly selected stocks and cryptocurrency

4.3 Data Preprocessing

For both stocks and cryptocurrencies, data between January 25, 2016 to January 25, 2021 were downloaded and they span totally five years, 1258 days for stocks and 1828 days for cryptocurrencies. The reason why only five years data were selected is that cryptocurrency does not have historical data before 2009 when the first cryptocurrency was adopted as bitcoin. Since then, more and more newly generated cryptocurrencies have come out. To make more use of stable cryptocurrency data in the model, our research restricted the data selection to the recent five years.

There was a problem when data for the random selected portfolio was processed, since the length of two types of investment data do not have the same length and has a gap of 569 trading days. In this research, for the days which cryptocurrency trading occurred while the stock trading did not occur, the price information for cryptocurrency was removed.

The data is divided as below:
Total days: 1258
Training days: 60% of total days
Validation days: 20% of total days
Test days: 20% of total days

Normal indicators for stock include "Date," "Open," "High," "Low," "Close," "Adj Close" and "Volume," but for this research, only "Open," "High," "Low" and "Close" were used. Strictly speaking, the market for cryptocurrency never closes, but there is still data available as "Open price" and "Close price" provided for cryptocurrency. "Open" generally refers to the price at 12:01 AM UTC of any given day and "close" generally refers to the price at 11:59 PM UTC of any given day. Hence, stock data and cryptocurrency data can be mixed together.

Figure 5 shows the data set weight distribution for this research, testing data set contains 253 days.

4.3.1 CNN Price Tensor Pre-Processing. The input data are arranged into 3D in shape of (4,10,1258) as shown in Figure

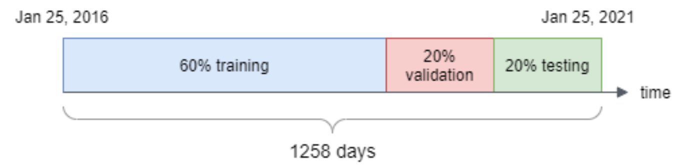


Figure 5: Dataset weight distribution

6. The first dimension 4 represents the number of features, and they are:

- Close(t-1)/Open(t-1)
- High(t-1)/Open(t-1)
- Low(t-1)/Open(t-1)
- Open(t)/Open(t-1)

Here the prices are normalized to the same scale. The second value for the dimension is 10, where it represents ten asset items (stocks+cryptocurrency) in the portfolio. The third value for the dimension is 1258, which represents the time steps (days). 1258 was sliced to 50 days each before passing to the network.

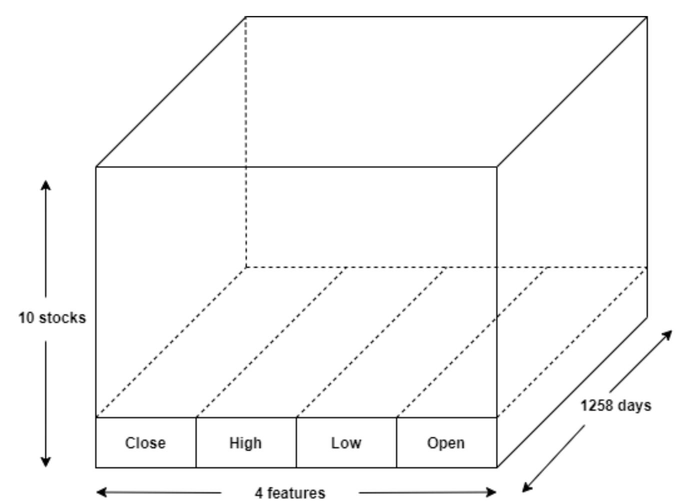


Figure 6: CNN data pre-processing

4.3.2 RNN and RNN+CNN Price Tensor Pre-Processing. For the architecture of recurrent neural network, and the architecture for the mixed (RNN+CNN) neural network, these two networks have the same input shape. There is only one feature (the mean value for the open price, close price, high price and the low price) used. The mean prices are passed to MinMaxScaler which helps to make all the value in the same scale.

$$MeanPrice = \frac{OpenPrice + ClosePrice + HighPrice + LowPrice}{4}$$

After data scaling, data was grouped to 50 days each. In the previous section's CNN policy network, the length of tensor was set to be 50 days. In order to do a comparative analysis with the CNN policy network, the tensor should be processed as the same length for batch processing. After transposing and dimension expanding, pre-processed input data has the shape (25,1,10,50), which is illustrated in Figure 7. The diagram shows that there are 25 batches of (1,10,50), for each step, one batch (1,10,50) from (25,1,10,50) was passed to the policy network to calculate the weight distribution. 1 means there is only one feature–mean value, 10 represents 10 stocks, and 50 is the length of the tensor(50 days).

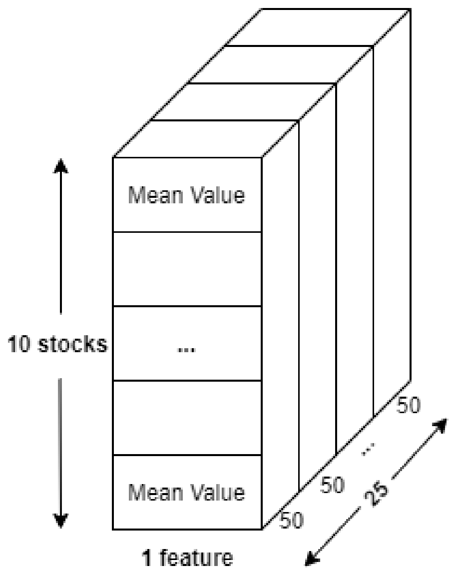


Figure 7: RNN and RNN+CNN data pre-processing

4.3.3 Linear Regression Price Tensor Pre-Processing. Data processing for linear regression is similar as for RNN. However, in linear regression, it needs X (features) and y (target values). In this research, the initial investment and mean value (open price, close price, high price and low price) of 10 stocks are chosen for X, which has 11 features in total. Initial investment is set to 10000 for every data time, and y is the portfolio value which is randomly generated between values of (10000,13000). After setting the portfolio value for each data time, the initial investment of each day will be adjusted to the previous portfolio value. Figure 8 shows the shape of pre-processing data for linear regression model.

Money	Stock1	Stock2	...	Stock10	Portfolio Value
10000	12.3	25.6		28.8	12500
12500	12.9	22.6		15.8	12907
12907	17.8	11.1		15.9	11008
...	...	...		...	...

Figure 8: Linear regression data pre-processing

4.4 Reinforcement Learning Framework

Markov Decision Process (MDP) provided the framework for describing the reinforcement learning problem. The state of the agent is the input matrix X_t and previous portfolio weights at time *t-1*.

$$state = S_t = (X_t, w_{t-1})$$

The action of the agent is the weight vector of portfolio at time *t*.

$$action = w_t = (w_1, ..., w_n)$$

The reward function is defined such as it is the agent's return minus a baseline return minus a term proportional to the maximum of the weight (this term is set up to make the agent avoid investing fully in one stock), and baseline is an equal weighted agent which divides the investment into the same proportion.

$$r_t = \sum w_i' y_i - \frac{1}{m} \sum y_i - \alpha max(w_1, ..., w_m)$$

where:

y_i is the return of single stock or cryptocurrency.

$\sum w_i' y_i$ is the sum of each stock or cryptocurrency's return. The formula for calculating the expected rates of return was used as reference.

m is number of items(stocks and cryptocurrency) in the portfolio.

α is the parameter of regularization, which can avoid investing fully in one stock.

The input price tensor is the state for the deep reinforcement learning model, and action is the weight distribution which was calculated by the policy network. Reward is the reward function in the overall architecture which helped to build the objective function. The three components work together to train the model to pursue maximum return.

4.5 Network Structure

The convolutional neural network structure was designed by Amrouni et al [15]. Motivated by the work in [15], this research also studied recurrent network, and a mixed network with RNN+CNN. The network structure design for the recurrent neural network and the mixed network is a new attempt for the training process which will be introduced in the next section.

4.5.1 Convolutional Neural Network. The detailed parameters for the convolutional neural network model are listed in Table 1. The price tensor starts in a shape of $[None, nb\_feature, m, n]$, where $nb\_features$ represents number of features, m represents the number of stocks and cryptocurrency, n is the length of tensor. It is then transposed to $[None, n, m, nb\_features]$ where it is used as the input for the convolutional neural network.

The policy network which uses convolutional neural network comprises input layer and three convolution layers where conv1 extracted and integrated effective features to 2 kernels, conv2 kept on the process and output 20 kernels, and in the end the conv3 will output the last feature map, which concatenated with the cash bias. After the concatenation with the cash bias, the tensor got squeezed, and in the end, softmax function is used to get the probability distribution which is also the weight distribution of the portfolio. The convolutional neural network model is shown in Figure 9. Batch size was set to 50 in this research, which is a common number picked by most researchers and it can also be changed to another value.

The output of the network is then used to calculate the adjusted reward. At first it needs to calculate the instantaneous reward as shown in the formula below:

$$InstantaneousReward = \frac{PortfolioValue - PrevPortfolioValue}{PrevPortfolioValue}$$

In the formula, the trading cost is also taken into consideration when the portfolio value is calculated.

Considering the formulation of the objective function, since there is no maximization function for the reward, minimizing the negative of the original value is equivalent to maximizing the original value; therefore, minimize (-adjusted reward) was chosen to maximize the reward over the batch.

Table 1: Parameters for convolutional neural network model

Input	(?,4,10,50)	
Conv1	(?,50,10,2)	activation=tf.nn.relu, filters=2, strides= (1, 1), kernel_size=(1,3), padding='same'
Conv2	(?,1,10,20)	activation=tf.nn.relu, filters=20, strides= (50, 1), kernel_size= (1, 50), padding='same'
Conv3	(?,1,10,1)	activation=tf.nn.relu, filters=1, strides= (21, 1), kernel_size= (1, 1), padding='same'
Tensor4	(?,1,11,1)	Concatenate conv3 with cash_bias
Squeezed_tensor	(?,11)	
Action	(?,11)	tf.nn.softmax

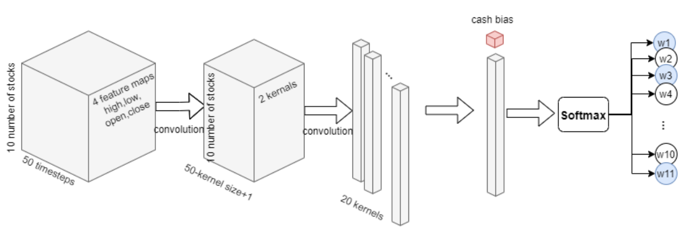


Figure 9: Convolutional neural network model

4.5.2 Recurrent Neural Network. The detailed parameters for the recurrent neural network model are listed in Table 2. The tensor which is the input to the network has the shape of (None, m, n), where *m* is the number of stocks and *n* is the length (time step) of the tensor. It has two recurrent layers. For the first recurrent layer, the number of recurrent units was set to 20, and it came from the inspiration from figure 4.9's CNN of filters=20. For the second recurrent layer, the number of the recurrent units was set to 1. The output dimension after the two recurrent layers was expanded to 4D. And after that it can be concatenated with a 4D-shape cash-bias. The following steps such as to squeeze the tensor, to use softmax to get the probability distribution, and to calculate the reward are the same as in the convolutional neural network. The recurrent neural network model is shown in Figure 10.

Table 2: Parameters for recurrent neural network model

Input	(?,10,50)	
Rnn1	(?,10,20)	num_units=20, activation=tf.nn.relu
Rnn2	(?,10,1)	num_units=1, activation=tf.nn.relu
Outputs(expand dimension)	(?,1,10,1)	
Tensor4	(?,1,11,1)	Concatenate Outputs with cash_bias
Squeezed_tensor	(?,11)	
Action	(?,11)	tf.nn.softmax

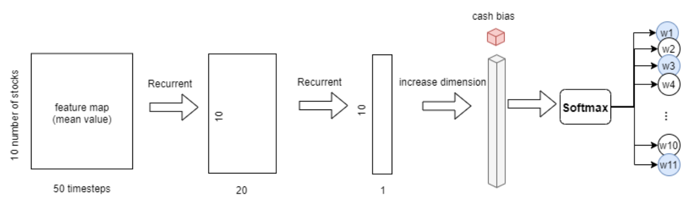


Figure 10: Recurrent neural network model

4.5.3 Mixed Neural Network. The mixed neural network model combines both the convolution layers and recurrent layers when processing the price tensor. Table 3 shows the details of the parameters. The input tensor has the same shape as the one of the recurrent neural networks and it is (None,10,50). First, it is passed to the recurrent layer, which has 20 recurrent units. The output of the recurrent layer's dimension is then expanded in order to be processed by the convolutional neural network. For the convolution layer, there is one filter. After the dimension expansion, all the following steps are the same as the convolution neural network's. The mixed neural network structure is as the Figure 11 below.

Table 3: Parameters for CNN+RNN model

Input	(?,10,50)	
Rnn1	(?,10,20)	num_units=20, activation=tf.nn.relu
Outputs(expand dimension)	(?,1,10,20)	
Conv3	(?,1,10,1)	activation=tf.nn.relu, filters=1, strides=(21, 1), kernel_size=(1, 1), padding='same'
Tensor4	(?,1,11,1)	Concatenate conv3 with cash_bias
Squeezed_tensor	(?,11)	
Action	(?,11)	tf.nn.softmax

4.5.4 Linear Regression Model. The training data is passed to the standard scaler first to be processed into the same scale.

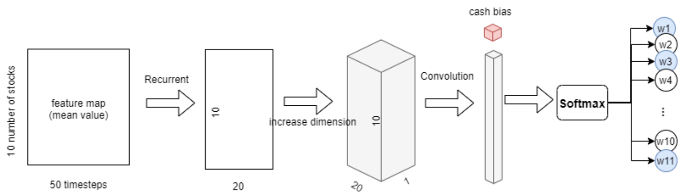


Figure 11: Mixed neural network model

Next, the processed data passes into the linear regression model in order to get the coefficient list which is also the weight list. The accumulated value for each value in the portfolio weight list should be 1, so the proportional value is calculated. The linear regression model is shown in Figure 12.

$$sum\_coef = |coef_1| + |coef_2| + ...|coef_{11}|$$

$$weight\_list = [\frac{|coef_1|}{sum\_coef}, \frac{|coef_2|}{sum\_coef} ... \frac{|coef_{11}|}{sum\_coef}]$$

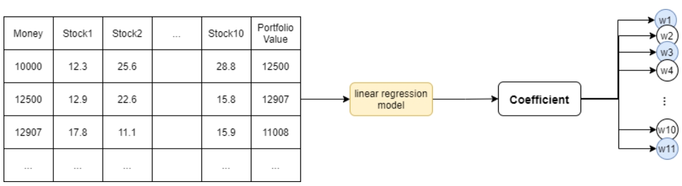


Figure 12: Linear regression model

4.6 Network Training Process

Portfolio Vector Memory was introduced by Jiang et al. in [9], and was implemented by Selim Amrouni in [10] for convolutional neural network model training. Figure 13 shows the CNN model training process. For each episode (total of 2 episodes in this paper), the PVM was initialized with training parameters. For each batch (total of 10 batches in this paper, 1 batch has 50 days), the starting point of the batch, initial portfolio value and initial weights were drawn from the PVM at time t-1. Based on the starting point, a price tensor which has the length of 50 was passed to the policy network to get the weight distribution. The weight distribution which is also the action wrote to the PVM at time t and made the agents step to the next day. The whole process recursively performed until all the batches and episodes ended.

This paper also implemented the PVM with the recurrent neural network model and the mixed RNN+CNN model. Figure 14 shows the process. The main difference between them is that for RNN and the mixed model, the price tensor retrieval and days traversal are separated. The starting point which was fed from the PVM was generated from the RNN pre-processed data, instead of generating the price tensor from CNN pre-processed data. CNN pre-processed data was here to help traverse the days. One price tensor which has length of 50 was retrieved and sent to the policy network to get the weight distribution. The remaining steps were similar as the CNN's.

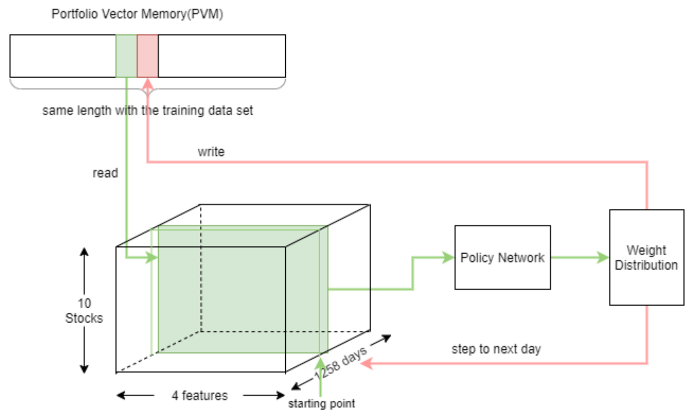


Figure 13: CNN training process

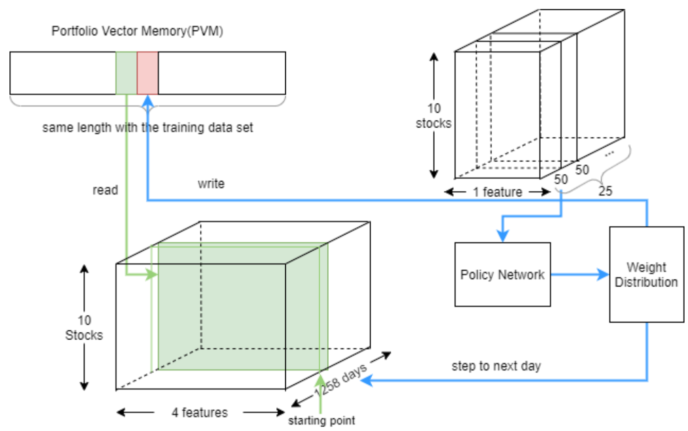


Figure 14: RNN and RNN+CNN training process

5 Results

5.1 Overall Results

Three different types of portfolio and three different networks were tested. For linear regression, the random selected portfolio was tested. The portfolio value evolution, min and max value, which represents minimum and maximum portfolio value, and the weight evolution during testing are presented in the following subsections. For testing, result figures are presented in the next subsections, except three deep reinforcement learning agents, where are other financial portfolio management agents being tested. Equal-weighted agent is an agent used for all the financial assets. Initial investment is 10000 dollars which acts as a baseline to show a clear performance comparison. For the other "full stock AAL, DAL, LUV..." means the money fully invest on the single stock.

5.1.1 Negatively Influenced Portfolio. Figure 15 shows the negatively influenced portfolio value evolution for 253 days. For both three types of deep reinforcement learning agents, there exist certain periods in which portfolio value is greater than the initial investment. But overall, they do not show an increasing trend. Among the three models, RNN model has the best performance(most return) at the end of the testing day, however it still loses 45.58% off the initial investment. Table 4 shows the maximum portfolio value and the minimum portfolio value for testing data set, CNN reached 10749 dollars which is the maximum portfolio value compared to the other two deep reinforcement learning agents. Tables 5, 6 and 7 shows the weight evolution every 50 days during testing for these three deep reinforcement learning agents.

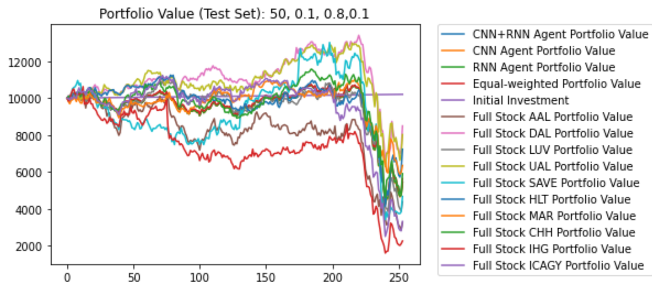


Figure 15: Negatively influenced portfolio

Table 4: Negatively influenced portfolio

	RNN	CNN	RNN+CNN
MAX	10744.384170389269	10749.582654440033	10746.332953411205
MIN	4296.119453575351	4257.566520880737	4282.784544420906

Table 5: Weight evolution during testing of CNN network

	Money	AAL	DAL	LUV	UAL	SAVE	HLT	MAR	CHH	IHG	ICAGY	PortfolioValue
Start	1	0	0	0	0	0	0	0	0	0	0	10000
50days	0.04202312	0.09757042	0.09566913	0.09583758	0.09606861	0.09529205	0.09550374	0.09521315	0.0962307	0.09488089	0.09571061	9869
100days	0.04302808	0.09447327	0.09794327	0.09643734	0.09649739	0.09397549	0.0966779	0.09640097	0.09637725	0.09359015	0.09459889	9043
150days	0.04197768	0.09555684	0.09728851	0.09613146	0.0950163	0.09509232	0.09558895	0.0955992	0.09481207	0.09690978	0.09602688	10111
200days	0.04253444	0.09510022	0.09617578	0.09571166	0.09539249	0.09750834	0.09640942	0.09539837	0.09467928	0.09564289	0.09544712	10471
250days	0.04445002	0.09371601	0.10153517	0.09098017	0.09698123	0.09955532	0.093565	0.09537336	0.10010976	0.09723239	0.08650158	4815

5.1.2 Positively Influenced Portfolio. Figure 16 shows the positively influenced portfolio value evolution for 253 days. Due to the explosive increasing trend for the full stock on BABA portfolio value, the graph does not show a clear trend for the

Table 6: Weight evolution during testing of RNN network

	Money	AAL	DAL	LUV	UAL	SAVE	HLT	MAR	CHH	IHG	ICAGY	PortfolioValue
Start	1	0	0	0	0	0	0	0	0	0	0	10000
50days	0.05115532	0.0966403	0.09475713	0.09492398	0.09515281	0.09438365	0.09459332	0.0943055	0.09531335	0.09397641	0.09479822	9870
100days	0.05236406	0.09355161	0.09698776	0.09549653	0.09555599	0.09305869	0.09573473	0.09546051	0.09543702	0.0926771	0.09367601	9052
150days	0.05109973	0.09464697	0.09636215	0.09521612	0.09411158	0.09418687	0.09467877	0.09468893	0.0939093	0.09598703	0.09511254	10111
200days	0.05176996	0.0941829	0.09524809	0.09478844	0.09447235	0.09656779	0.09547947	0.09447817	0.09376602	0.09472033	0.09452645	10469
250days	0.05408524	0.09277103	0.10051135	0.09006277	0.09600332	0.09855146	0.09262155	0.09441167	0.09910031	0.09625195	0.08562935	4854

Table 7: Weight evolution during testing of RNN+CNN network

	Money	AAL	DAL	LUV	UAL	SAVE	HLT	MAR	CHH	IHG	ICAGY	PortfolioValue
Start	1	0	0	0	0	0	0	0	0	0	0	10000
50days	0.04798819	0.09696288	0.09507342	0.09524083	0.09547042	0.09469869	0.09490906	0.09462028	0.09563149	0.09429009	0.09511465	9870
100days	0.04912673	0.0938712	0.09731909	0.09582276	0.09588243	0.09337659	0.09606178	0.09578662	0.09576306	0.09299371	0.09399602	9049
150days	0.0479361	0.09496253	0.09668343	0.09553357	0.09442535	0.09450089	0.09499443	0.09500462	0.09422239	0.09630705	0.09542965	10111
200days	0.04856724	0.09450101	0.0955698	0.0951086	0.09479144	0.09689396	0.09580196	0.09479728	0.09408273	0.09504026	0.09484572	10470
250days	0.05074487	0.09309864	0.10086629	0.09038082	0.09634235	0.09889948	0.09294863	0.09474508	0.09945027	0.09659185	0.08593173	4841

other agents. Actually, for the three types of deep reinforcement learning agents, during more than half of the testing days, the portfolio value seldom exceeds the initial investment, but at the end of the testing days, all agents show an increasing trend. For CNN model which has the best performance (most return) at end of the testing days, it gains 110.87% on the initial investment. Table 8 shows the maximum portfolio value and the minimum portfolio value for the testing data set, CNN reached 22784.7 dollars which is the maximum portfolio value compared to the other two deep reinforcement learning agents. Tables 9, 10 and

11 show the weight evolution during testing for these three deep reinforcement learning agents.

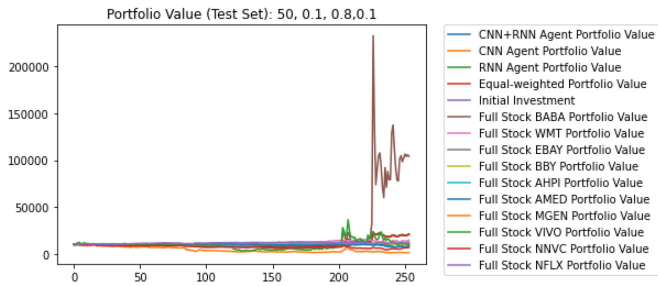


Figure 16: Positively influenced portfolio

Table 8: Positively influenced portfolio

	RNN	CNN	RNN+CNN
MAX	22696.510164433843	22784.707536769543	22604.74045890975
MIN	7621.239917016383	7607.848561432198	7635.092109920798

Table 9: Weight evolution during testing of CNN network

	Money	BABA	WMT	EBAY	BBY	AHPI	AMED	MGEN	VIVO	NNVC	NFLX	PortfolioValue
Start	1	0	0	0	0	0	0	0	0	0	0	10000
50 days	0.04045456	0.09421504	0.0965675	0.09705374	0.09904441	0.09744479	0.09691681	0.09784962	0.08812246	0.09594888	0.09638219	9509
100days	0.04151763	0.09493009	0.09508458	0.09717265	0.09681507	0.09684529	0.09672782	0.09159769	0.09385134	0.09725982	0.09819801	8612
150days	0.0406314	0.09541863	0.09672963	0.09640315	0.09806951	0.09753956	0.09593144	0.08993582	0.09915782	0.09297967	0.09720337	8168
200days	0.04009868	0.09856441	0.09598776	0.09656359	0.09593668	0.09655176	0.09554054	0.08520287	0.1065174	0.09353385	0.09550245	9840
250days	0.04078758	0.10094165	0.09346066	0.09511324	0.09493542	0.09390534	0.09331918	0.09932903	0.08692816	0.10371155	0.09756818	19948

Table 10: Weight evolution during testing of RNN network

	Money	BABA	WMT	EBAY	BBY	AHPI	AMED	MGEN	VIVO	NNVC	NFLX	PortfolioValue
Start	1	0	0	0	0	0	0	0	0	0	0	10000
50 days	0.04617789	0.09365308	0.09599151	0.09647485	0.09845364	0.09686356	0.09633874	0.09726599	0.08759685	0.09537658	0.0958073	9512
100days	0.04738496	0.09434898	0.09450253	0.09657781	0.09622242	0.09625245	0.0961357	0.09103698	0.09327683	0.09666445	0.0975969	8621
150days	0.04638161	0.09484672	0.09614986	0.09582533	0.09748171	0.09695493	0.09535646	0.08939677	0.09856349	0.09242237	0.09662076	8179
200days	0.04577628	0.09798143	0.09542002	0.09599244	0.09536923	0.09598068	0.09497544	0.08469892	0.10588738	0.09298062	0.09493757	9843
250days	0.04656188	0.100334	0.09289805	0.09454068	0.09436393	0.09334004	0.09275742	0.09873109	0.08640487	0.10308722	0.09698083	19888

Table 11: Weight evolution during testing of CNN+RNN network

	Money	BABA	WMT	EBAY	BBY	AHPI	AMED	MGEN	VIVO	NNVC	NFLX	PortfolioValue
Start	1	0	0	0	0	0	0	0	0	0	0	10000
50 days	0.05211806	0.09306983	0.0953937	0.09587403	0.0978405	0.09626032	0.09573876	0.09666024	0.08705131	0.0947826	0.09521064	9515
100days	0.05347262	0.09374604	0.09389861	0.09596063	0.09560751	0.09563735	0.09552135	0.09045521	0.09268075	0.09604672	0.0969732	8630
150days	0.05234932	0.09425317	0.09554816	0.09522566	0.09687167	0.09634819	0.09475972	0.08883733	0.09794668	0.091844	0.09601611	8190
200days	0.05166957	0.09737629	0.0948307	0.09539959	0.09478023	0.0953879	0.09438887	0.08417582	0.10523341	0.09240637	0.09435123	9846
250days	0.05255428	0.09970339	0.09231418	0.09394649	0.09377084	0.0927534	0.09217443	0.09811056	0.08586181	0.10243931	0.09637131	19825

5.1.3 Randomly Selected Portfolio. Figure 17 shows the randomly selected portfolio value evolution for 100 days. For the three types of deep reinforcement learning agents, they all show a stable increasing trend. The deep reinforcement learning agents have the best performance on this type of portfolio compared to the other two portfolios. The RNN model, which has the best performance (most return) gained 47.72% initial investment for 253 days. Table 12 shows the maximum portfolio value and the minimum portfolio value during testing, RNN reached 14772 dollars which is the maximum portfolio value compared to the other two deep reinforcement learning agents. Tables 13, 14 and 15 show the weight evolution during testing for these three deep reinforcement learning agents.

5.1.4 Linear Regression on Randomly Selected Portfolio. Since randomly selected portfolio shows a stable increasing performance, we tested linear regression using the randomly selected portfolio. The linear regression agent's result is compared with the equal-weighted portfolio value and the initial investment in Figure 18. From the comparison, it can be seen

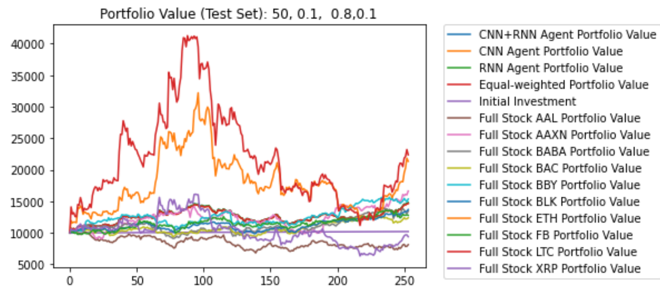


Figure 17: Randomly selected portfolio

Table 12: Randomly selected portfolio

	RNN	CNN	RNN+CNN
MAX	14772.48989557328	14679.754959892462	14764.415960353592
MIN	10000	10000	10000

that linear regression model shows a bad performance which has a clear downward trend. The weight evolution during testing is shown in Figure 19, the weight adjustment during testing is volatile for the linear regression agent.

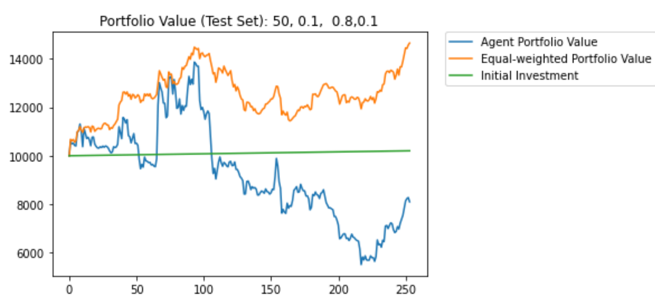


Figure 18: Linear regression on random selected portfolio

5.2 Portfolio Comparison Results

In this section, the performance results from the three different types of portfolios are discussed. Figure 20 shows the CNN model results for the three types of portfolio. Figure 21 shows the RNN model's results, and Figure 22 shows the RNN+CNN model's results. Among the three models, the negatively selected portfolio value presents a decreasing trend, while the positively selected portfolio value presents an explosive increasing trend around day 200, and the randomly selected portfolio presents an overall stable state with positive returns.

5.3 Activation Functions

The performances of activation functions were compared using random selected portfolio. Three different activation functions were compared: sigmoid, ReLU and tanh. *X* axis shows the 253 days, and the *y* axis is the portfolio value. Figure

Table 13: Weight evolution during testing of CNN network

	Money	AAL	AAXN	BABA	BAC	BBY	BLK	ETH	FB	LTC	XRP	PortfolioValue
Start	1	0	0	0	0	0	0	0	0	0	0	10000
50days	0.07450122	0.09033511	0.09407107	0.09289475	0.09233032	0.09176801	0.09219527	0.09322398	0.09446668	0.09161997	0.09259363	12429
100days	0.07540939	0.09131607	0.09225253	0.09355897	0.09278396	0.09257833	0.09294329	0.09269756	0.09382129	0.09080885	0.09182977	14157
150days	0.07423953	0.09346664	0.09401267	0.0927813	0.09427755	0.09246214	0.09293781	0.09294169	0.09060269	0.09087423	0.09140374	12331
200days	0.07616618	0.0939748	0.09440434	0.09629674	0.09472538	0.09163484	0.09343745	0.08661038	0.09466618	0.08670246	0.09138125	12397
250days	0.07331774	0.09164177	0.0896015	0.08934477	0.09293649	0.09313318	0.09140766	0.09780822	0.08983148	0.09658493	0.09439227	14475

Table 14: Weight evolution during testing of RNN network

	Money	AAL	AAXN	BABA	BAC	BBY	BLK	ETH	FB	LTC	XRP	PortfolioValue
Start	1	0	0	0	0	0	0	0	0	0	0	10000
50days	0.05744246	0.09200017	0.09580498	0.09460698	0.09403215	0.09345947	0.09389461	0.09494228	0.09620789	0.0933087	0.09430031	12477
100days	0.058157	0.09301998	0.09397392	0.09530473	0.09451527	0.09430579	0.09467756	0.09442725	0.09557194	0.09250329	0.09354326	14243
150days	0.05723598	0.09518336	0.09573942	0.09448543	0.09600916	0.0941604	0.09464481	0.09464876	0.0922668	0.09254333	0.09308257	12372
200days	0.05874964	0.09574646	0.09618409	0.09811217	0.09651118	0.09336238	0.09519897	0.0882432	0.09645087	0.08833702	0.09310402	12439
250days	0.05651004	0.09330393	0.09122665	0.09096526	0.09462212	0.09482238	0.09306556	0.09958222	0.0914608	0.09833674	0.09610431	14563

23 shows the RNN activation results. Figure 24 shows the combined model activation result, and Figure 25 shows the CNN activation results.

5.4 Optimizer

The performances of various optimizers are compared using random selected portfolio. Four types of optimizers are compared: Adam, RMSProp, Gradient Descent and Adagrad. *X* axis shows the 253 days, and *y* axis is the portfolio value. For all the models, Adam optimizer and Gradient Descent optimizer

Table 15: Weight evolution during testing of RNN+CNN network

	Money	AAL	AAXN	BABA	BAC	BBY	BLK	ETH	FB	LTC	XRP	PortfolioValue
Start	1	0	0	0	0	0	0	0	0	0	0	10000
50 days	0.05896482	0.09185157	0.09565025	0.09445417	0.09388027	0.09330852	0.09374295	0.09478894	0.0960525	0.093158	0.09414801	12472
100days	0.05969696	0.09286789	0.09382027	0.0951489	0.09436073	0.0941516	0.09452276	0.09427286	0.09541568	0.09235205	0.09339031	14236
150days	0.05875333	0.09503016	0.09558533	0.09433335	0.09585463	0.09400885	0.09449248	0.09449642	0.0921183	0.09239438	0.09293275	12369
200days	0.06030452	0.09558829	0.0960252	0.0979501	0.09635175	0.09320815	0.09504171	0.08809743	0.09629154	0.08819109	0.09295021	12435
250days	0.05800961	0.09315563	0.09108165	0.09082068	0.09447173	0.09467167	0.09291765	0.09942394	0.09131543	0.09818044	0.09595156	14555

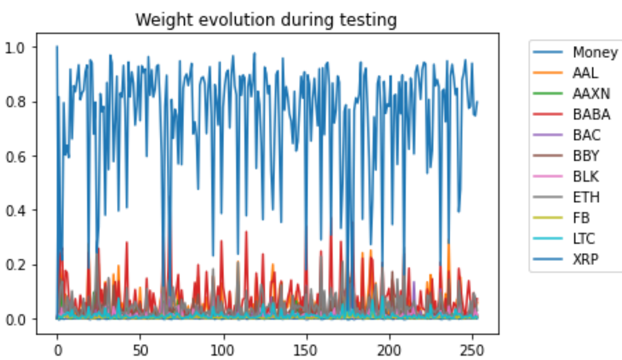


Figure 19: Weight evolution of linear regression

showed the best result. Figure 26 shows the RNN activation results, Figure 27 shows the combined model activation result, and Figure 28 shows the CNN activation results.

5.5 Learning Rate

The performances of different learning rates are also compared. Four different learning rates with 0.1, 0.01, 0.001 and 0.0001 were calculated. Figure 29 shows the RNN learning rate results. Figure 30 shows the combined model learning rate result, and Figure 31 shows the CNN learning rate results.

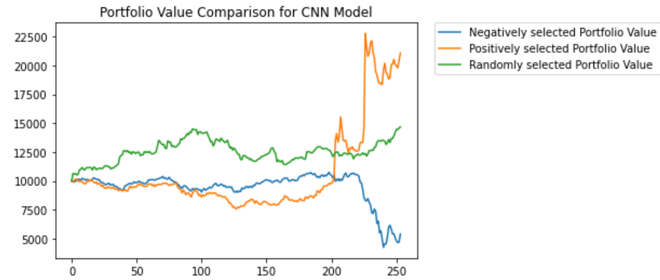


Figure 20: CNN model results

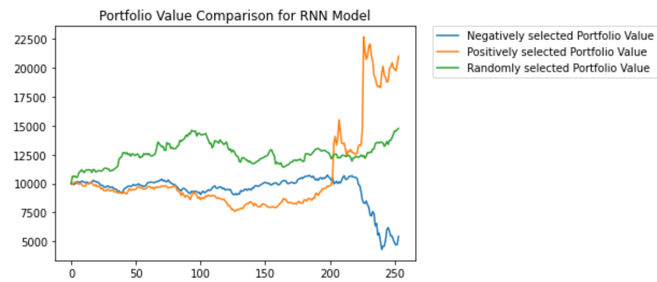


Figure 21: RNN model results

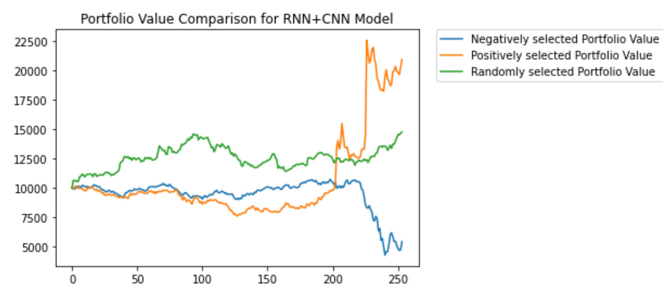


Figure 22: RNN+CNN model results

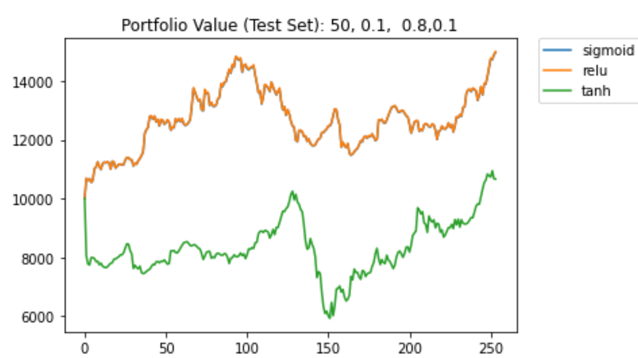


Figure 23: RNN results with activation functions

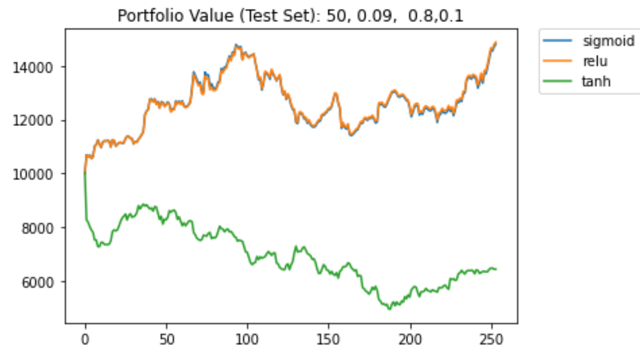


Figure 24: RNN+CNN results with activation functions

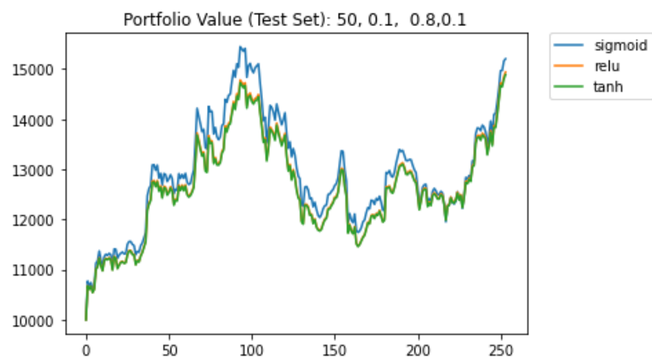


Figure 25: CNN results with activation functions

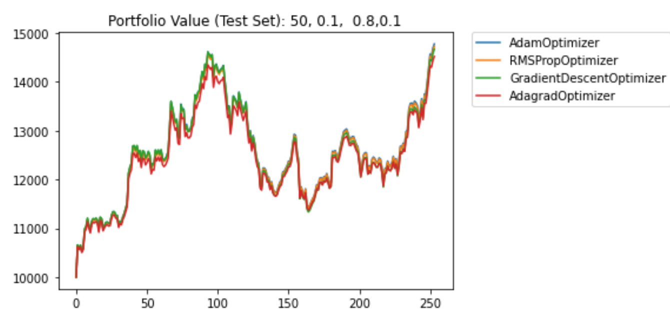


Figure 26: RNN results with different optimizers

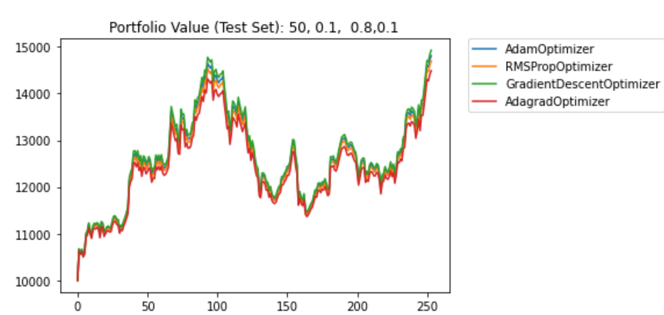


Figure 27: RNN+CNN results with different optimizers

Figure 28: CNN results with different optimizers

Figure 29: RNN results with different learning rates

Figure 30: RNN+CNN results with different learning rates

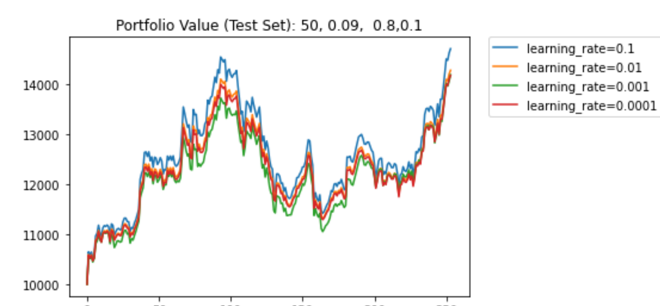


Figure 31: CNN results with different learning rates

6 Conclusion

This paper implemented deep reinforcement learning method to allocate weights for stocks and cryptocurrency in portfolio management and compared different policy network models. The training results from various policy networks constructed with convolutional neural network, recurrent neural network, or the mixed one show similar results. In this research, three different portfolios were considered, and they include stocks that have been negatively influenced by the Covid-19, and the stocks that have been positively influenced by Covid-19, and the stocks and cryptocurrency that were randomly selected. Among the three portfolios, it was found that all the agents have the best performance on the randomly selected portfolio, which presents overall stable moving-up portfolio values as time goes on. The average return rate for all three deep reinforcement learning agents reached 47.38%. It is known that the financial market is filled with uncertainties and unpredictability due to many different factors, so the training results from this research also show the instability when dealing with positively influenced portfolio and negatively influenced portfolio. The linear regression model for portfolio optimization shows poor results, but the weight allocation changed more when compared with deep reinforcement learning agents. Overall, the linear regression model does not meet the goal of financial portfolio management. On the other hand, the deep reinforcement learning agents show a better performance when cryptocurrency was added to the portfolio.

This paper considers that cryptocurrency does not have a long history, so only 5 years data were used. For future studies, data from 10 or more years can be collected to research the performance of deep reinforcement learning agents. The paper only applied to policy gradient in reinforcement learning; in the future, deep Q network and Deep Deterministic Policy Gradient can also be added to do a comparison study with the agents in the paper. For financial fields, there are many financial portfolio optimization methods, but in this research only equal weighted allocation strategy was used to compare with deep reinforcement learning agents. To do a further study, other portfolio optimization methods can also be applied and tested.

References

[1] A. Jonsson, "Deep Reinforcement Learning in Medicine," *Kidney Diseases*. 5:18-22, 2019.

[2] A. Tianello, *Using a Multi-Factor Regression Model to Determine Asset Allocation*, https://ufdcimages.uflib.ufl.edu/AA/00/05/75/50/00001 /altianello-Thesis.pdf.

[3] C. Adjasi and N. Biekpe, "Stock Market Development and Economic Growth: The Case of Selected African Countries," *African Development Review*, 18:144-161, 2006.

[4] G. Ban, N. El Karoui, and A. Lim, "Machine Learning and Portfolio Optimization," *Management Science*, 64:1136-1154, 2018.

[5] G. Lample, and D. Chaplot, "Playing FPS Games with Deep Reinforcement Learning," *Thirty-First AAAI Conference On Artificial Intelligence, San Francisco, California, USA*, 2017.

[6] J. Pineiro-Chousa, M. Vizcaiino-Gonzalez, and A. Perez-Pico, "Influence of Social Media over the Stock Market," *Psychology and Marketing*, 34: 101-108, 2017.

[7] L. Kaiser, M. Babaeizadeh, P. Milos, B. Osinski, R. Campbell, K. Czechowski, D. Erhan, C. Finn, P. Kozakowski, and S. Levine, "Model-Based Reinforcement Learning for Atari," *ArXiv Preprint ArXiv:1903.00374*, 2019.

[8] L. Schultz, and V. Sokolov, "Deep Reinforcement Learning for Dynamic Urban Transportation Problems," *ArXiv Preprint ArXiv:1806.05310*, 2018.

[9] L. Wang, and A. Kutan, "The Impact of Natural Disasters on Stock Markets: Evidence from Japan and the US," *Comparative Economic Studies*, 55:672-686, 2013.

[10] N. Kanwar, *Deep Reinforcement Learning-based Portfolio Management,* https://rc.library.uta.edu/uta-ir/bitstream/handle/10106/28108/KANWAR-THESIS-2019.pdf?sequence=1&isAllowed=y, 2019.

[11] N. Kumar, S. Rahman, and N. Dhakad, "Fuzzy Inference Enabled Deep Reinforcement Learning-Based Traffic Light Control for Intelligent Transportation System," *IEEE Transactions On Intelligent Transportation Systems*. 22:4919-4928, 2020.

[12] P. Henderson, R. Islam, P. Bachman, J. Pineau, D. Precup, and D. Meger, "Deep Reinforcement Learning that Matters," *Proceedings Of The AAAI Conference On Artificial Intelligence, Louisiana, USA*, 392:3207-3214, 2018.

[13] P. Santa-Clara, and R. Valkanov, "The Presidential Puzzle: Political Cycles and the Stock Market," *The Journal Of Finance*, 58:1841-1872, 2003.

[14] P. Yu, J. Lee, I. Kulyatin, Z. Shi, and S. Dasgupta, "Model-Based Deep Reinforcement Learning for Dynamic Portfolio Optimization," *ArXiv Preprint ArXiv:1901.08740*, 2019.

[15] S. Amrouni, A. Moulin, and P. Mizrahi, *A Deep Reinforcement Learning Framework for the Financial Portfolio Management Problem*, https://selimamrouni.github.io/portfolio/air-elle.html, 2018.

[16] S. Bartram, G. Brown, and R. Stulz, "Why are US Stocks more Volatile?" *The Journal Of Finance*, 67:1329-1370, 2012.

[17] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing Atari with Deep Reinforcement Learning," *ArXiv Preprint ArXiv:1312.5602*, 2013.

[18] X. Liu, Z. Ding, S. Borst, and A. Walid, "Deep Reinforcement Learning for Intelligent Transportation Systems," *ArXiv Preprint ArXiv:1812.00979*, 2018.

[19] Yahoo Finance. https://finance.yahoo.com/.

[20] Y. Liu, B. Logan, N. Liu, Z. Xu, J. Tang, and Y. Wang, "Deep Reinforcement Learning for Dynamic Treatment Regimes on Medical Registry Data," *2017 IEEE International Conference On Healthcare Informatics (ICHI), Park City, Utah, USA*, pp. 380-385, 2017.

[21] Y. Zhang, P. Zhao, Q. Wu, B. Li, J. Huang, and M. Tan, "Cost-Sensitive Portfolio Selection via Deep Reinforcement Learning," *IEEE Transactions On Knowledge And Data Engineering*, 34: 236-248, 2020.

[22] Z. Liang, H. Chen, J. Zhu, K. Jiang, and Y. Li, "Adversarial Deep Reinforcement Learning in Portfolio Management," *ArXiv Preprint ArXiv:1808.09940*, 2018.

[23] Z. Liang, K. Jiang, H. Chen, J. Zhu, and Y. Li, "Deep Reinforcement Learning in Portfolio Management," *ArXiv Preprint ArXiv:1808.09940*, 2018.

[24] Z. Liu, C. Yao, H. Yu, and T. Wu, "Deep Reinforcement Learning with its Application for Lung Cancer Detection in Medical Internet of Things," *Future Generation Computer Systems*. 97: 1-9, 2019.

[25] Z. Jiang, D. Xu, and J. Liang, "A Deep Reinforcement Learning Framework for the Financial Portfolio Management Problem," *ArXiv Preprint ArXiv:1706.10059*, 2017.

[26] Z. Jiang and J. Liang, "Cryptocurrency Portfolio Management with Deep Reinforcement Learning," *2017 Intelligent Systems Conference (IntelliSys), London, United Kingdom*, pp. 905-913, 2017.

Yue Ma is a PhD student at Northern Illinois University, where she is pursuing her passion for computer science. She received her Master's degree from Southeast Missouri State University, bringing a wealth of knowledge and experience to her studies. Her research interests include data visualization, augmented reality (AR), and machine learning. She is dedicated to furthering her knowledge, and actively working towards making more publications and contributing to her research fields in meaningful ways.

Ziping Liu received her PhD in Engineering Science from Southern Illinois University at Carbondale in 1999 and began her computing career at Motorola, where she developed software for mobile phones. Currently, she is a professor of Computer Science at Southeast Missouri State University, where she has been teaching since 2001. Her research interests encompass a wide range of topics, including machine learning, cloud computing, secured software design, wireless ad hoc networks and sensor networks, distributed computing and game development.

Charles D. McAllister is currently a professor of business analytics at Southeast Missouri State University. He has previously served the university in such roles as vice provost and dean of graduate studies, interim dean of business, interim chairperson of computer science, and interim director of institutional research. He received a Ph.D. in industrial engineering and operations research from The Pennsylvania State University in 2002, M.S. in industrial engineering from Purdue University in 1999, and B.S. in industrial and management systems engineering from the University of Nebraska in 1998. His research interests include optimization, analytics, simulation, pedagogy, and higher education administration.

Polygon Formation in Distributed Multi-Agent Systems

Rui Yang*, Azad Azadmanesh*, and Hassan Farhat*
University of Nebraska, Omaha, NE USA

Abstract

A primary challenge to form a geometric pattern for a fleet of randomly distributed agents is the determination of their locations on the boundary of a formation. The challenge is more acute if the agents need to be uniformly distributed on the formation. This study addresses this challenge for polygon formations via a two-phase procedure. In the first phase, the agents form an enclosing circle to the location of the formation. This prevents the agents from location-conflicts on the polygon and hence each agent can uniquely ascertain its projection point on the formation. In the second phase, the agents establish their projected points and move toward their locations on the polygon, while mitigating collisions. The circle formation is also used as a regrouping feature before the agents reconfigure themselves into a different polygon formation. The formation control laws developed are verified through simulation for circle formation, convex polygons, and some categories of concave polygons. The control laws include the cases for rotation, translation (relocation), and scaling of polygons as well.

Key Words: Consensus, control systems, cyber physical systems, formation control, multi-robot systems, peer-to-peer networks.

1 Introduction

The research in networked multi-agent systems (MASs) has flourished in recent years due to their potential in various applications [2, 3, 12, 19, 20]. Some examples are cooperative control of unmanned aerial vehicles (UAVs), autonomous underwater vehicles (AUVs), and surveillance and reconnaissance missions to accomplish a common goal [4, 9, 18, 22, 24]. A mainstream of research in MASs is on the structure (formation) of agents in a distributed manner. More specifically, each agent following a control mechanism communicates with its neighbors only. Therefore, the view of the field by each agent is local not global. Through the local interactions, the agents can form a particular geometric pattern [16, 23].

The three common approaches to formation controls are: leader-follower, behavioral, and virtual structure. These approaches may also be combined with artificial intelligence approaches for improving flexibility and performance. In the leader-follower approach [4, 15], some agents act as leaders and the rest as following the leaders. The leaders' trajectories (e.g., a formation pattern) are transformed by the followers into coordinates under local control laws (e.g., keeping a certain distance from the leaders). Although the salient feature of this approach is its simplicity, the leader-follower approach suffers from having single points of failures. For example, if a leader fails, the formation becomes difficult. In addition, there is no feedback from the followers to the leaders, e.g., a follower is not able to inform the leader (s) if it runs into an obstacle. In the behavioral approach [10, 24], the agents follow a set of predefined control laws to control their behavior in certain conditions such as avoiding collision, avoiding obstacles, and keeping a certain distance from the neighbors. An advantage of this approach is its flexibility in systems with a substantial number of agents. A disadvantage is the complexity of mathematical analysis to create control laws to guarantee precise formation.

In the virtual structure formation [11, 24], the structure is treated as a virtual rigid structure, e.g., a circle. The agents' positions are determined according to reference points on the rigid structure. If the agents can track the reference points, the formation can be preserved. In addition, if the agents can follow their own specific reference points, the precise formation of the structure can be maintained. Several studies have focused on the same formation since it becomes difficult to reconfigure the formation into a different virtual structure, particularly if reconfiguration is needed often.

This study borrows elements from the behavioral and virtual structures to form polygons. The behavioral approach is utilized to formulate control laws for traveling and avoiding collisions based on local interaction with neighbors. Virtual structures are embedded by points that are formed into polygons. Agents travel to the coordinates of these points. The agents determine the coordinates on the polygons individually and in a distributed manner.

To form a polygon, a two-phase approach is employed. In the first phase, the agents are randomly distributed in a field. Using the local control laws, the agents will then form a circle circumscribing the location of the polygon structure. The agents will also have the option to uniformly distribute themselves on the circle. In the second phase, each agent identifies the coordinates of its projected point on the polygon. Taking advantage of linear function properties, an agent's projected point is the intersection of its closest trajectory toward the circle center and the polygon.

The two-phase approach is adopted to make the reconfiguration into polygons simpler. The enclosing circle formation provides conflict free projected points. Besides, the

*E-mail: ryang@unomaha.edu, azad@unomaha.edu, hfarhat@unomaha.edu.

circle formation can be used as the pre-requisite to a number of symmetric formations and in general regular polygons, as will be seen later in the study. Noteworthy is that if the agents are uniformly distributed on the circle and the circle passes through the polygon vertices, the agents, can be uniformly distributed on the polygons as well.

This study is restricted to two-dimensional space and is the continuation of the work done in [23]. The focus in [23] was on the first phase, i.e., on circle formation, whereas this study is mostly concerned with the second phase. Section 2 sets the network model and provides the background information on circle formation based on [23], in addition to some research works on polygon formations. Section 3 introduces some perspectives on polygons and the approaches defining the circle that circumscribes a polygon. Section 4 lays out the foundation on polygon formations. It discusses the general approach for various convex and concave polygon formations, which can be applied to the classic regular polygons such as squares and equilateral triangles as well. Section 5 discusses polygon transformations such as rotation, relocation (translation), and scale adjustment of polygons while maintaining their formation. Section 6 is about the simulation of the results in Sections 4 and 5. The section includes some examples of reconfiguration and transformation of polygons. Section 7 concludes the study with a summary and some avenues for future studies.

2 Background

The subject of pattern formations and in general formation control has been investigated in numerous studies. A number of these studies have been devoted to polygon formations. Among these, [7] proposes an algorithm for a group of homogenous robots that gradually form into patterns such as polygons and circles. Through local interactions and differentiating tasks that each robot plays, the authors have shown that the robots can start making simple patterns such as a line that gradually turns into more complex patterns like a circle and then into polygon patterns. Although the authors claim that it is theoretically possible to generate more complex patterns, the approach is a time-consuming process, e.g., allowing the robots to find each other and forming themselves into a line from randomly distributed robots.

The focus of [13] is on pattern formation via machine learning, specifically using the Q-learning algorithm. In this process, the agents are rewarded based on their actions. An agent is rewarded with the highest reward once it reaches its final target, and it is penalized if it moves away from the target. The authors use six agents to form the vertices of a hexagon. For polygons with lower sides such as a square or a triangle, some agents are removed. So, the number of agents must represent the number of vertex points, and not able to form polygons with the number of agents higher than the sides. The study does not allude to any discussion on collision or obstacle avoidance.

The study in [5] proposes a distributed control strategy to form regular polygons with a specified scale and with arbitrary number of agents. Polygon formations are achieved using local measurements. These measurements are relative and cyclic in the sense that an agent k's neighbors are agents $k-1$ and $k+1$. Under this sensing strategy, each agent moves toward the end point of a vector that is perpendicular to the midpoint of the line segment that is connecting the agent with its neighbor. Like [13], the agents forming the polygons are stationed only at the vertices of the polygons. So, if there are n agents, the formation will be a n-sided polygon. Additionally, the agents are treated as point agents, so collision avoidance is not considered.

In [8], a two-stage process for forming static polygons is presented. The approach is based on the leader-follower principle. In the first stage, the leader agent provides the orientation and the distance information for each agent with respect to itself (the leader). In the second stage, the robots are moved in circulation motion until they are uniformly distributed. In this stage, the distance between any agent and the leader, and the angle formed between any two neighbors with respect to the leader are updated repeatedly until the formation stabilizes. Like the previous discussions, the agents' final positions are at the vertices of the polygons.

Like the work in [5], the authors in [25] assume the sensing topology is cyclic, so that each agent has only two neighbors. No discussion is made as to how the sensing topology is established or can be established. In their approach, some external control input is injected to specific agents (vertex agents) in accordance with the desired formation. Since local measurements are used, the agents only need to keep their orientations with respect to their nearest neighbors.

Contrary to these studies, this study does not assume any predefined relationship among the agents, does not use any leader-follower strategy, does not assign any agents to the polygon vertices, and includes collision avoidance. In addition, the work herein specifically discusses regular versus irregular polygons, convex versus concave, and provides tangible insight into rotation, translation, and scaling of polygons. An advantage of the work is that most of the delay and collision avoidance operations take place during the first phase. Once the circle is formed, the delay and collisions in the polygon formation including polygon reconfigurations, rotation, translation, and scaling are negligible since the agents' transitions occur with the same relative distance from their neighbors. However, the overhead delay in the first phase could increase, e.g., more agents translates into more collision avoidance operations.

2.1 Circle Formation

Studies on circle formation make various modeling assumptions such as global observation versus limited observation and collision free versus collision avoidance. The approach in [23] investigates the circle formation in a more formal fashion with realistic assumptions in mind. These assumptions include limited visibility, autonomous behavior, and collision avoidance while traveling to the final destinations. The only requirement, which is beyond the scope of this study, is to assume the agents can compute their positions using a coordinate positioning system.

In [23], the multi-agent system is a dynamic network in which the changes in the topology are caused by the mobility of the agents noted as $\{A_0, A_1, \ldots, A_n\}$, where A_0 is the virtual agent representing the center of the circle and n is the number of real agents. Two agents A_i and A_j are neighbors if they are within the sensing range of each other. The distance between the two agents is denoted as D_{ij}. Any pair of agents keep a minimum distance from each other to avoid collisions, which is called the collision distance (CD).

A two-dimensional coordinate system is used, in which the location of an agent A_i is (x_{a_i}, y_{a_i}) and its mobility at any time t is shown by its coordinates as $x_{a_i}(t)$ and $y_{a_i}(t)$. For simplicity, when there is no confusion, the time is dropped but assumed, e.g., as x_{a_i} and y_{a_i}.

The location of the circle center, as (x_c, y_c), needs to be agreed upon by an appropriate consensus protocol among agents. The circle center (x_c, y_c) might also be shown as (x_0, y_0). Once agreed on a circle center, no message communication to form the circle takes place. An agent A_i continuously travels toward the circle boundary on the shortest path if it is not in the collision path with any of its neighbors. The agent moves toward the circle by changing its coordinates according to the following:

$$x_{a_i}(t+dt) = x_{a_i}(t) + \Delta_d(r - D_{i0})\cos(Angle_{i0}(t)) \quad (1)$$

$$y_{a_i}(t+dt) = y_{a_i}(t) + \Delta_d(r - D_{i0})\sin(Angle_{i0}(t)) \quad (2)$$

where dt is a small-time value, Δ_d is a small positive value, r is the circle radius, and $Angle_{i0}$ is the angle formed at the circle center (x_0, y_0) with respect to agent A_i. In other words, the angle $Angle_{i0}$ has the vertex at (x_0, y_0) with the end-legs of the angle at the coordinates $(x_c + r, y_c)$ and (x_{a_i}, y_{a_i}). If an agent A_i is in collision with some neighbors A_j as it moves toward the circle, it repels from those agents by changing its coordinates according to the following:

$$x_{a_i}(t+dt) = x_{a_i}(t) - x_{a_i}^{dif}(t) \quad (3)$$

$$y_{a_i}(t+dt) = y_{a_i}(t) - y_{a_i}^{dif}(t) \quad (4)$$

where $x_{a_i}^{dif}$ is the sum of $(x_{a_j} - x_{a_i})/D_{ij}$ with respect to each agent A_j that is in collision course with. The $y_{a_i}^{dif}$ is defined similarly.

Once the agents are on the circle, they move counterclockwise on the circle keeping a distance called segment distance (SD) from each other. The SD is the size of the circle perimeter divided by the number of agents. Thus, the agents will be uniformly distributed on the circle. The SD can also be manually set if one wishes to change the scale of the circle. An agent A_i moves counterclockwise on the circle according to the following:

$$x_{a_i}(t+dt) = x_c(t) + r\cos\big(Angle_{i0}(t) + \Delta_\phi \phi\big) \quad (5)$$

$$y_{a_i}(t+dt) = y_c(t) + r\sin\big(Angle_{i0}(t) + \Delta_\phi \phi\big) \quad (6)$$

Since the circle center does not change with time, $x_c(t)$ and $y_c(t)$ can be replaced with x_c and y_c, respectively. The Δ_ϕ is a small percentage of ϕ, where $\phi = 360°$.

Because of the importance of (5) and (6) in the movement and rotation of agents discussed later, Figure 1 shows the process of obtaining (5) and (6) for better insight. The figure shows the situation where A_i at time t with its coordinates at $(x_{a_i}(t), y_{a_i}(t))$ would like to move to a new location $(x_{a_i}(t+dt), y_{a_i}(t+dt))$ on the circle at time $(t+dt)$ and make a change in its angle by $\Delta_\phi \phi$ (angular speed). At time t, the agent has the angle of $Angle_{i0}$ with respect to 0°. From the figure, since $r\cos\big(Angle_{i0}(t) + \Delta_\phi \phi\big)$ is the x-coordinate offset of the agent at time $(t+dt)$, adding the x-coordinate of the circle center results in the equation (5). Similarly, $r\sin\big(Angle_{i0}(t) + \Delta_\phi \phi\big)$ is the y-coordinate offset from the y-coordinate of the circle center, and so adding the offset to the circle center y-coordinate yields (6). It is important to realize that smaller values of $\Delta_\phi \phi$ provides for smoother transitions. Similarly, the transitions become smoother as dt is reduced, as opposed to discrete changes.

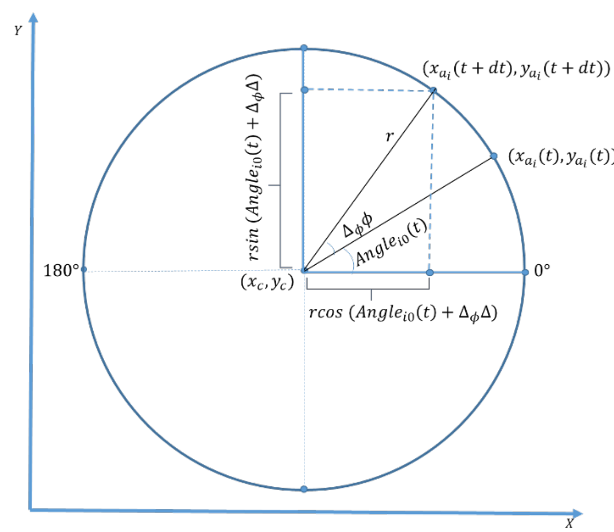


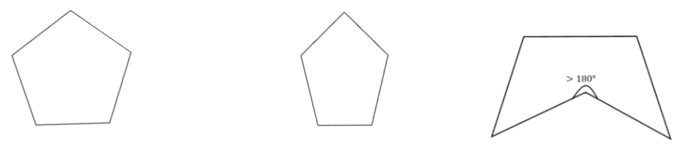
Figure 1: Travel of an agent A_i on the circle

For further detail on circle formation, the reader is referred to [23].

3 Preliminary Perspectives on Polygons

A polygon has a number of line segments that are connected to form a closed region. Polygons can be regular or irregular. In a regular polygon , every internal angle is of the same degree and every side is of the same length. Otherwise, the polygon is irregular. On the other hand, a polygon can be concave or convex. A concave polygon has at least one vertex pointed inward. In other words, it has an internal angle with a degree larger than 180°. Otherwise, the polygon is convex.

A concave polygon can never be regular because the interior angles are of different measures. Another property of a concave polygon is that a line passing through the polygon can touch the polygon more than twice. In contrast, a line can cross a convex polygon at most twice. Figure 2 shows an example of a five-sided (pentagon) polygon. Figure 2a and 2b are convex polygons, whereas Figure 2c is a concave polygon because one of its internal degrees is greater than 180°.



a) Convex regular b) Convex irregular c) Concave

Figure 2: Examples of polygon structures

3.1 Circle Enclosed Polygons

Since this study employs a two-stage approach to polygon formations, it is necessary to discuss the notion of a circle enclosing a polygon. The circle enclosing a polygon can be the Smallest Enclosing Circle (SEC) that covers the entire polygon in the field [17, 21]. A polygon is inscribed in the SEC if all the vertices of the polygon are on the circle. It can also be said that the SEC circumscribes a polygon if the circle passes through the vertices of the polygon.

An inscribed polygon is called a *cyclic polygon*. Not all polygons are cyclic. All regular polygons can be inscribed and are thus cyclic. Some examples of regular polygons are squares, equilateral triangles, and equilateral pentagons. An irregular polygon can be cyclic as well. For example, all rectangles are cyclic.

For a MAS, if the polygon formation is irregular, obtaining the SEC may not be a feasible solution due to its involved mathematical algorithm. Even if the SEC can be obtained, the polygon may not be cyclic. Besides, obtaining the SEC may not be desirable as its circumference size may depend on the number of agents and the minimum distance between each pair of agents on the circle. In addition, as mentioned, concave polygons cannot be inscribed and thus not cyclic.

Consequently, a feasible approach would be to find a circle that may not be the SEC but reasonably covers the entire fleet of agents. Figure 3 illustrates how the circle can be obtained. The figure is a heptagon (7-sided polygon) that cannot be inscribed. The reason is that this heptagon is formed by replacing one of the regular (cyclic) hexagon sides by two outward sides. For the purpose of visualization, the regular hexagon is shown inside of the heptagon by thin broken lines. Therefore, this heptagon is not cyclic.

Since the coordinates of the heptagon needs to be known a-priori, the maximum and minimum x- and y-coordinates are used to form a rectangle. In the figure, the maximum x- and y-coordinates are x_3 and y_5, respectively. Similarly, the minimum x- and y-coordinates are x_6 and y_1, respectively. Using these

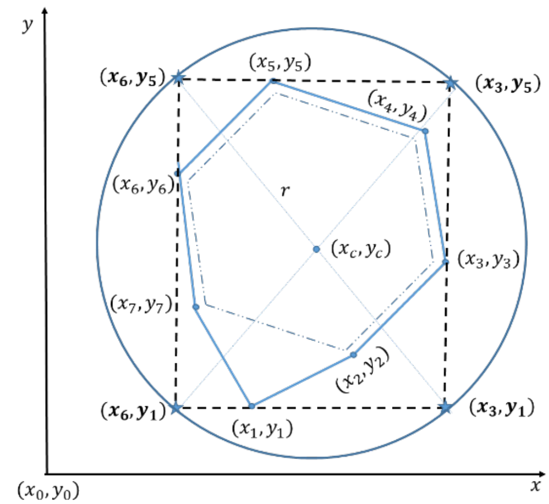


Figure 3: Example for finding a circle that covers a non-regular polygon

figure, the rectangle is shown as a thick dashed box. Its vertices are marked with stars along with their coordinates in bold. Generalizing this approach and using the subscripts of max and min in representing the maximum and minimum x- and y-coordinates, the vertices of the rectangle, starting from the lower left and moving counterclockwise, become: (x_{min}, y_{min}), (x_{max}, y_{min}), (x_{max}, y_{max}), and (x_{min}, y_{max}).

As the figure shows, since any rectangle is cyclic, a circle circumscribing the rectangle can be formed. Obviously, the circle covers all the agents as well. In addition, the rectangle can determine the coordinates of the circle center and the radius of the circle. As the example in the figure shows, the location of the circle center is halfway between the maximum and minimum x- and y-coordinates, In general,

$$(x_c, y_c) = (\frac{x_{min} + x_{max}}{2}, \frac{y_{min} + y_{max}}{2})$$

Since the rectangle diagonal passes through the circle center, the circle radius is half the rectangle diagonal, i.e.,

$$r = \frac{\sqrt{(x_{max} - x_{min})^2 + (y_{max} - y_{min})^2}}{2}$$

It should be noted that, although the example in Figure 3 shows a convex polygon, the same approach can be used for concave polygons as well for obtaining the enclosed circle. Also, as shown in Figure 3, the formation field is assumed to be in the positive quadrant of the xy-plane.

4 Polygon Reconfiguration Phase

This section applies the formation approach discussed in the previous section to regular and non-regular polygons including concave polygons.

4.1 Regular Convex Polygons

To have a firm understanding of the approach, an example using a regular pentagon polygon is shown in Figure 4a. Since the polygon is regular, it is cyclic and thus can be inscribed by the SEC, or the approach described above may be used to obtain the circle for covering the polygon.

It is assumed that the agents have already formed a circle. Recall that a polygon formation with s sides is a two-phase process. In the first phase, the agents form a circle. The second phase is about the reconfiguration in which an agent, let's say with the coordinates (x_a, y_a), obtains its reconfigured location $(x_?, y_?)$ on the polygon. Once $(x_?, y_?)$ is determined, the agent moves to its new reconfigured location by assuming $(x_?, y_?)$ is the center of a circle with radius set to 0. This allows the agent to follow the same procedure for forming a circle, discussed in [23], but the agents end up forming a polygon without the need to perform any redistribution once they reach their specified locations $(x_?, y_?)$.

In the reconfiguration process for regular polygons, it is assumed that the agents are aware of an anchor point with degree α. In Figure 4a below, $\alpha = 270°$, and the location of the anchor point is shown as (x_{anchor}, y_{anchor}). Furthermore, assume the polygon vertices are defined as v_k, $1 \leq k \leq s$, numbered counterclockwise, with the corresponding coordinates (x_k, y_k).

The figure shows an agent between v_5 and v_1 with the corresponding coordinates (x_5, y_5) and (x_1, y_1). For the sake of generality, let's refer to them as v_i and v_{i+1} with the corresponding coordinates (x_i, y_i) and (x_{i+1}, y_{i+1}), where $(i+1)$ is calculated in modulo s. For an agent to attain the coordinates of the end vertices v_i and v_{i+1}, the agent needs to compare its degree on the circle against the degree of each pair of adjacent vertices until a match is found that satisfies $deg_{v_i} \leq deg_a \leq deg_{v_{i+1}}$, $deg_{v_i} \leq deg_a \geq deg_{v_{i+1}}$, or $deg_{v_i} \geq deg_a \leq deg_{v_{i+1}}$, where $(i+1)$ is done in modulo s. The latter two cases might happen because of the rotation from 360° back to 0°. Once a match is found on deg_{v_i} and $deg_{v_{i+1}}$, the agent will use the corresponding coordinates.

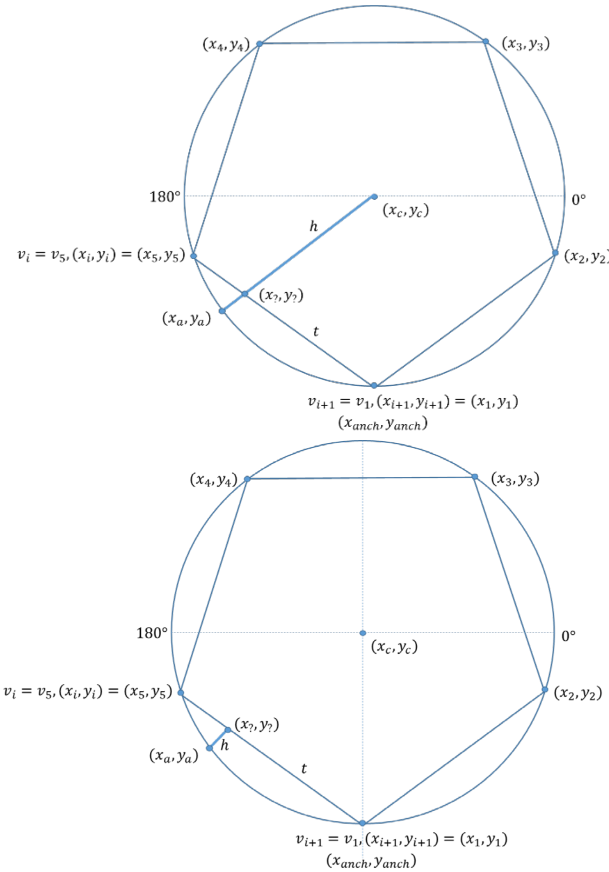
The following shows how to obtain the corresponding coordinates of a vertex v_i with degree deg_{v_i}:

$$(x_i, y_i) = (x_c + r\cos(deg_{v_i}), y_c + r\sin(deg_{v_i}))$$

where r is the circle radius. As an example, in Figure 4a, since the anchor point is at $\alpha = 270°$ and the adjacent vertices are $360°/s = 360°/5 = 72°$ degrees apart from each other,

$$(x_i, y_i) = (x_c + r\cos(\alpha + (i-1) \times 360°/s), y_c + r\sin(\alpha + (i-1) \times 360°/s))$$

To show the general approach in obtaining $(x_?, y_?)$ shown in Figure 4a, let the equation for the line t formed between these two vertices v_i and v_{i+1} be: $y_t = m_t x_t + b_t$. Similarly, let the



a) Using the trajectory to circle center
b) Using the perpendicular trajectory to the polygon

Figure 4: The approach for obtaining the reconfigured point $(x_?, y_?)$

equation for the line h between the circle center and the agent be: $y_h = m_h + b_h$.

Since $(x_?, y_?)$ is at the intersection of both lines t and h, $(x_?, y_?)$ is a valid point for both lines. Thus, two equations with two unknowns are formed that can be solved to obtain $(x_?, y_?)$:

$$y_t = m_t x_t + b_t \Rightarrow y_? = m_t x_? + b_t$$

$$y_h = m_h x_h + b_h \Rightarrow y_? = m_h x_? + b_h$$

This leads to:

$$x_? = \frac{b_h - b_t}{m_t - m_h} \quad (7)$$

$$y_? = \frac{m_t b_h - m_h b_t}{m_t - m_h} \quad (8)$$

Furthermore, the slopes of the lines are:

$$m_t = \frac{y_{i+1} - y_i}{x_{i+1} - x_i} \quad (9)$$

$$m_h = \frac{y_c - y_a}{x_c - x_a} \quad (10)$$

And the y-intercepts of the lines are:

$$b_t = y_i - m_t x_i \quad (11)$$

$$b_h = y_a - m_h x_a \quad (12)$$

In (7) and (8), b_h, b_t, m_t, and m_h are known, as shown in (9) – (12). Also, (x_a, y_a), (x_i, y_i), and (x_c, y_c) are known. Consequently, $(x_?, y_?)$ in (7) and (8) can be calculated easily.

An interesting, revised version of the approach for obtaining $(x_?, y_?)$ is for an agent to continuously travel to the polygon on the shortest path. This implies that the trajectory h would be perpendicular to the polygon side t. This is shown in Figure 4b. The slope of y_t is:

$$m_t = \frac{y_i - y_{i+1}}{x_i - x_{i+1}} \quad (13)$$

On the other hand, the y-intercept of t is b_t, which can be obtained as:

$$y_i = m_t\, x_i + b_t \Rightarrow b_t = y_i - \frac{y_i - y_{i+1}}{x_i - x_{i+1}} x_i \quad (14)$$

Since the slope of h is the negative inverse of m_t,

$$m_h = -m_t^{-1} = -\frac{x_i - x_{i+1}}{y_i - y_{i+1}} \quad (15)$$

Since the location of the agent is known as (x_a, y_a),

$$y_h = m_h x_h + b_h \Rightarrow y_a = -m_t^{-1} x_a + b_h \Rightarrow b_h = y_a + m_t^{-1} x_a \quad (16)$$

Thus,

$$y_h = -m_t^{-1} x_h + (y_a + m_t^{-1} x_a) \quad (17)$$

Since $(x_?, y_?)$ is at the intersection of both lines t and h, $(x_?, y_?)$ is valid for both lines. Thus, two equations with two unknowns are formed that can be solved to obtain $(x_?, y_?)$:

$$y_t = m_t x_t + b_t \Rightarrow y_? = m_t x_? + b_t \quad (18)$$

$$y_h = m_h x_h + b_h \Rightarrow y_? = m_h x_? + b_h \quad (19)$$

This leads to:

$$x_? = \frac{b_h - b_t}{m_t - m_h} \quad (20)$$

$$y_? = m_t \frac{b_h - b_t}{m_t - m_h} + b_t \quad (21)$$

Since b_t, b_h, m_t, and m_h are all known, $x_?$ and $y_?$ can be calculated easily as well. However, two special cases need to be checked for when a line segment happens to be horizontal or vertical. The line is horizontal if $y_i = y_{i+1}$. In that case, the travel path h for the agent is vertical, so the x-coordinate of the agent stays the same. Therefore,

$$x_? = x_a$$

$$y_? = y_i$$

If the line segment is vertical, i.e., $x_i = x_{i+1}$, the y-coordinate does not change as the agent travels since the travel is horizontal. Thus,

$$x_? = x_i$$

$$y_? = y_a$$

The discussion for Figure 4a and its revised version in Figure 4b are also applicable to cyclic non-regular polygons if the angles of the vertices are known to the agents. A notable ramification to the revised version is that if the agents are uniformly distributed on the circle, they would be distributed uniformly on the polygon structure as well.

4.2 Irregular Convex Polygons

The process for forming irregular convex polygons is like that of regular polygons. Figure 5 below shows a non-regular polygon inscribed in a circle. In this example, the circle circumscribing the polygon is the SEC. Compared to Figure 4, an agent (x_a, y_a) can easily apply the approach discussed in the previous subsection to determine its reconfiguration point $(x_?, y_?)$. However, in terms of input to the agents, since the polygon is irregular, having the location of the anchor point to determine the entire structure of the polygon is no longer sufficient. More specifically, since the polygon is not regular,

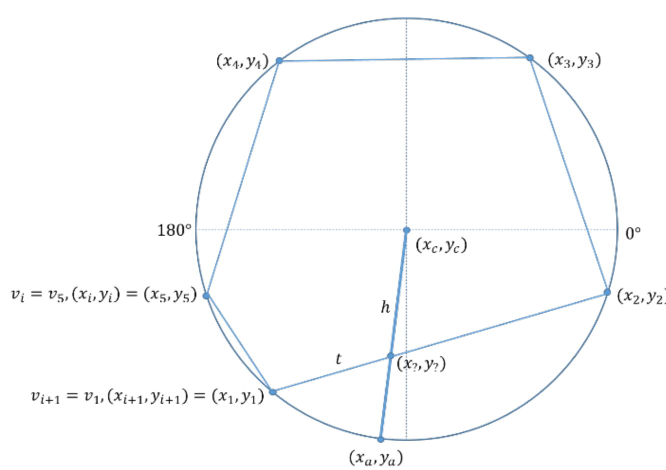


Figure 5: Example showing the approach for obtaining the reconfigured point $(x_?, y_?)$ for irregular polygons

the locations of the vertices need to be known before the reconfiguration can take place.

The approach for non-regular convex polygons that cannot be inscribed is the same, except the circle is not necessarily a SEC (see Figure 3). It should be noted that as the fleet of agents becomes larger the polygon formations become more pronounced.

The next section discusses the case for some categories of concave polygons.

4.3 Concave Polygons

Since concave polygons are not inscribable, a process similar to what was described in Section 3 needs to be applied to create a reasonable circle that covers the polygon vertices. Consider the two concave polygon examples in Figure 6 that have used the process in Section 3 for creating the circles around the vertices.

In Figure 6a, the trajectory between the agent location (x_a, y_b) and the circle center (x_c, y_c) intersects the polygon in one location. Therefore, the agent is able to determine its reconfigured point $(x_?, y_?)$ uniquely. However, in Figure 6b, the trajectory cuts through the polygon in multiple locations and thus it is not clear which intersection point to use as the reconfigured point $(x_?, y_?)$. The reason is that the degree of the polygon vertices in Figure 6b are not in increasing order. Even if the agent is able to randomly decide on one of the intersection points, the formation of the polygon would not be deterministic. Consequently, the formation of concave polygons in this research is limited to cases such as the one in Figure 6a. Specifically, the concave polygons considered are limited to those with $deg_{v_{i+1}} \geq deg_{v_i}$, where $1 \leq i \leq s$ and $(i+1)$ is in module s.

5 Polygon Transformations

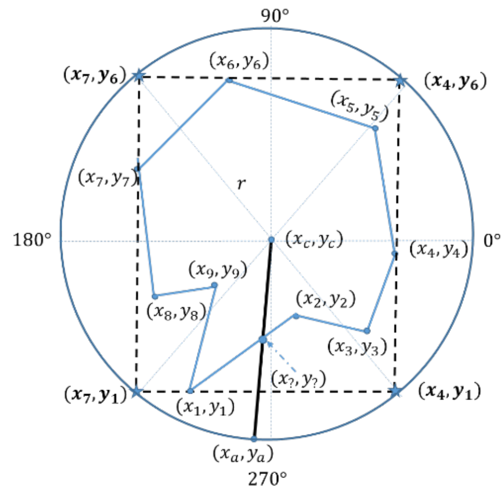
This section illustrates how the agents, once located on the polygon, would maintain the polygon structure while rotating in place, translating (moving) to a different location, rotating and translating simultaneously, or scaling in its final location of rotation and translation.

5.1 Polygon Rotation

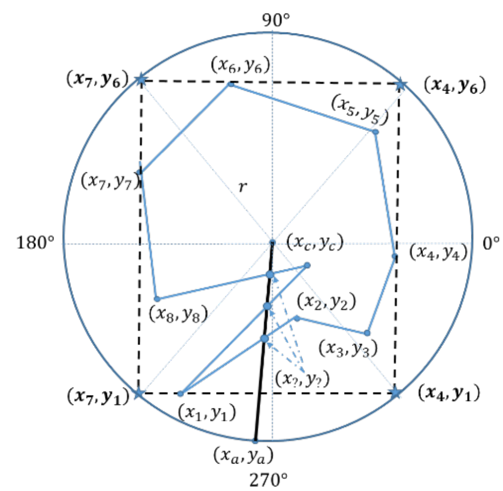
The algorithm for rotating the polygon structure in place is relatively simple since (5) and (6) have shown how to move an agent from one location to another on the circle boundary. These equations are modified slightly to account for incremental rotations. From Figure 1, the location of A_i at time t before being relocated is:

$$x_{a_i}(t) = x_c(t) + r\cos\big(Angle_{i0}(t)\big) \quad (22)$$

$$y_{a_i}(t) = y_c(t) + r\,\sin\big(Angle_{i0}(t)\big) \quad (23)$$



a)



b)

Figure 6: Agent trajectory for two examples of a concave polygon: a) cutting only one side of the polygon, b) cutting multiple sides of the polygon

After dt time, the agent's new location, as shown in Figure 1, is:

$$x_{a_i}(t+dt) = x_c(t) + r\cos\big(Angle_{i0}(t) + \Delta_\phi \phi\big) \quad (24)$$

$$y_{a_i}(t+dt) = y_c(t) + r\sin\big(Angle_{i0}(t) + \Delta_\phi \phi\big) \quad (25)$$

Taking advantage of cosine and sine properties, the cosine and sine portions of (24) and (25) can be expanded as:

$$\mathrm{Cos}\big(Angle_{i0}(t) + \Delta_\phi \phi\big) = \cos\big(Angle_{i0}(t)\big)\cos\big(\Delta_\phi \phi\big) - \sin\big(Angle_{i0}(t)\big)\sin\big(\Delta_\phi \phi\big) \quad (26)$$

$$\sin\big(Angle_{i0}(t) + \Delta_\phi \phi\big) = \cos\big(Angle_{i0}(t)\big)\sin\big(\Delta_\phi \phi\big) + \sin\big(Angle_{i0}(t)\big)\cos\big(\Delta_\phi \phi\big) \quad (27)$$

Substituting (26) in (24) yields:

$$x_{a_i}(t+dt) = x_c(t) + r\big(\cos\big(Angle_{i0}(t)\big)\cos\big(\Delta_\phi \phi\big) - - \sin\big(Angle_{i0}(t)\big)\sin\big(\Delta_\phi \phi\big)\big) \quad (28)$$

From (22) and (23):

$$\big(x_{a_i}(t) - x_c(t)\big) = r\cos\big(Angle_{i0}(t)\big),$$

$$(y_{a_i}(t) - y_c(t)) = r \sin\big(Angle_{i0}(t)\big) \quad (29)$$

Substituting (29) in (28) yields:

$$x_{a_i}(t+dt) = x_c(t) + (x_{a_i}(t) - x_c(t))\cos(\Delta_\phi \phi) - (y_{a_i}(t) - y_c(t))sin(\Delta_\phi \phi) \quad (30)$$

Using (27) and taking advantage of (29), a similar substitution along the y-coordinate in (25) results in:

$$y_{a_i}(t+dt) = y_c(t) + (x_{a_i}(t) - x_c(t))\sin(\Delta_\phi \phi) + (y_{a_i}(t) - y_c(t))cos(\Delta_\phi \phi) \quad (31)$$

Note that the center of the formation is invariant under the rotation transformation equations (30) and (31). Additionally, the equations can be used to prove the transformation is a rigid-body transformation. That is, the agents on the same line of the polygon will be on the rotated line with the same relative locations and the same distances from each other with all the corresponding angles preserved.

5.2 Polygon Translation

Translation is the changes in x- and y-coordinate of a polygon location, which translates to the same changes of location for every agent on the polygon. In other words, translation is a displacement of location added to each of the agents' location. As a result, the formation topology is preserved as in the case of rotation. For instance, a translation by (x_{trns}, y_{trns}) for an agent A_i currently at location (x_{a_i}, y_{a_i}) results in the agent's new location at $(x_{a_i} + x_{trns}, y_{a_i} + y_{trns})$.

Consequently, an agent A_i with the coordinates $(x_{a_i}(t), y_{a_i}(t))$ and a desired translation by (x_{trns}, y_{trns}), at time $(t + dt)$, will have the new coordinates:

$$x_{a_i}(t+dt) = x_{a_i}(t) + x_{trns},\ y_{a_i}(t+dt) = y_{a_i}(t) + y_{trns} \quad (32)$$

Using (32) and the previous equations (30) and (31), the polygon formations can be rotated and translated incrementally until the desired rotation and translation are achieved. The order of transformation is application dependent. If translation by (x_{trns}, y_{trns}) is performed first, then the invariant point (the modified circle center) in (30) and (31) is adjusted first. The new updated (x_c, y_c) will be:

$$x_c(t+dt)x_c(t) + x_{trns} \quad , \quad y_c(t+dt) = y_c(t) + y_{trns} \quad (33)$$

5.3 Polygon Scaling

Scaling is a feature that enhances the flexibility of the polygon formation, by expanding or shrinking the area coverage of the formation about the center of the formation. To keep the topology of the formation unchanged, uniform scaling can be applied to each agent's location.

By fixing the center of the formation as the invariant location under uniform scaling, the shape of the formation is maintained. However, the agents' distances from the formation center and the relative distances among agents need to be adjusted appropriately.

If the agents' coordinates are measured relative to the origin coordinates $(0, 0)$, then uniform scaling is the process of multiplying the xy coordinates by a scale factor s. Under this transformation, $(x_{a_i}(t), y_{a_i}(t))$ is transformed at time $(t + dt)$ to:

$$\big(x_{a_i}(t+dt), y_{a_i}(t+dt)\big) = \big(sx_{a_i}(t), sy_{a_i}(t)\big) \quad (34)$$

For s between 0 and 1, the formation area is decreased. For $s > 1$, the formation area is increased.

In (34), the location (0, 0) is the invariant point under scaling. To make the center of the formation $(x_c(t), y_c(t))$ as the invariant point, (34) should be adjusted to:

$$\big(x_{a_i}(t+dt), y_{a_i}(t+dt)\big) = \big(x_c(t), y_c(t)\big) + s\big(x_{a_i}(t) - x_c(t), y_{a_i}(t) - y_c(t)\big) \quad (35)$$

$$= \big((sx_{a_i}(t) + (1-s)x_c(t)), (sy_{a_i}(t) + (1-s)y_c(t))\big) \quad (36)$$

Equations (35) and (36) are the concatenation of three transformations: a translation by $-(x_c(t), y_c(t))$, so that the scaling transformation by s in (34) can be applied, followed by a translation of $(x_c(t), y_c(t))$.

5.4 Incremental Polygon Transformation

The equations for the three primary transformations described above can be revised to achieve incremental transformations.

This allows the agents to move at a desired rate before settling on their final transformation. In addition, the incremental transformation can be applied to accomplish a composite transformation. That is, the formation structure can go under a sequence of rotation, translation, scaling, or any predefined order.

According to the translation properties, an incremental translation by (dx_{trns}, dy_{trns}) of duration dt followed by another incremental translation (dx'_{trns}, dy'_{trns}) of duration dt', and applied to the center of the formation $(x_c(t), y_c(t))$, produces a translation with the new center of formation as:

$$(x_c(t + dt + dt'), y_c(t + dt + dt') = (x_c(t) + dx_{trns} + dx'_{trns},, y_c(t) + dy_{trns} + dy'_{trns}) \tag{37}$$

Equation (37) shows that the concatenation of translations is additive. This result, and other similar results, can be proved using matrix representations of transformations applied to homogeneous coordinate system. A similar analysis can be applied to rotation, to show that a sequence of rotations about the same invariant point, results in a rotation by a degree that is the sum of the degrees used for the individual rotations. Unlike translation and rotation, composite scaling is not additive [1, 6].

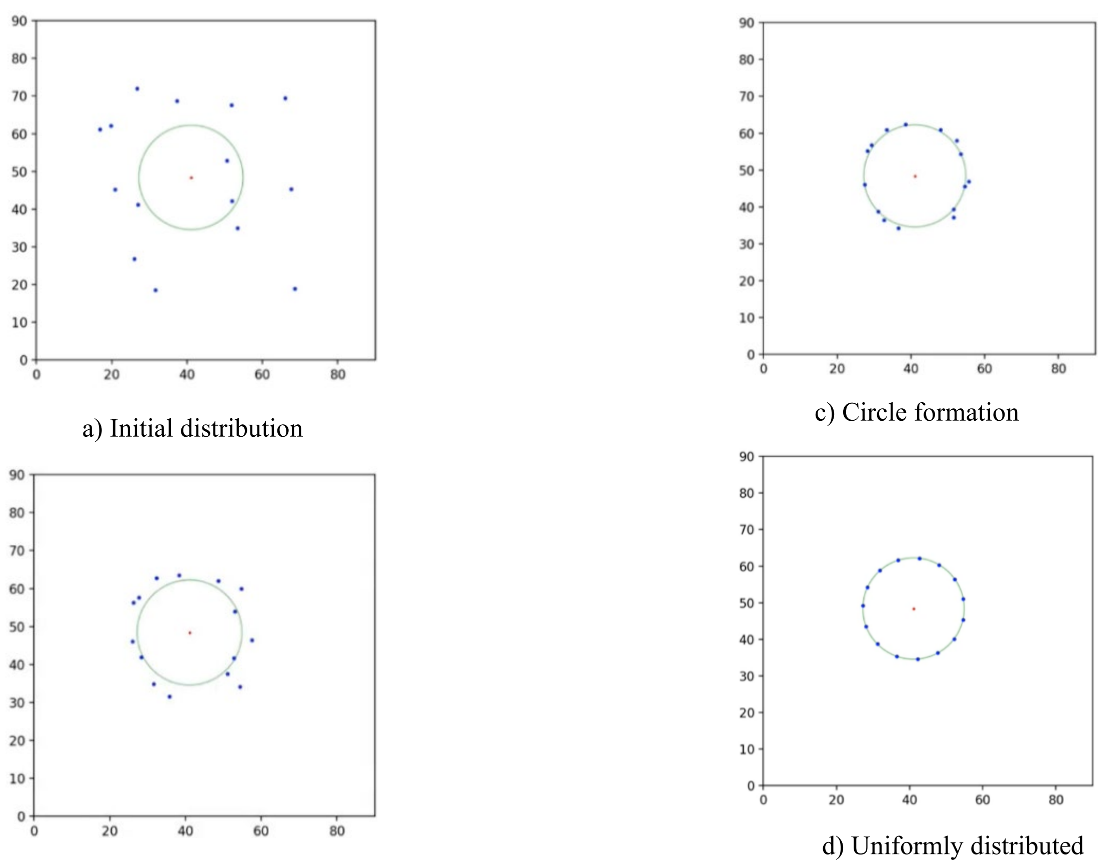
6 Simulation

This section provides some simulation experiments in accordance with the discussion in the previous section.

6.1 Two-Phase Formation

The Python simulations illustrate the two-phase approach of circle formation followed by the reconfiguration process for cyclic polygons to achieve a triangle formation, a rectangle, and then a pentagon. Figure 7 displays four snapshots for achieving a circle formation. Figure 7a shows the initial, random distribution of the agents. Figure 7b shows the agents moving toward the circle while avoiding collisions. In Figure 7c, the agents have almost reached the circle. In Figure 7d, the agents have formed the circle and repositioned themselves into a uniformly distributed formation, while avoiding collisions.

During the entire process of formation, the agents are entirely distributed with no assistance from any external entity. The only external input received is the number of agents and the minimum collision distance for avoiding collision with their neighbors. The agents are optionally able to receive input as to how large the circle formations should be. It should be mentioned that the drawing of the circle in the figure is not



a) Initial distribution

b) Reaching the circle

c) Circle formation

d) Uniformly distributed

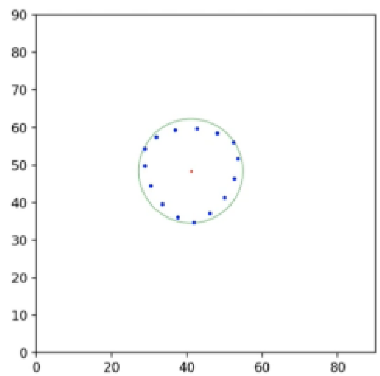
Figure 7: Snapshots of circle formation

necessary. It is merely drawn as a reference to assist in better observation.

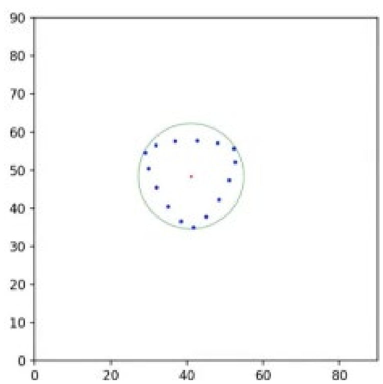
Figure 8 shows the snapshots of agents reconfiguring themselves into a triangle immediately following the Figure 7d circle formation. In Figure 8a, the agents have determined their reconfigured locations $(x_?, y_?)$ and about to move toward those locations. In Figure 8b, the triangle formation is clearly visible. Figure 8c illustrates the complete formation of agents into a triangle. The agents on the reconfigured formation are still uniformly distributed because they were distributed uniformly on the circle and move toward the polygon sides perpendicularly. If they were not, then their distribution on the triangle would not be uniform either.

Figure 9 shows the continuation of Figure 8c, where the agents continuously reconfigure themselves to a rectangle, to a square, and finally to a pentagon.

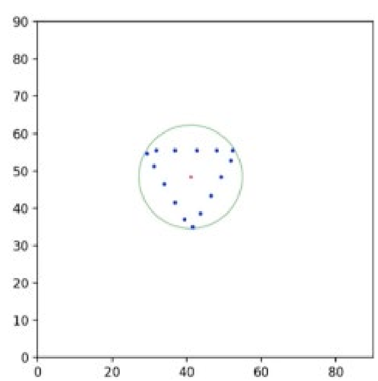
These simulations followed the process shown in Section 4. However, the study revealed other alternatives for obtaining $(x_?, y_?)$. Although the calculations are more involved, one approach that we have developed and simulated with success is taking advantage of the triangulation process. More specifically, once the coordinates of v_i and v_{i+1} are determined,



a) Starting to move

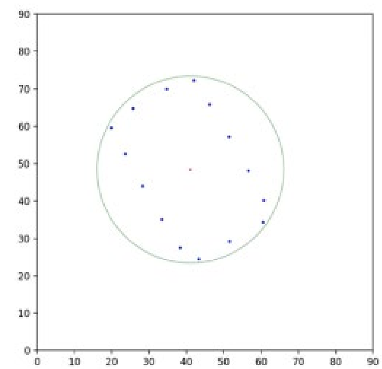


b) Close to completion

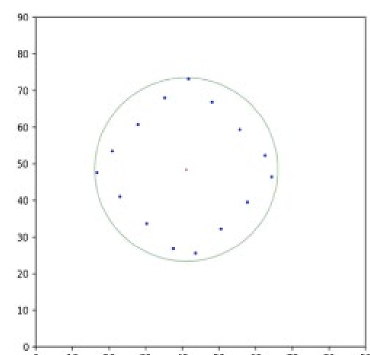


c) Triangle formed

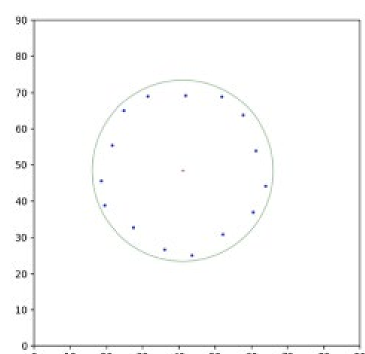
Figure 8 : Snapshots of triangle reconfiguration



a) Rectangle



b) Square



c) Octagon

Figure 9: Snapshots of changing reconfiguration

Heron's formula [14] is applied to find the distance h (see Figure 4b). The location $(x_?, y_?)$ is then triangulated using the three circles centered at the agent, v_i and v_{i+1}.

6.2 Transition of Formation through Transformations

Following the two-phase approach to the polygon formation, incremental transformation can be applied in the form of translation, rotation, or scaling of the entire formation, while preserving the formation structure. For a better insight and before providing some snapshots of a pentagon transformation in Subsection 6.3, the following illustrates the Excel implementation of rotation with translation using the incremental procedure described in Subsection 5.4.

Suppose it is desired to translate a circle formation from $(x_c(t), y_c(t))$ to the new location centered at $(x_c'(t'), y_c'(t'))$, for some time $t' > t$, and rotated by deg degrees. For this to happen, the formation is incrementally rotated by a small degree $d\phi$ followed by a small translation displacement (dx_{trns}, dy_{trns}). Suppose these steps are repeated n times to reach the desired transformation. Accordingly, each incremental transition uses:

$$d\phi = \frac{deg}{n}, \quad dx_{trns} = \frac{x_c(t') - x_c(t)}{n}, \quad dy_{trns} = \frac{y_c(t') - y_c(t)}{n},$$

$$dt = \frac{t' - t}{n} \tag{38}$$

In (38), $deg = \Delta_\phi \phi$ (See Figure 1). In addition, the number of steps n does not have to be the same for rotation, translation, and time. For example, if the number of steps for rotation and translation are n and m, respectively, where $n > m$, the rotation will continue without any translation for the remaining $(n - m)$ steps. The following shows the steps in carrying out the task:

For $k = 1$ to n do {

// Apply incremental rotation
For each A_i do {
$x'_{a_i}(t + dt) = (x_{a_i}(t) - x_c(t))\cos(d\phi) - (y_{a_i}(t) - y_c(t))\sin(d\phi) + x_c(t)$ // See (30)
$y'_{a_i}(t + dt) = (x_{a_i}(t) - x_c(t))\sin(d\phi) + (y_{a_i}(t) - y_c(t))\cos(d\phi)$ // See (31)
}

// Apply incremental translation by updating the coordinates
For each A_i do {
$(x_{a_i}(t), (y_{a_i}(t)) = (x'_{a_i}(t + dt), y'_{a_i}(t + dt))$
}
$(x_c(t), y_c(t) = (x_c(t) + dx_{trns}, y_c(t) + dy_{trns})$

}

In the code above, it is important to update the center of the formation so as to maintain the correct rotation about the updated center of formation.

Figure 10 below shows the results of the above code execution. In the simulation, four agents are distributed around a circle of radius 3 centered at $(4, 6)$. The formation was rotated by 30° and translated by $(5, 3)$, with $n = 30$. The circles on the graph are not part of the simulation. They are drawn to enhance visualization.

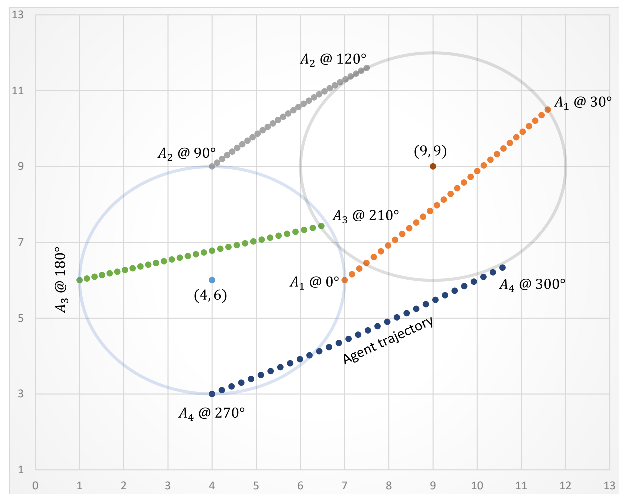
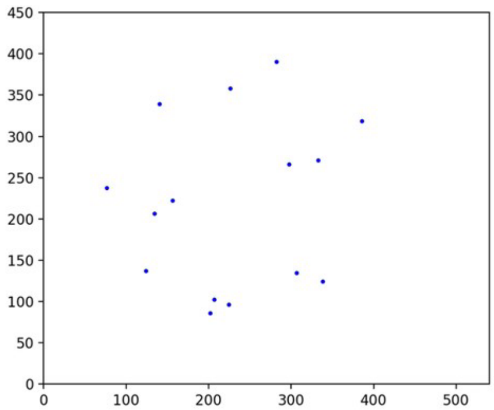


Figure 10: Rotation and translation of four agents

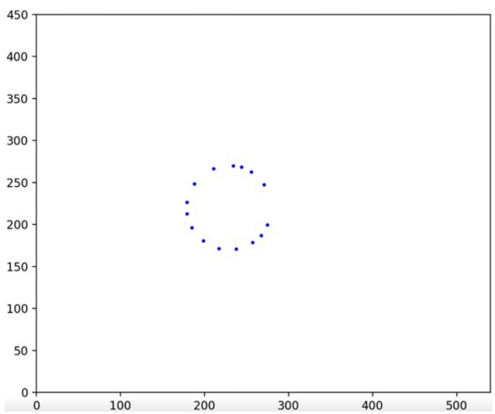
6.3 Pentagon Transformation

Having a better understanding of the discussion in the previous subsection, the following illustrates the formation of a pentagon followed by its transformation of rotation and translation as described in Section 5.

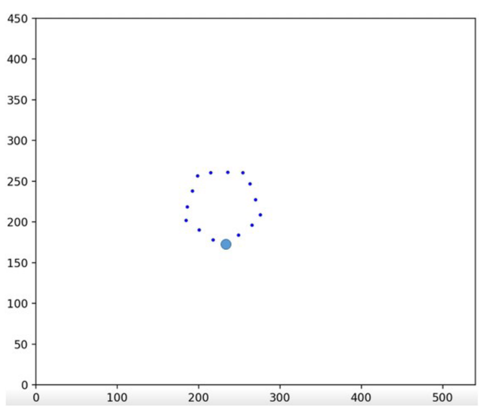
Figure 11a shows the initial distribution of some agents in a field. Figure 11b shows that the agents have formed a circle in phase 1 and attempting to distribute themselves on the circle boundary uniformly. In Figure 11c, the agents have formed a pentagon in phase 2. For better visualization in the follow up figures, the bottom agent is shown thicker. Figure 11d displays the fact that the pentagon has rotated to the left and moved to a different place while maintaining its shape. In Figure 11e, the pentagon has made about half a turn compared to Figure 11c. In Figure 11f, the pentagon has made a full turn, while moving to a different location. Once reaching its position in Figure 11f, the pentagon keeps rotating in place, but its rotation is not shown in this figure.



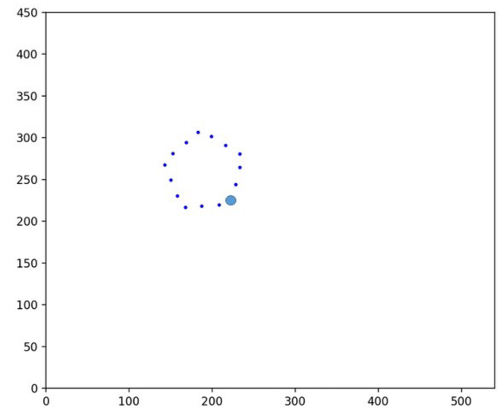
a) Initial distribution of agents



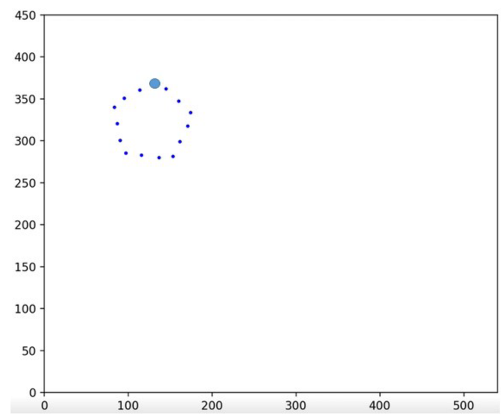
b) Agents attempting to distribute uniformly



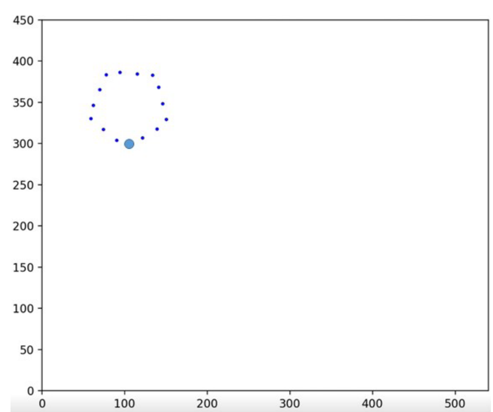
c) Agents forming the pentagon



d) The pentagon rotating while translating



e) The pentagon almost at half turn while translating



f) The pentagon making a full turn while translating

Figure 11: Transformation of a pentagon through rotation and translation

7 Conclusion

A major challenge in the design of multi-agent formations is the identification of the agents' positions on the formation structure. The proposed research offers a two-phase approach to polygon formations, which borrows elements from the behavioral and virtual structures principles. The approach has played a fundamental role in improving performance and mitigating the impractical assumptions. It enables the agents to identify their positions on a polygon uniquely, autonomously, and avoid collisions during the reconfiguration phase. In the first phase, the agents form an enclosing circle over the formation structure. For better and more concise polygon formations, the agents have the option of uniformly distribute themselves on the circle. During the second phase, the agents reconfigure themselves into the desired polygon formation. In addition to polygon formations, the two-phase approach simplified the design for rotation, translation, and scaling of polygons.

The approach identified the types of polygons that can be formed. Specifically, the proposed approach handles convex and concave polygons under certation constraints. For example, concave polygons can be formed if the degrees of the vertices are in ascending order. As the number of agents increases, the structure of polygon formations become more pronounced, especially for concave polygons. In contrast to some studies, the number of agents deployed does not depend on that of polygon vertices. Furthermore, no distinction is made between agents for conducting special tasks or allocating special agents to form the polygon vertices.

Several future studies are anticipated. One is to remedy or reduce the current restrictions to include a wider range of concave polygons. Also, early results indicate that the current approach can be adjusted to better distribute the agents on non-cyclic polygons. Another avenue of research is to modify the control laws to handle faults, e.g., if an agent does not adhere to the consensus protocol or to the collision avoidance operations. In addition, the current research is testing the proposed approach in Robotic Operating System (ROS) as the stepping-stone to prototyping the control laws.

References

[1] E. Angle, *Interactive Computer Graphics*, 5th Ed., Addison Wesley, 2009.

[2] F. Dignum, P. Mathieu, J. M. Corchado, and F Prieta, "Advances in Practical Applications of Agents, Complex Systems Simulation. The PAAMS Collection," 20th Int'l Conf., PAAMS 2022, Springer, *LNCS*, 13616, 2022.

[3] A. Dorri, S.S. Kanhere, and R. Jurdak, "Multi-Agent Systems: A Survey," *IEEE Access*, 2018.

[4] X. Chen, A. Serrani, and H. Ozboy, "Control of Leader-Follower Formation of Terrestrial UAVs," *IEEE Conf. on Decision and Control*, pp. 498-503, 2003.

[5] K. Fathian, N. R. Gans, W. Krawcewics, and D. Rachinskii, "Regular Polygon Formations with Fixed Size and Cyclic Sensing Constraint," *IEEE Transactions on Automatic Control*, 64(12):5156-5163, 2019.

[6] D. Hearn and, M. P. Baker, *Computer Graphics*, 2nd Ed., Prentice Hall, 1994.

[7] Ikemoto Y. Y. Hasegawa, T. Fukuda, and K. Matsuda, "Gradual Spatial Pattern Formation of Homogenous Robot Group," *Information Sciences: An Int'l J.*, 17(14):431-445, 2005

[8] B. A. Issa, A. T. Rashid, and M. T. Rashid, "Leader-Neighbor Algorithm for Polygon Static Formation Control," *Int'l Conf. on Electrical Communication, and Computer Eng.*, 2020.

[9] T. Kopfstedt, M Mukai, M. Fuita, and C. Ament, "Control of Formations of UAVs for Surveillance and Reconnaissance Missions," *IFAC Proceedings Volumes*, 41(2):5161-5166, 2008.

[10] J. R. Lawton, R. W. Beard, and B. Young, "A Decentralized Approach to Formation Control Maneuvers," *IEEE Transactions on Robotics and Automation*, 19(6):933-941, 2004.

[11] C. B Low., "A Dynamic Virtual Structure Formation Control for Fixed Wing UAVs," *IEEE Int'l Conf. on Control and automation*, 2011.

[12] M. Oprea, "Applications of Multi-Agent Systems," *IFIP Congress Tutorials*, 2004.

[13] B. K. S. Prasad, A. G. Manjunath, and H. Ramasangu, "Multi-Agent Polygon Formation Using Reinforcement Learning," *Int'l Conf. on Agents and Artificial Intelligence*, 2017.

[14] C. H. Raifaizen, "A Simple Proof of Heron's Formula," *Mathematics Magazine*, 44(1):27-28, 1971.

[15] W. Ren, "Consensus strategies for cooperative control of vehicle formation", *Control Theory and Applications*, 1(2), pp. 505-512, 2015.

[16] S. B. Sarsilmaz and T. Yucelen, "Control of Multiagent Systems with Local and Global Objectives," *IEEE Conf. on Decision and Control*, 2018.

[17] S. Skyum, "A Simple Algorithm for Computing the Smallest Enclosing Circle," *Information Processing Letters*, 37(3):121-125, 1991.

[18] Y. H. Tan, S. Lai, K. Wang, and B. M. Chen, "Cooperative Control of Multiple Unmanned Aerial Systems for Heavy Duty Carrying," *Annual Reviews in Control*, 46:44-57, 2018.

[19] A. Uhrmacher, D. Weyns, and P.J. Mosterman, "Multi-Agent Systems: Simulation and Applications," CRC Press, 2009.

[20] X. K. Wang, X. Li, and Z .Q. Zheng, "Survey of Developments on Multi-Agent Formation Control Related Problems," *Control and Decision*, 28(11):1601-1613, 2013.

[21] E. Welzl, "Smallest Enclosing Disks (Balls and Ellipsoids)," H. Maurer (ed.), New Results and New Trends in Computer Science, *LNCS*, 555:359-370, 1991.

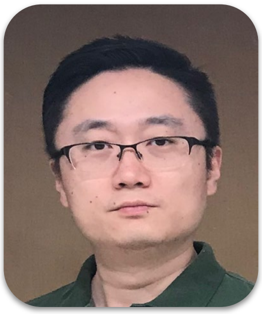
[22] B. Xu, J. L. Bai, Y. L. Hao, W. Gao, and YlL. Liu, "The Research Status and Progress of Cooperative Navigation for Multiple AUVs," *Acta Automatica Sinica*, 41(3):445-461, 2015.

[23] R. Yang, A. Azadmanesh, and H. Farhat, "A New Approach to Circle Formation in Multi-Agent Systems," *Int'l Conference on Wireless Networks*, 2019.

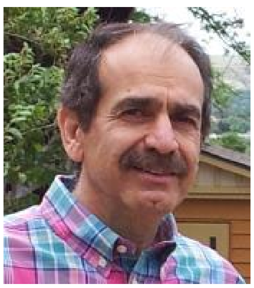
[24] Ziquan Yu, Y. Zhang, B. Jiang, J. Fu, and Y. Jin, "A Review on Fault-Tolerant Cooperative Control of Multiple Unmanned Aerial Vehicles," *Chinese J. of Aeronautics*, 35(1):1-18, 2022.

[25] B. Zhou, Q. Yang, L. Dou, H. Fang, and J. Chen, "An Attempt to Self-Organized Polygon Formation Control of

Swarm Robots Under Cyclic Topologies,”, *IFAC-PapersOnLine*, 53(2):11000-11005, 2020.



Rui Yang is a Ph.D. candidate in the college of IS&T at the University of Nebraska-Omaha. His research interests are in multi-agent systems with emphasis on formation control and their applications in AI and robotic related areas.



Azad Azadmanesh received the Ph.D. degree in computer science from the University of Nebraska-Lincoln, USA. He is currently a Professor in the Department of Computer Science at the University of Nebraska-Omaha. His research interests include fault-tolerance, network technologies, distributed agreement, and reliability modeling.

Hassan Farhat (photo not available) received his Ph.D. in Computer Science and Engineering from the University of Nebraska-Lincoln, USA. His research interests and publications are in the areas of computer graphics, VLSI design and testing, and recently in multi-agent systems. His current research includes comparative studies of the numerical solutions to the consensus problems in multi-agent systems. Dr. Farhat has also been involved with active learning, which has resulted in textbook publications.

Accelerating Dynamic Programming by P-Fold Pipeline Implementation on GPU *

Susumu Matsumae†
Saga University, Saga 840-8502, JAPAN

Abstract

In this paper, we show the effectiveness of pipeline implementations of Dynamic Programming (DP) on Graphics Processing Unit (GPU). We deal with a simplified DP problem where each element of its solution table is calculated in order by semi-group operations among several of already computed elements in the table. We implement the DP program on GPU in a pipeline fashion, i.e., we use GPU cores for supporting pipeline-stages so that several elements of the solution tables are partially computed at one time. Further, to accelerate the pipeline implementation, we propose a p-fold pipeline technique, which enables larger parallelism more than the number of pipeline-stages.

Key Words: Dynamic programming; GPGPU; pipelining.

1 Introduction

In this paper, we show the effectiveness of pipeline implementations of Dynamic Programming (DP) on Graphics Processing Unit (GPU). We deal with a simplified DP problem where each element of its one-dimensional solution table of size n is calculated in order by semi-group computations among several of k already computed elements in the table. Since the size of solution table is n and each element requires computations of k elements, the simplified DP problem is solved sequentially in $O(nk)$ steps.

It has been studied to speed up DP programs using GPU (e.g. [2, 10]), where they mainly focus on optimizing the order of accessing data by proposing novel techniques avoiding memory access conflicts. In our study, however, we consider adopting a pipeline technique and implementing the DP program on GPU in a pipeline fashion. The pipeline computation technique [11] can be used in situations in which we perform several operations $\{OP_1, OP_2, \ldots, OP_n\}$ in a sequence, where some steps of each OP_{i+1} can be carried out before operation OP_i is finished. In parallel algorithms, it is often possible to overlap those steps and improve total execution time.

In our previous studies [5, 6], we solved the simplified DP problem on GPU in a pipeline fashion, i.e., we use GPU cores for supporting pipeline-stages so that several elements of the solution tables are respectively computed partially at one time. Our pipeline implementation determines one output value per one computational step with k threads in a pipeline fashion and solves the simplified DP problem on GPU in $O(n)$ steps, which is k times faster than the $O(nk)$ steps for the sequential algorithm.

Here in this paper, we try to further accelerate the pipeline implementation with larger parallelism. In our previous studies [5, 6], the parallelism is up to k, the number of pipeline-stages. Here, to accelerate the pipeline implementation, we propose a p-fold pipeline technique, and show that it enables larger parallelism more than k. We also discuss the effectiveness of our method with explaining the 0-1 knapsack problem by DP as an example.

The rest of this paper is organized as follows. Section 2 introduces problem definitions and base algorithms. Section 3 explains our pipeline implementations for DP on GPU. Section 4 proposes our p-fold pipeline technique and explains how to accelerate the pipeline implementation by increasing the parallelism more than the number of pipeline-stages. Section 5 provides experimental results. And finally, Section 6 offers concluding remarks.

2 Preliminaries

In this section, we introduce some preliminary definitions and base algorithms. We first define a simplified DP problem to be solved on GPU, and then explain a naive and standard GPU implementations of programs.

2.1 Simplified DP Problem

In this study, we implement a typical DP program on GPU. To simplify the exposition, we focus on programs that solve such a simplified DP problem defined as follows:

Definition 1. *(Simplified DP Problem) A one-dimensional array* $\mathtt{ST}[0], \mathtt{ST}[1], \ldots, \mathtt{ST}[n-1]$ *of size n as a solution table, a set* $\mathscr{A} = \{a_0, a_1, \ldots, a_{k-1}\}$ *of k integers representing offset numbers, and a semi-group binary operator* $\otimes$ *over integers are given. Without loss of generality, every element of set* $\mathscr{A}$

*This is the extended version of the paper presented at the 35th International Conference on Computer Applications in Industry and Engineering, 2022 (CAINE 2022) [7].

†Department of Information Science. Email: matsumae@is.saga-u.ac.jp

satisfies the following inequality:

$$a_0 > a_1 > \cdots > a_{k-1} > 0. \tag{1}$$

Then, a simplified DP problem (S-DP problem) is to fill all the elements of array ST *in such a way that each* $\mathtt{ST}[i]$ *is computed by the following equation:*

$$\mathtt{ST}[i] = \otimes_{0 \le j < k} \mathtt{ST}[i - a_j] \tag{2}$$

where $\mathtt{ST}[0], \mathtt{ST}[1], \ldots,$ *and* $\mathtt{ST}[a_0 - 1]$ *are preset with initial values.* ■

For example, Fibonacci number problem can be seen as the S-DP problem where $k = 2, a_0 = 2, a_1 = 1, \otimes = +$, and ST[0]=ST[1]=1.

Even if a problem is solved by DP with two-dimensional solution table, the problem may be reduced to the S-DP problem of one-dimensional solution table. The interested reader should refer to [6], where the matrix-chain multiplication (MCM) problem (e.g., [1]) is discussed as an example.

2.2 Conventional Approach to S-DP Problem

To begin with, we show a straightforward sequential algorithm that solves the S-DP problem. The algorithm is shown in Figure 1. The outer loop computes values from $\mathtt{ST}[a_0]$ to $\mathtt{ST}[n-1]$ in order. It should be noted that $\mathtt{ST}[0], \mathtt{ST}[1], \ldots, \mathtt{ST}[a_0 - 1]$ are initially given as input values. The inner loop computes ST[i] for each i by equation (2). Since the outer loop takes $n - a_0 = O(n)$ iterations and the inner loop requires $O(k-1)$ steps, this sequential algorithm solves the S-DP problem in $O(nk)$ steps in total.

A Sequential Algorithm for S-DP Problem

for i = a_0 **to** $n-1$ **do**

  ST[i] = ST[i−a_0];
  for j = 1 **to** $k-1$ **do**

    ST[i] = ST[i] $\otimes$ ST[i−a_j];

Figure 1: A sequential algorithm for S-DP problem

Next, we consider parallelizing the sequential algorithm for S-DP problem. The straightforward approach is to parallelize the inner loop by using GPU cores. We can easily write a multi-thread program that executes the inner loop-body, $\mathtt{ST}[\mathtt{i}] = \mathtt{ST}[\mathtt{i}] \otimes \mathtt{ST}[\mathtt{i} - a_\mathtt{j}]$, for each j in parallel using $k-1$ threads at one time. Such an implementation, however, does not improve the time cost, because every thread has access to the same ST[i] and thus memory access conflicts occur. As a result, those memory conflicts should be automatically solved at run-time by the serializing mechanism of GPU, and consequently the whole time-cost stays in $O(nk)$ steps, which is the same time cost as that of the sequential implementation. Further, even if those $k-1$ threads could operate in concurrent read and concurrent write (CRCW) mode, obviously we would not get the correct answer because all $k-1$ computations of $\otimes$ in equation (2) would be performed at one time.

To avoid the possible memory access conflicts, we can use a standard parallel prefix computation algorithm (e.g., [3, 4]), in which the computations of $\otimes$ over the k values are executed in a tournament fashion. Since the parallel prefix computation runs in $O(\log k)$ steps for k values, the entire time cost can be improved to $O(n \log k)$ steps even when we use k threads.

Although we can successfully reduce the time cost from $O(nk)$ to $O(n \log k)$ by using the parallel prefix computation, it is not work-time optimal because there are many idle threads during the computations in a tournament fashion. In the next section we propose other parallel implementation strategy and show that we can improve the time cost further.

3 Pipeline Implementation on GPU

In this section, we explain our parallel implementation technique for S-DP problem on GPU. Our program runs in a pipeline fashion.

3.1 Pipeline Implementation for S-DP problem

First, we introduce the pipeline implementation technique shown in our previous studies [5, 6]. In the next section, we will further accelerate the algorithm.

In our parallel implementation, we use a group of k threads to establish k-stage pipeline, and this thread group treats k consecutive elements of array ST at one time in parallel. Figure 2 describes our pipeline algorithm for the S-DP problem. The index variable i of the outer loop stands for the head position of the elements calculated by the k-thread group. The inner loop controls each thread's behavior in such a way that the thread j executes computation for $\mathtt{ST}[\mathtt{i} - \mathtt{j}]$ using the value stored in $\mathtt{ST}[\mathtt{i} - \mathtt{j} - a_\mathtt{j}]$.

A Pipeline Algorithm for S-DP Problem

for i = a_0 **to** $n+k-2$ **do**

  for j = 0 **to** $k-1$ **do** **in parallel**

    Thread j executes the following operation
    if $a_0 \le i_\mathtt{j} < n$ where $i_\mathtt{j} = \mathtt{i} - \mathtt{j}$:

$$\mathtt{ST}[i_\mathtt{j}] = \begin{cases} \mathtt{ST}[i_\mathtt{j} - a_\mathtt{j}]; & (\mathtt{j} = 0) \\ \mathtt{ST}[i_\mathtt{j}] \otimes \mathtt{ST}[i_\mathtt{j} - a_\mathtt{j}]; & (\mathtt{j} > 0) \end{cases}$$

Figure 2: A pipeline algorithm for S-DP problem

An execution example is shown in Figure 3, where $k = 3$, $a_0 = 6$, $a_1 = 3$, and $a_2 = 1$ hold and the initial values are stored in ST[0], ST[1],..., ST[5].

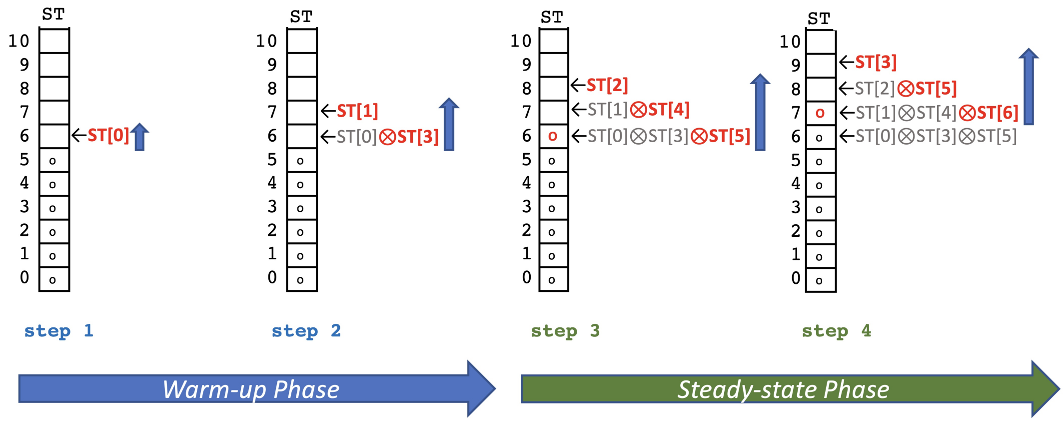


Figure 3: An execution example for the case where $k = 3$, $a_0 = 6$, $a_1 = 3$, and $a_2 = 1$ hold and the initial values are preset to ST[0], ST[1], ..., and ST[5]

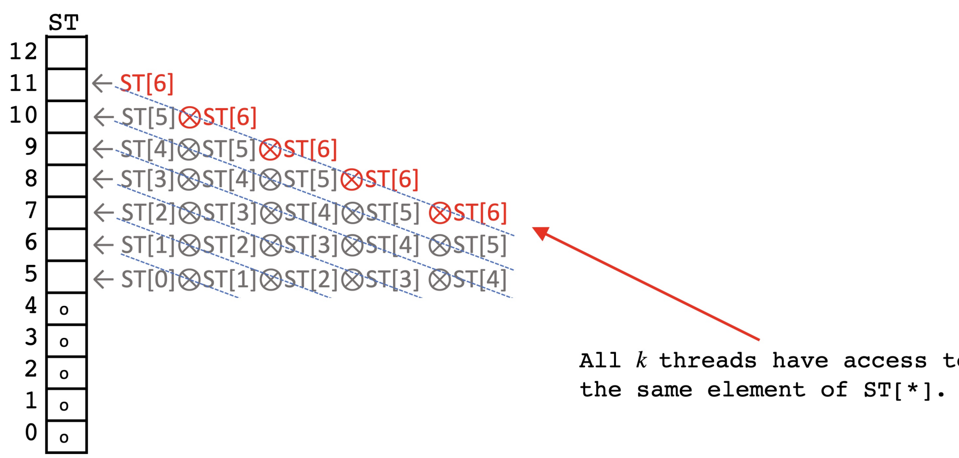


Figure 4: The worst-case example where the offset numbers are consecutively given. In this example, we have $k = 5$ threads and offset numbers are $a_0 = 5$, $a_1 = 4$, $a_2 = 3$, $a_3 = 2$, and $a_4 = 1$

In step 1, the head position i of the elements to be computed is 6. In this step the only one thread is activated and executes ST[6] ← ST[0]. In step 2, the head position is incremented to 7, and two threads are activated. The first thread treats ST[7] and the second thread works on ST[6]. We say that these steps 1 and 2 are in warm-up phase of pipeline, for not all threads are working yet and the number of active threads is gradually incremented one by each step. In step 3, the head position becomes 8, and now all $k = 3$ threads actively execute operations for ST[8], ST[7], and ST[6] respectively. From step 3, all k threads are steadily working and hence we say these steps are in steady-state phase. It should be noted that finally in step 3 the content of ST[6] is completely determined while those of ST[8] and ST[7] are partially computed and not yet fixed. From step 3, all the $k = 3$ threads are active until step $n - a_0$ when the head position i reaches $n - 1$, and after that step the number of active threads decreases one by each step. As you can see there is no memory access conflict in this example.

As for the time-complexity of our pipeline implementation, from a theoretical viewpoint, it takes only $O(n)$ steps if there is no memory access conflict, because the outer loop takes $n + k - a_0 - 1 = O(n)$ iterations and the inner loop requires $O(1)$ time.

From a practical viewpoint, the inner loop may take more time steps, because the possible memory access conflicts occur. In the worst-case when consecutive offset numbers are given, those ST$[i_j - a_j]$, in the right-hand side of the assignment statement, coincidentally become the same element of array ST and hence the worst memory access conflicts occur. In such a case, all threads in the inner loop are serialized and it takes time proportional to k. See Figure 4 for such a worst-case example. In this example, all five threads try to have access to the same ST$[i - 5]$ at one time in the inner loop. For such a case, we have proposed a *2-by-2 pipeline implementation* technique to avoid memory conflicts, where each thread invoked in the inner loop executes two computations for each element of array ST. The details can be found in [5].

3.2 Correctness of Pipeline Algorithm

Before we improve and accelerate the pipeline implementation in the next section, we consider the sufficient condition for the pipeline algorithm to work correctly and prove that the condition always holds and that our algorithm works correctly.

In the inner loop of the algorithm, in order to (partially) calculate the value of $\mathtt{ST}[i_\mathtt{j}]$, each thread j uses the value of $\mathtt{ST}[i_\mathtt{j} - a_\mathtt{j}]$ assuming that this value has been computed at the time it is referenced. In what follows, we show that the assumption is always true if equation (1) is valid.

In the inner loop (executed in parallel), the values of $\mathtt{ST}[i_\mathtt{j} - a_\mathtt{j}]$ used by the k threads are $\mathtt{ST}[\mathtt{i} - a_0]$, $\mathtt{ST}[\mathtt{i} - 1 - a_1]$, $\mathtt{ST}[\mathtt{i} - 2 - a_2]$,.... , $\mathtt{ST}[\mathtt{i} - (k-1) - a_{k-1}]$. By the equation (1), we can say that the largest value among these k referenced indices is $\mathtt{i} - (k-1) - a_{k-1}$. On the other hand, when the index of the outer loop is i, the elements of array ST have already been calculated up to $\mathtt{ST}[\mathtt{i} - k]$. With these observations, the following inequality must be true for the pipeline algorithm to work correctly:

$$\mathtt{i} - (k-1) - a_{k-1} \;\leq\; \mathtt{i} - k,$$

which leads to

$$1 \leq a_{k-1}.$$

Since the inequality $1 \leq a_{k-1}$ is always true by the equation (1), we can say that our pipeline algorithm correctly calculates the values of array ST.

4 Accelerating Pipeline Implementation on GPU

In the pipeline algorithm introduced in the preceding section, the parallelism is k. In this section, we consider increasing the degree of parallelism further to accelerate the computation.

4.1 Problem with Simultaneous Execution of Multiple Iterations

A simple idea for the acceleration is to let several iterations in the outer loop of the pipeline algorithm be executed in parallel at one time. Such an idea, however, may not work because of the following reason. Each computation of $\mathtt{ST}[i_\mathtt{j}]$ uses the value of $\mathtt{ST}[i_\mathtt{j} - a_\mathtt{j}]$ which had been computed up to that operation. See step 4 of Figure 3 as an example. In step 4, ST[3], ST[5], and ST[6] are referred to, though the computation of ST[6] had just been completed in the preceding step 3. Hence, in the case shown in Figure 3, while step x being executed, the value of ST that had just been calculated in the preceding step $x-1$ is immediately used for the step x, which means that multiple consecutive iterations of the outer loop of the pipeline algorithm cannot be executed at the same time.

However, if the offset values are sufficiently large, it is expected that the required elements for the next calculation had already been computed much earlier. In such a case, there is a possibility that several consecutive iterations of the outer loop may be executed together at the same time. For example, in Figure 5, since all elements up to ST[9] had already been calculated at the time of step 3, all the elements of ST required for the future steps 4, 5, 6, and 7 are already obtained. Hence, in addition to step 3, it seems that it is possible to execute these steps from step 4 to 7 at the same time as well. However, such an idea still arises two problems as follows: possible memory read conflicts and memory write conflicts. The first problem simply leads to the slowdown of execution because those possible memory conflicts should be automatically solved at run-time by the serializing mechanism of GPU. The second problem is from the concurrent write problem, and it is fatal because the calculation cannot be performed correctly. For example, let us consider the case where we simultaneously execute those 5 steps from step 3 to 7 in Figure 5. In such a case, ST[12], for example, is to be written by 3 threads: thread 0 in step 3, thread 1 in step 4, and thread 2 in step 5 try to execute $\mathtt{ST}[12] = \mathtt{ST}[2]$, $\mathtt{ST}[12] = \mathtt{ST}[12] \otimes \mathtt{ST}[4]$, and $\mathtt{ST}[12] = \mathtt{ST}[12] \otimes \mathtt{ST}[7]$, respectively. Obviously, we would not get the correct value by the simultaneous execution of these 3 operations.

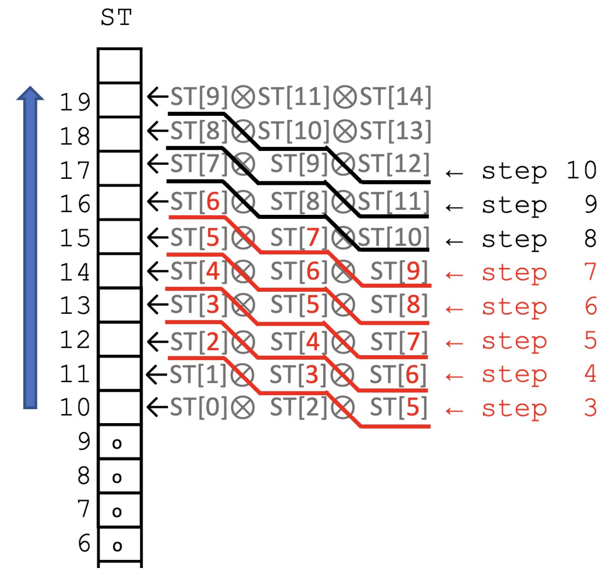


Figure 5: An example of problems with simultaneous execution of multiple iterations, where $k = 3$, $a_0 = 10$, $a_1 = 8$, and $a_2 = 5$

4.2 p-fold Pipeline Implementation on GPU

In this section, we propose a p-fold pipeline technique. Instead of simultaneous execution of multiple iterations of the outer loop in the pipeline algorithm of Figure 2, we modify the operations in the inner loop. The idea is to increase the number of threads for the inner loop.

Figure 6 describes the p-fold pipeline algorithm we propose. Here, the number of threads working in the inner loop is increased to pk from k. Figure 7 shows an execution example when $p = 3$. In the example, 9 ($= pk$) threads are activated during the inner loop, and the parallelism is increased by a factor of 3 ($= p$) compared to the original pipeline implementation where only 3 ($= k$) threads are activated. In our p-fold pipeline implementation, possible memory read conflicts may occur depending on the offset values, but the memory write conflicts does not. Hence, we can avoid the fatal problem discussed in the preceding subsection and say that our p-fold pipeline algorithm works correctly.

A p-fold Pipeline Algorithm for S-DP Problem

for i $= a_0 + (p-1)$ **to** $(n-1)+(k-1)p$ **by** p **do**

  for j $= 0$ **to** $pk-1$ **do** **in parallel**

    Thread j executes the following operation
    if $a_0 \le i_{\mathtt{j}} < n$ where $i_{\mathtt{j}} = \mathtt{i} - \mathtt{j}$:

$$\mathtt{ST}[i_{\mathtt{j}}] = \begin{cases} \mathtt{ST}[i_{\mathtt{j}} - a_{\lfloor \mathtt{j}/p \rfloor}]; & (\lfloor \mathtt{j}/p \rfloor = 0) \\ \mathtt{ST}[i_{\mathtt{j}}] \otimes \mathtt{ST}[i_{\mathtt{j}} - a_{\lfloor \mathtt{j}/p \rfloor}]; & (\lfloor \mathtt{j}/p \rfloor > 0) \end{cases}$$

Figure 6: A p-fold pipeline algorithm for S-DP problem

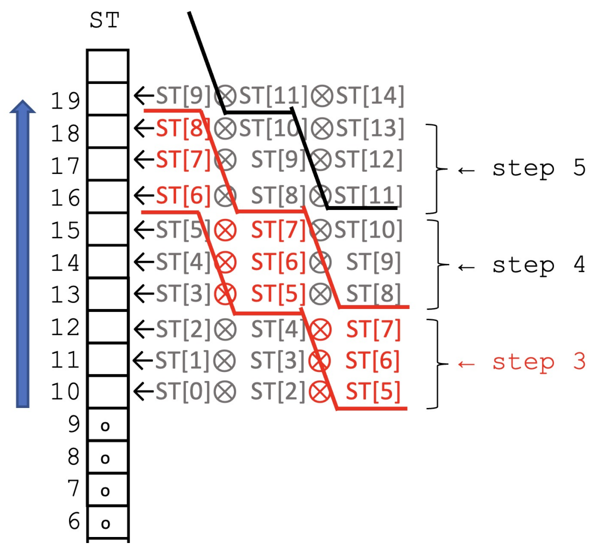


Figure 7: An execution example of the p-fold pipeline algorithm for the case where $p = 3$, $k = 3$, $a_0 = 10$, $a_1 = 8$, and $a_2 = 5$

As for the time-complexity of the p-fold pipeline algorithm, from a theoretical viewpoint, it takes $O(n/p)$ steps if there is no memory read conflict, because the outer loop takes

$$\frac{\{(n-1)+(k-1)p\} - \{a_0 + (p-1)\} + 1}{p} = O\left(\frac{n}{p}\right)$$

iterations and the inner loop requires $O(1)$ time. Since the number of threads is pk, the work is $O(pk \times n/p) = O(nk)$, which is equal to that of sequential algorithm in Figure 1.

4.3 Example: 0-1 Knapsack Problem

Let us consider an example of solving the 0-1 knapsack problem by DP as follows. Given a set of K items (each item is given a weight and a value) and the capacity C of a knapsack, the 0-1 knapsack problem is a problem to find a best way to pack the knapsack so that the total value is maximized within the capacity of the knapsack, by deciding which items to pack and which to exclude.

It is well-known that the 0-1 knapsack problem can be solved by dynamic programming using a two-dimensional solution table $\mathtt{ST}[*,*]$. Each $\mathtt{ST}[i,j]$ stores the best solution for the subproblem $\mathrm{P}(i,j)$, where $\mathrm{P}(i,j)$ is the problem of finding the best way to pack the knapsack that maximizes the value under the two conditions: 1) only up to the i-th item of given items are considered to be packed, and 2) the total weight capacity of the knapsack is limited to j. It should be noted that the original problem is $\mathrm{P}(K,C)$.

The $\mathrm{P}(K,C)$ can be solved by DP as follows. The solution for each $\mathrm{P}(i,j)$ is stored in the two-dimensional solution table $\mathtt{ST}[i,j]$ $(0 \le i \le K, 0 \le j \le C)$. Then, $\mathrm{P}(i,j)$ can be easily solved if the two solutions of its subproblems $\mathrm{P}(i-1,j)$ and $\mathrm{P}(i-1,j-w_i)$ are already known where w_i is the weight of the i-th item. That is, $\mathtt{ST}[i,j]$ can be easily obtained by only checking the two elements $\mathtt{ST}[i-1,j]$ and $\mathtt{ST}[i-1,j-w_i]$. To obtain $\mathrm{P}(K,C)$, we need to fulfill the two-dimensional solution table of size of $(K+1)$ rows and $(C+1)$ columns as Figure 8 in row-major order from the top-left corner to the bottom-right corner, which takes $O(KC)$ steps.

To apply our p-fold pipeline algorithm to this 0-1 knapsack problem, we need to convert the two-dimensional solution table $\mathtt{ST}[*,*]$ to the liner solution table $\mathtt{ST}[*]$. This conversion can be easily done by rearranging elements of $\mathtt{ST}[*,*]$ in row-major order so that each $\mathtt{ST}[i,j]$ is mapped to $\mathtt{ST}[i(C+1)+j]$. It should be noted that in the case of reducing the 0-1 knapsack problem to the S-DP problem, one of the two offset values, a_0, may dynamically change because it depends on w_i. In this sense, the 0-1 knapsack problem is not formulated strictly as an S-DP problem, but it is essentially the same and we can use the p-fold pipeline algorithm. In such a mapping, the calculation of $\mathtt{ST}[i]$ can be executed by using only the two elements of $\mathtt{ST}$ whose indices are at least $(C+1)$ smaller than i. In other words, the offset values of S-DP problem are at least $(C+1)$, and thus we can expect to use the p-fold pipeline algorithm with a large parallelism p (e.g., at least more than C) for solving this 0-1 knapsack problem.

5 Experimental Results

In this section, we show experimental results about the expected number of memory access collision and the number p of the p-fold pipeline implementation.

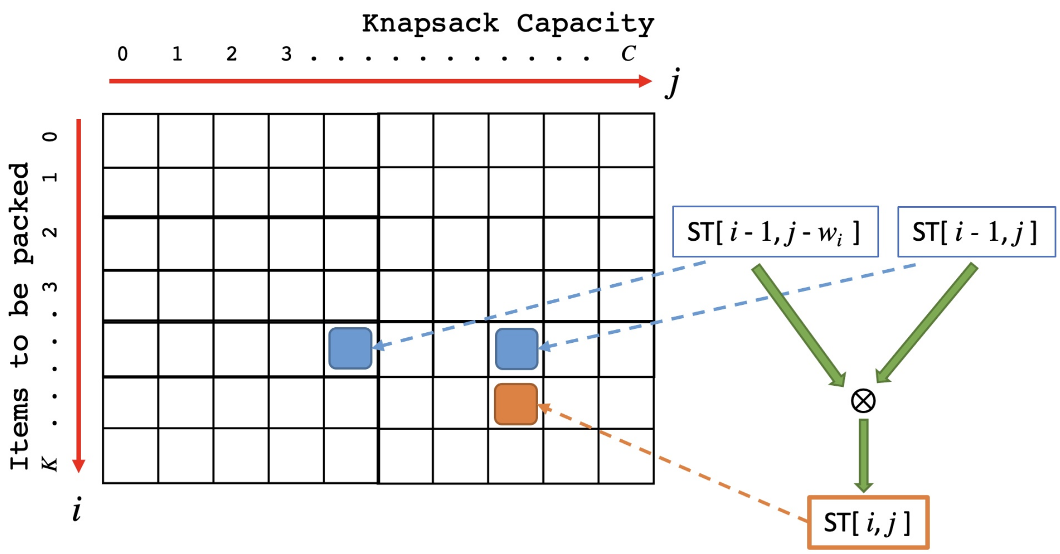


Figure 8: A two-dimensional solution table for the dynamic programming solving 0-1 knapsack problem. Each $\mathtt{ST}[i,j]$ can be calculated by the two elements $\mathtt{ST}[i-1,j]$ and $\mathtt{ST}[i-1,j-w_i]$ in the row one above

5.1 Collision Count of 1-Fold Pipeline Implementation

While the sequential algorithm described in Figure 1 only executes one $\otimes$ operation per step, our pipeline implementation in Figure 2 can process k number of $\otimes$ operations in a step by using k threads. That is, our pipeline implementation solves an S-DP problem at a parallel degree of k. However, from a practical viewpoint, there is a possibility that memory access congestions (concurrent reads to the same element $\mathtt{ST}[i]$) may occur in the pipeline implementation, resulting in performance degradation.

Figure 9 shows simulation results showing collision counts with various K and r. The parameter K is the upper limit of offset values and r is the ratio for controlling the density of selected offset values from the offset value pool $\{1,2,\ldots,K\}$. That is, we choose rK number of offset values from the integer set $\{1,2,\ldots,K\}$. For ratio r we take 0.1, 0.2, 0.3, 0.4, and 0.5. We count the collision counts with moving K from 10 to 640, at a doubling rate of increase. For each pair (r,K), we randomly generate 1000 sets of offset values, and count the maximum number of read collisions occurred in a step. From Figure 9, we can see that the collision counts get larger as K and/or r increase. It should be noted that the rate of increase, however, is not large when K increases.

For more detail on the rate of increase with the parameter r, see Figure 10, which shows the number of expected collisions with increasing the density of offset values in its range $\{1,2,\ldots,K\}$. Here, we fix $K=50$. When the ratio r is approaching to 1.0, the number of collisions is getting closer to K. This is because we must choose all K offset values from $\{1,2,\ldots,K\}$ when $r=1.0$, which is the worst-case (the case where all offset values are consecutively selected) scenario.

5.2 Number p of p-Fold Pipeline Implementation

In the p-fold pipeline implementation in Figure 6, the larger the value of p, the greater the degree of parallelism is expected, and the program execution speed becomes faster. However, it is not possible to choose an arbitrarily large value of p. This is because at each step of execution, each of the kp threads respectively reads out $\mathtt{ST}[*]$ for computation and those $\mathtt{ST}[*]$ values must have been completely computed at that point in time. See Figure 11 for example. Although the offset value settings are the same in Figure 7 and 11, in Figure 7 the 3-fold pipeline implementation is feasible, while in Figure 11 the 4-fold pipeline is not.

Figure 12 shows simulation results of how large p can be expected for p-fold pipeline implementation with increasing the lower limit (gap) g of offset values. Here, we fix $K=100$ and $r=0.2$, and select rK offset values from integer set $\{g,g+1,\ldots,K+g\}$. As discussed in Section 4.3, we can see that the number p becomes larger as the gap value g increases. For example, it is expected that we can execute 9-fold pipeline implementation for an S-DP problem of $k=20$ $(=r*K)$ offset values if all offset values are more than 100. In such a case, we can expect higher degree of parallelism like 180 $(=p*k=9*20)$ by using possibly 9-fold pipeline implementation, while the degree of parallelism is only 20 $(=k)$ for the simple pipeline implementation in Figure 2.

Figure 13 shows simulation results of how large p can be expected for p-fold pipeline implementation with increasing the density of offset values in its range $\{g,g+1,\ldots,K+g\}$. Here, we fix $K=100$ and $g=20$. We can see that the p decreases rapidly when ratio r is approaching to 1.0. This is because we must choose all K offset values from $\{g,g+1,\ldots,K+g\}$ when $r=1.0$, and as a result the worst-case scenario where all offset values are consecutively selected occurs.

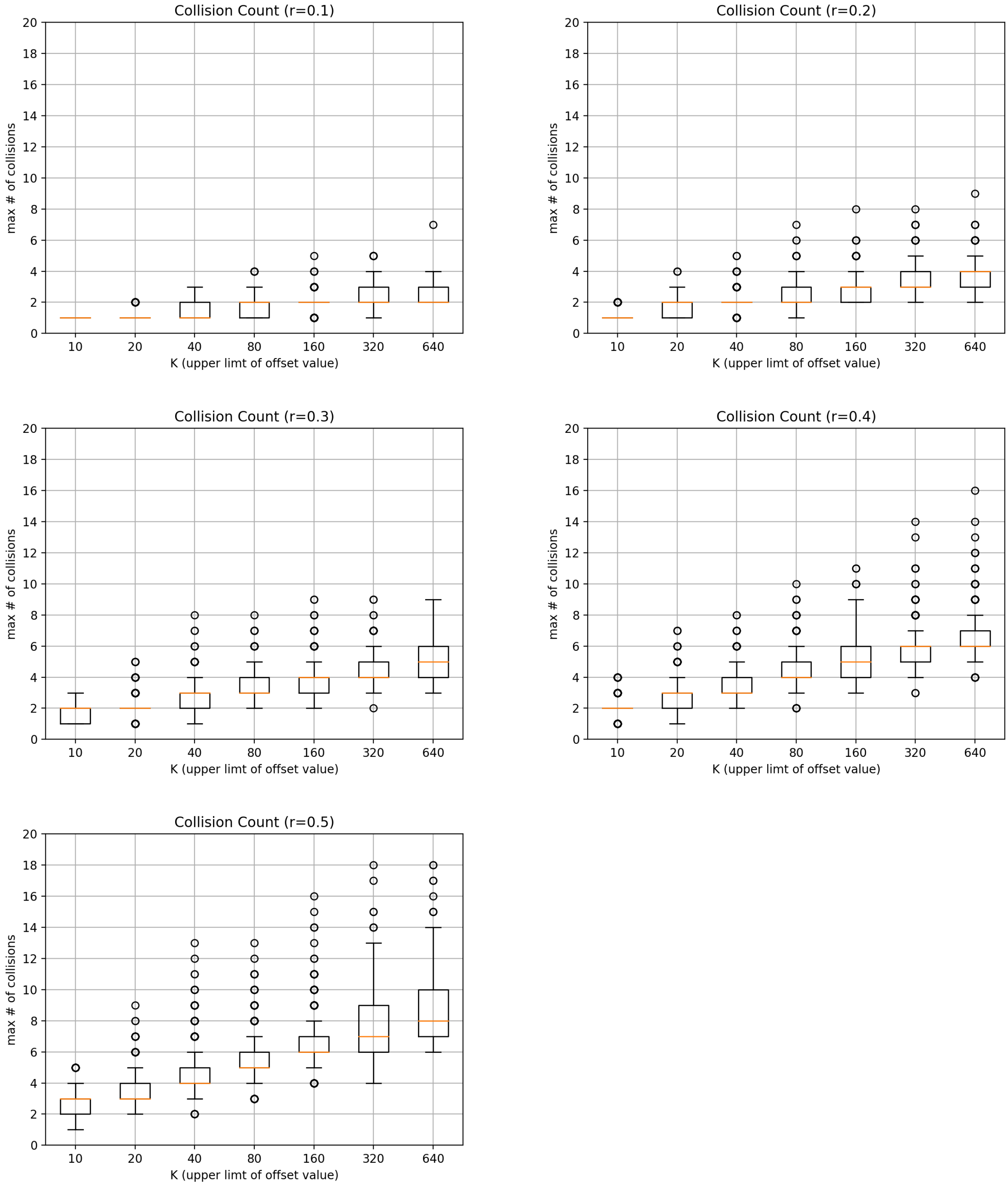


Figure 9: Number of collisions with increasing the upper limit K of offset values. From integer set $\{1,2,\ldots,K\}$, $r*K$ number of offset values are selected. The horizontal axes are of logarithmic scale

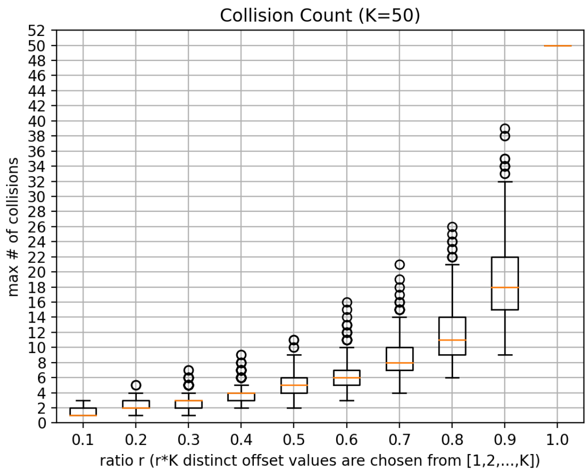


Figure 10: Number of collisions with increasing the density of offset values in its range $\{1,2,\ldots,K\}$. Here, we set $K = 50$

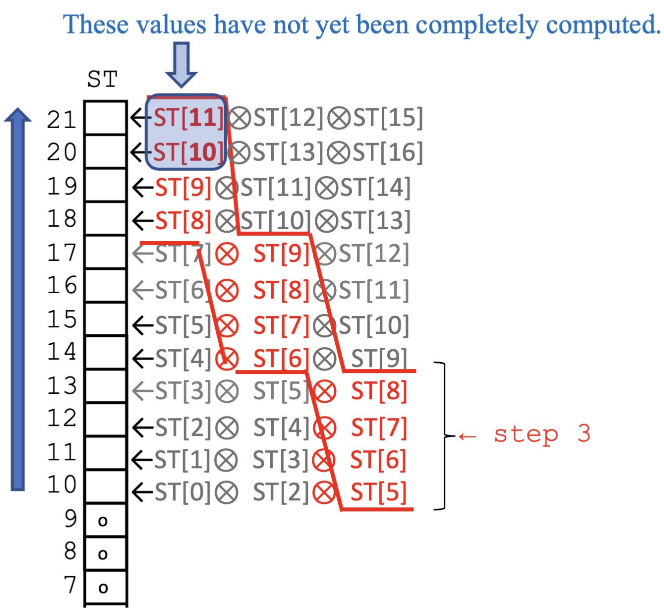


Figure 11: An IMPOSSIBLE execution example of the p-fold pipeline algorithm for the case where $p = 4$, $k = 3$, $a_0 = 10$, $a_1 = 8$, and $a_2 = 5$

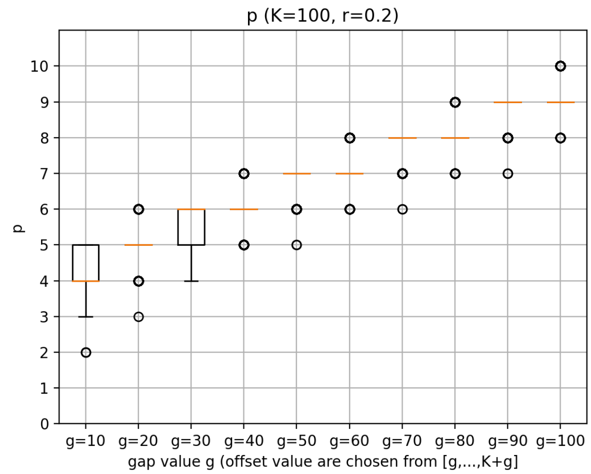


Figure 12: Number p of p-fold pipeline implementation with increasing the lower limit (gap) g of offset values. From integer set $\{g, g+1, \ldots, K+g\}$, $r * K$ offset values are selected. Here, we set $K = 100$ and $r = 0.2$

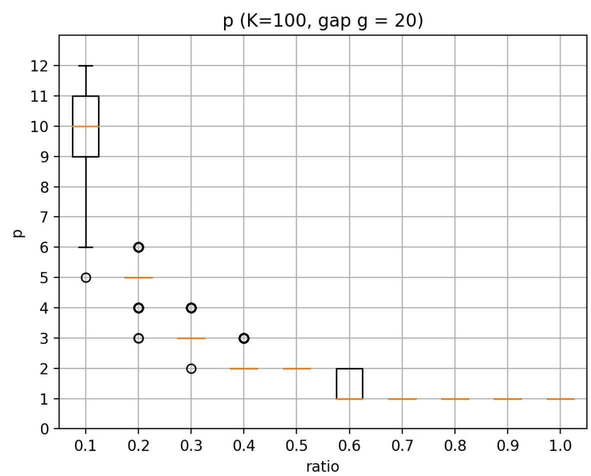


Figure 13: Number p of p-fold pipeline implementation with increasing the density of offset values in its range $\{g, g+1, \ldots, K+g\}$. Here, we set $K = 100$ and $g = 20$

5.3 Collision Count of p-Fold Pipeline Implementation

Figure 14 shows simulation results of the collision count with increasing the lower limit (gap) g of offset values. We can see that collision counts tend to decrease as g increases. The intuitive explanation is as follows. In the p-fold pipeline implementation, at each step, let (idx_1, idx_2, ..., idx_{kp}) be the index sequence of `ST[*]` read by those kp threads in order of thread number th. Then, if those indices are grouped by separating every p indices in order from the first idx_1, it is guaranteed that no index collision will occur within such a group of size p. Thus, since the increase of p implies the increase of size of collision-free index groups, the collision count is expected to decrease.

6 Concluding Remarks

In this study, we examined the effectiveness of pipeline implementations of Dynamic Programming (DP) on GPU. We dealt with a simplified DP problem where each element of its

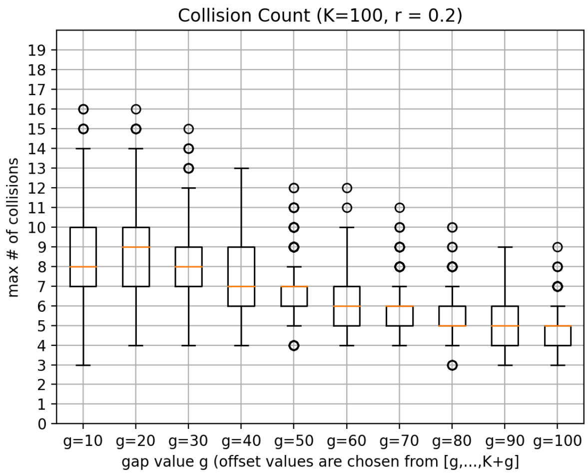


Figure 14: Number of collisions with p-fold pipeline implementation with increasing the lower limit (gap) g of offset value. From integer set $\{g, g+1, \ldots, K+g\}$, $r * K$ offset values are selected. Here, we set $K = 100$ and $r = 0.2$

one-dimensional solution table of size n is calculated in order by semi-group computations among several of k already computed elements in the table and proposed pipeline implementations on GPU model. In our previous studies [5, 6], the degree of parallelism of our pipeline is up to k, but in this paper, we proposed the p-fold pipeline implementation technique and successfully increased the parallelism toward pk, though the parameter p depends on the problem to be solved. As an application example, we explained how to apply our p-fold pipeline technique to the 0-1 knapsack problem. And our experimental results showed that we can expect a large p for the p-fold pipeline.

For future work, we plan to evaluate the performance of our pipeline implementations by conducting experiments on GPU. We also plan to study the performance of our pipeline implementation on theoretical GPU models (e.g., [8, 9]).

References

[1] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to Algorithms, Third Edition*. The MIT Press, 3rd edition, 2009.

[2] Y. Ito and K. Nakano. "A GPU Implementation of Dynamic Programming for the Optimal Polygon Triangulation." *IEICE Transactions on Information and Systems*, E96.D(12):2596–2603, 2013.

[3] V. Kumar, A. Grama, A. Gupta, and G. Karypis. *Introduction to Parallel Computing: Design and Analysis of Algorithms*. Benjamin-Cummings Publishing Co., Inc., Redwood City, CA, USA, 1994.

[4] F. T. Leighton. *Introduction to Parallel Algorithms and Architectures: Array, Trees, Hypercubes*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1992.

[5] M. Miyazaki and S. Matsumae. "A Pipeline Implementation for Dynamic Programming on GPU." In *International Workshop on Parallel and Distributed Algorithms and Applications (in Conjunction with CANDAR'18)*, Nov. 2018. doi: 10.1109/CANDARW.2018.00063

[6] S. Matsumae and M. Miyazaki. "Solving Dynamic Programming Problem by Pipeline Implementation on GPU." *International Journal of Advanced Computer Science and Applications*, 9(12):518–523, 2018. doi: 10.14569/IJACSA.2018.091272

[7] S. Matsumae. "Accelerating Pipeline Implementation of Dynamic Programming on GPU." In *the 35th International Conference on Computer Applications in Industry and Engineering (CAINE 2022)*, EPiC Series in Computing, 89:52–61, 2022. doi: 10.29007/6mgp

[8] K. Nakano, S. Matsumae, and Y. Ito. "The Random Address Shift to Reduce the Memory Access Congestion on the Discrete Memory Machine." In *2013 First International Symposium on Computing and Networking*, pp.95–103, Dec. 2013. doi: 10.1109/CANDAR.2013.21

[9] K. Nakano and S. Matsumae. "The Super Warp Architecture with Random Address Shift." In *2013 20th International Conference on High Performance Computing (HiPC)*, pp.256–265, Dec. 2013. doi: 10.1109/HiPC.2013.6799118

[10] K. Nakano. "A Time Optimal Parallel Algorithm for the Dynamic Programming on the Hierarchical Memory Machine." In *2014 Second International Symposium on Computing and Networking (CANDAR)*, pp.86–95, 2014.

[11] S. H. Roosta. *Parallel Processing and Parallel Algorithms: Theory and Computation*. Springer, 1999.

Susumu Matsumae is a professor at the Department of Information Science, Faculty of Science and Engineering, Saga University, Japan. He received the M.E. and PhD degrees in computer science from Osaka University in 1996 and 2000, respectively. From 2000 to 2001, he was a research associate of Graduate School of Engineering Science, Osaka University. From 2001 to 2007, he was with Department of Information Systems, Tottori University of Environmental Studies. In 2007, he joined Department of Information Science, Saga University. His research interests include parallel algorithms and architectures, logic, and computational complexity.

Development and Evaluation of a Process Management and Analytics Platform for Small and Medium-sized Enterprises

Jonas Kallisch
Emden/Leer University of Applied Sciences, Emden, Germany

Christoph Wunck†
Emden/Leer University of Applied Sciences, Emden, Germany
Iowa State University, Ames, USA
OFFIS Institute for Information Technology, Oldenburg, Germany

Abstract

This paper presents a prototype for an information exchange system, which allows information exchange between companies without actually sharing data. First, the need for such an intercompany exchange platform is explained and the value for supply chains resulting from such a platform is described. A literature review presents the existing concepts and techniques contributing to the development of an architecture. Finally, the information exchange concept and the prototype implementation are explained in detail.*

Key Words: Small and medium enterprises (SME), analytics, machine learning, data exchange.

1 Introduction

Machine learning, advanced analytics and other methods have become a major element of modern production systems. Reduction of waste, improvement of manufacturing timings and quality are some of the contributions that these methods have brought to the companies own manufacturing infrastructure [17]. Because of this fact, many companies have started to analyze their shop floor data to profit from the benefits described. The increasing availability of customizable analytic tools and the decreasing of their prices enables even small businesses to use them. As a result, companies are more empowered than ever to identify and address the vulnerabilities of their manufacturing infrastructure. Parallel to this development, supply chain cooperation has deepened, which means that value creation depends more than ever on the cooperation of the companies[11].

Despite this reliance on collaboration, many companies, especially small and medium enterprises (SME), do not share data with their customers or suppliers and so do not optimize their joint manufacturing [23]. There are many reasons for this lack of data exchange. An interview series conducted by the authors investigated into these reasons, found that a common issue is that companies have recognized the value of the data captured in manufacturing, and fear disadvantages from sharing it.

The fear is that the companies could lose control over the data shared in the supply chain (SC), for example if the data is stolen [5]. In some cases, the losing of control about their intellectual property could even threaten the existence of the affected company, because competitors could use the information to improve their own production. Therefore, the fear of many companies of sharing data from their manufacturing systems is justified and understandable. It can be concluded that the first requirement of companies for a system for collaborative analysis of data in a supply chain is the guarantee to maintain data sovereignty. Ideally, a solution should be sparing with the sharing of data or completely avoid the sharing of raw data. Requirement #1 is therefore the preservation of the company's data sovereignty.

Another issue for SMEs is their lack of technical knowledge to create their own platforms and systems to exchange data with other SMEs [32]. SMEs often work on a low budget with external technical service providers, which must be able to support the solution of the company to exchange data with a customer or a supplier. This situation, which particularly affects SMEs, results in the requirement that a solution must fit the company's budget and must also be available in a form that can be implemented and maintained by service providers. Requirement #2 can be described as companies expecting a lightweight solution.

The advantages of sharing data rely on the ability to analyze them and understand the results of the analysis. Many SMEs don't have these abilities within their businesses, as our interview partners told us and other studies found out as well [30]. Thus, the third requirement for a solution can be described as the ability of companies to perform cross-company analyses on their own or at least in cooperation with their IT service provider. Requirement #3 can be summarized with the term applicability with the capabilities of the companies.

However, there are numerous reasons for analyzing data across companies or even for sharing data, even if this sharing

* Funded by the Lower Saxony Ministry of Science and Culture under Grant Number ZN3489 within the Lower Saxony "Vorab" of the Volkswagen Foundation and supported by the Center for Digital Innovations (ZDIN).

† Department of Computer Science.

contradicts the claim to maintain data sovereignty. By waiving to share shop floor data in the supply chain they give up many opportunities to improve the overall production efficiency along the value creation inside their SC [23].

A first step to data integration is to reduce the costs of data generation inside the SC [1]. Much of the data used by the participants is collected multiple times for different purposes. This results in costs that could be lowered by sharing the data between the members of the SC. In addition, the companies that are involved in the SC can be seen as members of a community of trust, as they should have a shared interest to improve the results of the group like quality or efficiency.

A second step towards integrating SC data is to create new data through a combination of already existing data. By combining data from different stages relations between incidents appearing in late stages of the manufacturing process and data patterns appearing at earlier stages of the SC can be found [24]. This helps to reduce waste and optimize the overall production system.

To summarize, it can be stated that data exchange on a shop-floor level is not widespread and SMEs in particular rarely exchange manufacturing data with other companies. The authors propose that in addition to the lack of trust between companies, the existing solutions for data exchange do not meet all the requirements of companies.

To demonstrate this, the paper will first present the existing architectures and evaluate them with a view to the identified requirements #1, #2 and #3. Therefore, particular attention will be paid to the aspects of data security and usability for SMEs. The paper will show that existing architectures and concepts do not meet the requirements for all companies and that a concept is therefore needed for the secure networking of corporate data assets. Such a concept and a prototype have been developed by the "Zukunftslabor Produktion" [Futurelab Manufacturing] (ZLP) [35]. This interdisciplinary project is developing solutions and models especially for small and medium businesses in the manufacturing sector. Based on a use-case of the production of die-cast aluminium parts, the project examines how technical manufacturers can use data analytics and other industry 4.0 methods to improve their shop floor systems and create more resilient SCs. The concept and the reference implementation will be presented in the third part of this paper. The paper will close with a discussion of the findings.

2 Existing Concepts and Solutions

To identify existing concepts and implementations of intercompany communication in SCs, more than 100 papers have been examined and evaluated in a structured literature review. This section presents the concepts identified by the literature review and shows which SME requirements they do not meet.

2.1 Data Spaces

Data spaces have been defined by Franklin, Havely and Maier as a next step in the evolution of data integration architectures [8]. The evolutionary step is that data spaces combine storing of data with services, to merge data from different sources to extract information. Key of this approach is the integration of data from different domains and the mapping of their different data elements. Therefore, data spaces can represent a SC as a number of related data sources, which can be connected. This data connection allows the whole system to improve its value. Similar to a SC the data spaces' value depends on the level of compatibility – mapping and matching – between the different suppliers [27].

One implementation of these data Spaces is the International Data Space (IDS) [9]. The IDS is a system of data providers interacting on a platform. Each participant can be provider or user of data and has the right to negotiate about the rights on the data. The platform allows to connect the data on a contract base. The full model of the IDS is displayed in Figure 1.

The model consists of data providers and consumers that hold their data in their own physical storage locations. However, these physical storages do not interact directly with each other. The interaction is controlled by the broker, which offers two main services to the participants. Firstly, the broker lists and categorizes the data delivered by the data providers. This service enables the searching companies to find the data they need. The companies can search based on labeled data or based on the types of data they already have – e.g., a certain type of

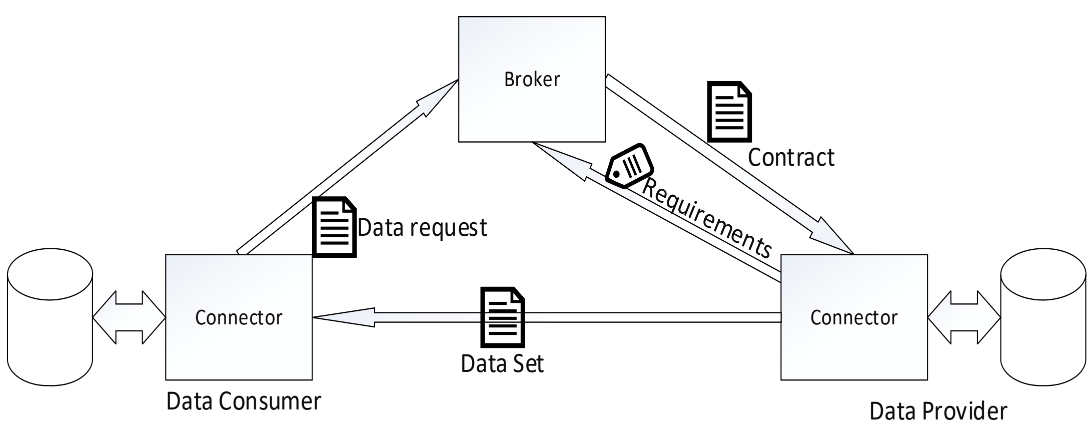


Figure 1: Model of the IDS [13]

application or asset. If the system includes a data provider with the needed data, the broker connects the two – or more – companies. As the data can be traded anonymously, it is even possible to buy data from trusted sources without knowledge about the very company supplying the data.

Eventually, a service provider intermediating between the entities is part of the model. A service provider can be part of a whole ecosystem of service providers. These can for example be infrastructure services – e.g. data storage, computing power or other infrastructure – that allow participants to be part of the data space without having their own hardware. [29] Beyond this Infrastructure as a Service (IaaS), other providers can add services – e.g., data analysis or searching services – or Software as a Service (SaaS) – e.g., analytics tools or data mining software – via a dedicated App Store. [24] Therefore, the IDS delivers a solution for sharing data between companies on a contract base. However, the concept explicitly assumes sharing of data and thus cannot preserve the data sovereignty of the participants. Therefore, the concept cannot be used to network companies without exchanging data and does not meet the requirement #1.

2.2 GAIA-X

The European GAIA-X project can be considered as a solution for connecting different kinds of data stores. These data stores might also be IDS infrastructure elements. It contributes an architecture concept that allows the sharing of data in a public catalog, where everybody can see the available data, but can only access data when the access is granted [10]. Similar to IDS brokers, GAIA-X enables companies to give permission for sharing data on an individual level.

Core component for the concept is the federated catalog. This is a public catalog, where companies can publish the data they are able to share. Every company connected with the catalog is able to see the data that is available, but can only access data when the access is granted by the data owner. Therefore, the companies can choose to give permission for sharing data, on an individual level.

The GAIA-X foundation guarantees this kind of data sovereignty as a service. This means that participants have the capability to fully self-determinate their data exchange and sharing. The secure exchange is realized by a function called Data Contract Transaction. This service initiates a handshake between the data provider and the requesting party. The service validates the contract and, if the content is valid and both parties have confirmed the transaction, the Data Contract Service distributes the Data Contract to both companies. After that, the requesting company can access the requested data and may analyze it. The distribution of data is observed by a function called Data Exchange Logging, which enables companies to restrict the usage of their data to a certain extent or for a specific purpose.

The model of GAIA-X allows sharing of data in a secure and customizable way but still needs to actually exchange the data to analyze them inside of the SC. A very interesting part of the solution is the way the catalogue combines data identification and services by self-description. The value of this for companies that are interested in sharing data has also been examined by Dumss et al. [6]. They suggest an architecture model called EuProGigant, which allows exchanging data in a scalable way. They also describe how services can enrich the generated data and emphasize the importance of self-description in GAIA-X. However, they did not give a suggestion how to secure the data exchange or how to keep the intellectual property of the companies. This means that the concept, as it is proposed at this point, is not able to protect the data ownership interests of the SC companies.

Summed up the GAIA-X foundation provides a reliable, effective and secured solution for sharing data. For companies that are interested in sharing or selling their data the GAIA-X Federated Catalogue is a fitting solution. In the case of an interconnected SC GAIA-X might also be a very good communication platform, but by itself is not able to create an information exchange without actually exchanging the data. By that, the author's opinion is that GAIA-X does not preserve data sovereignty and therefore misses requirement #1.

2.3 Catena-X

A development based on GAIA-X is the concept of Catena-X. As an architectural concept, Catena-X offers a decentralized data platform system that enables companies to exchange and use data securely. [12] The platform is based on a smart contract system and an open data architecture. In the concept of Cantena-X, data exchange is realized by the so-called Eclipse Dataspace Connector (EDC).

The EDC enables participants to exchange data from their internal systems, such as ERP, CRM or PLM systems. [22] To do this, a smart contract is created in the platform between the participants who want to exchange data. This contract defines the conditions of the exchange and manages the process of data sharing.

Figure 2 shows the structure of the Catena-X platform and how the data exchange is managed. As can be seen, the components of the federated catalog, as well as the Identity Service of GAIA-X are used in the concept. The contract is concluded via these and guarantees that the transaction partners and offered data are correct. Unlike in the GAIA-X concept, the data exchange itself does not take place via the central platform. Instead, each participant uses its individual EDC in one of two ways. Data can either be performed via direct transfer between the EDC systems of the respective companies, which is indicated in Figure 2 as Option A, or via a service provider, shown as Option B. The second option can also include other data services, such as conversion or analysis. With SMEs in mind, it is also possible to operate the EDC as software as a service. These can therefore participate in the common data analysis without having to build up their own know-how.

Overall, the concept of Catena-X enables the secure sharing of data between companies. The concept of locating data assets within the system is well suited for use between partners in a supply chain, as it is based on GAIA-X. The data can be transferred directly between the partners. The direct transfer of

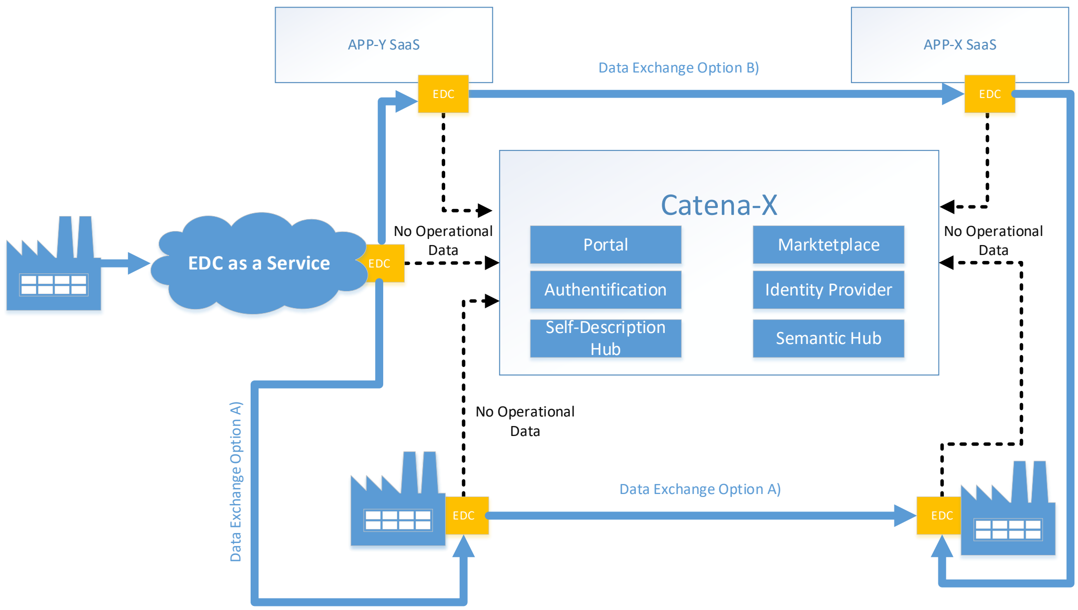


Figure 2: Architecture of Catena-X (based on [15])

data between companies addresses one of the caveats to analyzing data across companies. Because the data is not stored centrally, it is easier to protect data ownership and from unauthorized access to the data. Nevertheless, even in this architecture, sharing data on an individual level becomes necessary. This means that protection against unauthorized access becomes more difficult with each additional location or company where the data is stored as a copy. Besides this problem, the concept of Catena-X does not offer protection against misuse of the data by a partner within the supply chain. The fears of many SMEs of competitive disadvantages due to the loss of control over their know-how can therefore not be solved by Catena-X. Therefore, requirement #1 is not met, even if Cantena-X enables a better control of the data, since these are not stored in a central data storage.

2.4 Federated Learning

Federated Learning (FL) is a concept to analyze datasets, which are distributed over different devices that are connected with a central station [36]. It can be divided in horizontal and vertical FL [34]. The difference between these two types of FL is the selection of elements they share. As Figure 3 shows horizontal FL shares features, e.g., temperature measuring or other kind of data points, but not the samples – a concrete case of measurement, while vertical FL shares samples, but not features.

The more common case of FL is the horizontal FL as it is used in mobile devices, to improve their ability to analyze their user data, without transferring them. A horizontal FL starts with an initial algorithm, which is created on a data sample. In a second step, the model is decomposed into sub models, matching the data elements of the different storages. The different data stores, e.g., smartphones, then train the model on their own data. In the fourth and last step the results of the individual models can be transferred to the central application to improve the model. The result is that the different data stores are able to improve their analytical models of their data without sharing with each other. The limitation of the horizontal FL is that it requires similar

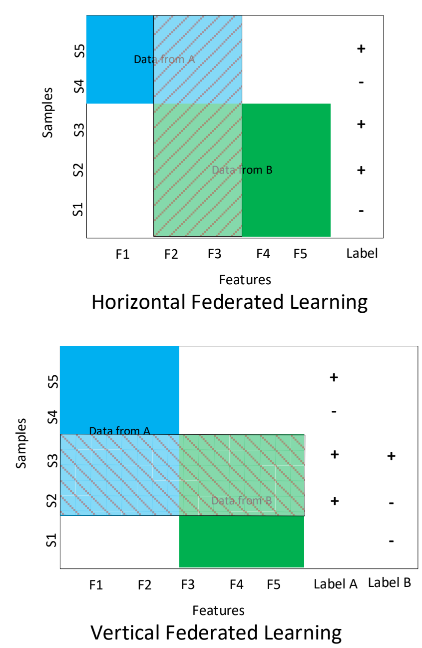


Figure 3: Difference between types of federated learning (based on [34])

data structures on the different devices.

In SCs, that do not share common data architectures between different companies, vertical FL can be used. As shown in Figure 3 the data stores in a vertical FL model are not sharing the same features – data structures – but the different data stores share the same samples. For example, two companies in the same city might not collect the same data, but collect the data from the same customers. If these companies share a common interest, they could combine their data and use it to improve the quality of their prediction algorithms.

Therefore, the usage of vertical FL requires to exchange the data or at least the labels of the data between the companies. A way to solve this problem of exchanging data between the company might lay in the work of Yang, Liu, Chen and Tong [34]. They propose a framework for secured vertical FL, which allows participants to exchange analytical models but frees them from having to share their data with each other. According to this, the issue could be considered as solved, but even these variant of vertical FL comes with some limitations.

The first limitation of the model is that it still requires the companies involved to share their labels of all samples to improve the training of the model. This might lead to leakage of information possible as Bagdasarya, et.al. have shown [4]. The second limitation comes with some assumptions on the data of the participants. One of these is, that all participants are sharing the same labels and have a combined goal, e.g. reduction of waste or overall cost reduction [16]. This would be an issue, if a company is involved in different SCs, with different goals or might profit from a certain kind of waste. The third and last-mentioned limitation relates to the value of information contributed by each individual party involved in the vertical FL. The problem is called ‘unbalanced clients’. This means that some participants of the SC are able to contribute more to the whole system than others. The issue of this is that a federated learning architecture isn’t able to balance these different features, without exchanging of datasets [36]. It should be mentioned that Zhang, et.al. have found that by selecting an adaptive number of local training rounds for each party can lead to better models, but this also increases the danger of data leakage. Obviously, FL does not deliver a lightweight solution within the capabilities of SMEs. Though FL bears great potential as a concept preserving data sovereignty, requirements #2 and #3 are not fully met.

2.5 Commercial Solutions

Beside the implementation of any of the concepts evaluated in the preceding sections the question remains if there is any existing commercial solution to securely connect SME SCs without giving up their individual data ownership interests. The authors have investigated various available products to see if commercial solutions are available that solve the problem away from scientific issues. The products investigated are described below and, in the authors’ view, represent a good range of the systems available on the market.

- SAP Business One

SAP Business One is focused on small businesses and is able to deliver a ERP system that does not share much with the main SAP Products S3 and S4 [28]. Companies can chose to run the software on their own server or use a cloud server hosted by SAP [26]. The product allows built-in analytics and SC automation of business transactions [25]. Business One is a solution that can help small companies to get to the earlier described state of the art of SC communication. It also allows the companies to get analysis of their business decisions, but is not able to solve the issue of intercompany communication.

- SAP Business ByDesign

SAP Business ByDesign aims at medium businesses [25]. The product is cloud based and provides a customizable ERP system, which can be integrated in a SC [14]. The solution is able to create build in analytics to create real time dashboards of the company´s situation [28]. In addition to the features of BusinessOne, it delivers the features of SC Management, like functions to support sourcing and purchasing [28]. As BusinessOne, Business ByDesign did not involve the SC partners manufacturing infrastructure.

- Microsoft Dynamics

Microsoft Dynamics delivers functions for integration of warehouses, material flow planning and collaboration with other companies [19]. The SC management component delivers similar functions to the given by the SAP product [7]. The review also found that all investigated ERP systems enable the integration of suppliers on the level of business communication. Another study also shows the state of ERP MES integration in Microsoft Dynamics and the lag of a integration of intercompany shop floor data exchange in the solution [21]. A current development of Microsoft Dynamics is the product Dynamics 365. The product allows the usage of so called “supply chain data sharing templates”. These templates provide options for cloud-based sharing of transactional information, like payment and logistics data, sharing of order pool, warehouse management for example for commission purposes or delivery management.

- Microsoft Azure

Microsoft Azure enables collaboration between companies in a SC with a cloud application that integrates inventory, factory status and logistical data in a twin of the SC elements [18]. The core focus is on using machine learning, optimization algorithms and artificial intelligence on distributed data to improve the decision making inside of the SC. An element called control tower creates a dashboard, which shows this state. The concept of these control towers is that a SC member delivers all relevant data to the gateway and connects them to the data of the other members. The results can be combined with public data, for example weather information [31]. An example for an SC using Azure can be found in the agriculture sector. These use cases show how the connection of different SC data storages can improve the whole SC's success, but also shows that the members involved must be willing to share their data, as the architecture provides the risk of losing control about the data. It should also be mentioned that the project focus was not on connecting SC members but on connecting singular IoT solutions and product used by different farmers involved. Therefore, the translatability to industrial SC´s may not be given.

2.6 Conclusion

This section has found that existing approaches to creating cross-enterprise data sharing do not meet all of the requirements identified in this paper. In particular, the guarantee of data sovereignty is insufficiently fulfilled by the existing solutions. Nevertheless, the existing concepts provide a basis for building a solution that meets all requirements. Table 1, on the next side, summarizes which requirements are met by which solutions.

3 Towards a Data Securing Information Exchanging Architecture

As the previous chapter has shown, none of the existing concepts or products meet all requirements to solve the problem of information exchange between companies without direct data exchange. Based on the requirements and the existing approaches, an architecture was developed that on the one hand offers the option to interact with data ecosystems such as IDS or GAIA-X in the future and on the other hand protects the control over the intellectual property of the companies. Our concept provides for the sharing of information via a linking platform, which neither requests data for centralized processing, nor passes it on to other participants in the platform. To create these links, existing connections in the data traffic can be used. In an SC, these usually consist of the flow of goods between the respective companies.

In many cases, these goods are identifiable, for example via a batch or serial number. These identifiers can be used to find and link the products within the cross-company workflow and the data collected during processing.

In theory, the data values recorded by one company *A* therefore, have a direct influence on the measurement results of a second company *B*. If company *B* finds that different results are determined for comparable process variables in its own data stock, the cause can therefore lie in a deviating process variable in company *A* or in a pre-production stage of *A*. By using a process support platform, these correlations can be analyzed without direct data exchange. In addition, the leakage of information is less likely, since only in defined cases, analysis will be started and the results do not even have to be shared between the companies.

This description of a possible network results in a concept whose core is a value-oriented relationship between at least two partners. As displayed in Figure 4 the partners are connected via a process support platform that distributes information based on the stored exchange relationships. The participants would be company *A* with data *A1* and the derived information *A2*, company *B* with data *B1* and the derived information *B2*. *A2* and *B2* are exchanged in the scenario. Thus, participant *A* holds *A1, A2* and *B2* and participant *B* holds *A2*, *B1* and *B2*, whereby each participant can determine the concrete information provided. Based on this core concept, the participants of the process support platform can improve their own analyses through the exchange and agree on individual agreement services. Such a service can be, for example, a cross-company fault cause recognition, which is passed on to supplier *A* in the event of a fault at *B* that cannot be explained by its own processes. *A* would then be obliged to determine whether this error can be traced back to processes in its production.

The concept for such a data saving and cross-company information sharing architecture consists out of three layers:

- Communication layer
- Service layer
- Analytics layer

The first of these layers, the communication layer, is intended to establish data exchange between the individual enterprise is to be transferred to link the data sets of the various supply

Table 1: Fulfillment of requirements

	Protection of data sovereignty	Lightweight solution	Capabilities of companies
Data Spaces	requirement not met	requirement not met	requirement partially met
GAIA-X	requirement not met	requirement partially met	requirement fully met
Catena-X	requirement partially met	requirement partially met	requirement partially met
Federated Learning	requirement fully met	requirement not met	requirement not met
Commercial Solutions	requirement not met	requirement fully met	requirement fully met

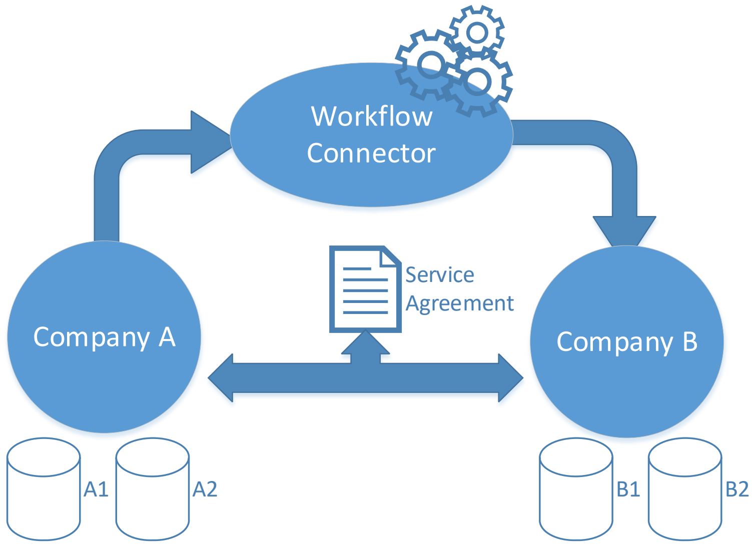


Figure 4: Concept of the project

chain participants. The correct selection of these data points and information is thus not monitored or determined by an internal or external actor. From the point of view of the concept developers, this is not necessary, since all participants in a supply chain should be interested in optimizing joint value creation. Nevertheless, it is conceivable to have the selection and set-up performed by a common, preferably independent, service provider. This would also have the advantage that knowledge of the entire data structure of all participants can be incorporated into the establishment of the link.

The second layer is used to implement services between the participating companies. These services allow companies to create individual applications for their supply chain. The layer in the interconnection platform describes these services and stores the rights of the respective participants to use the services. Requests to use services are forwarded via the communication layer and lead to the creation of an agreement to use the service. It would also be conceivable to extend the service system to an open system in which providers of data or analytics services could also interact in the supply chain. Essentially, however, this would require a partner to provide hosting of the platform and enable it for companies not in the supply chain. A connection with existing architecture models, such as IDS or GAIA-X would also be an option for the integration of these external services.

The final layer of the concept is for cross-enterprise data analytics. The layer is intended to provide a set of basic analytics that can be used by companies within their data platforms. Results of the cross-company analyses are then exchanged via the link in the communication layer. Here, participants should be able to restrict the sharing of data to prevent the extraction of data through targeted queries. This layer is implemented with the use of federated learning models in mind, as these provide a suitable basis for implementing cross-enterprise analyses. In order to compensate for the identified weaknesses of federated learning, companies should be able to configure the analysis models themselves and thus influence the respective parameters in their access area. The goal is thus to compensate for the problem of unbalanced clients.

This model resulted in a basis for the implementation of the architecture with one of ZLP's industrial partners. This forms the basis for answering further research questions and optimizing the model.

4 Prototype of a Supply-Chain Workflow Connector

The concept described in the previous section was developed in the form of a prototype in cooperation with an industry partner of the ZLP. This prototype is to be tested on industrial data to determine the potential of the ZLP's described data-saving concept. For this purpose, the analysis results of the prototype will be compared with a classical, central data analysis. By looking at the differences between the two approaches, any potential for optimizing the system will be identified.

In preparation for implementation, the data sets intended for exchange were transferred to a data platform. This serves to separate the operational infrastructure from the infrastructure connecting the supply chain. In view of the large number of different formats within the data sets provided by the practice partners, these individual database systems were realized in the form of customizable data platforms. For the selection of the data platform system, different systems were compared with each other. The architectures compared included the aforementioned commercial solution Microsoft Azure, which is used in a combination with Apache Kafka [2] as a data stream message broker, MongoDB [20], Hadoop [3] and Apache Superset [33]. Due to the industry partner’s specifications, only systems with OnPremise options were evaluated.

The various systems were built as examples and evaluated in terms of their utility values and the effort required to build them. Since one of the ZLP's goals is to provide a modular solution that is easy for companies to implement, the software's deployability for companies with little IT knowledge of their own was also considered a key criterion. All of the platform systems considered are basically capable of meeting the requirements of the ZLP. In terms of the complexity of the structure, the various systems examined did not differ significantly from one another. With regard to the performance of the platforms, differences can be seen in specific situations, which were also described by other studies.

As a decision for an architecture it was decided to first realize two platforms that represent two actors within the supply chain of the ZLP and to show the transferability of the concept. These form between the tool mold fabricator and the aluminum die caster. Hadoop and MongoDB were implemented as the systems. As a system, these offer many advantages in terms of extensions and existing analysis systems.

The systems are linked via the communication layer. This was realized with the help of the multiagent system Jade. Jade is a framework that enables the creation of agents that can also exchange messages and data packets across systems.

The respective agents are used to communicate with the central Workflow Connector and can make requests to other participants in the platform. Which rights and which relationships exist between the companies regulates how the requests are handled and forwarded in the platform. The companies are able to configure these relationships themselves and also select the respective data elements.

Based on this link with Jade, services or agents were first defined that enable the exchange of quality parameters via the platform. On the basis of these services, quality deviations of the created parts are to be compared with the history of a tool without merging the data. The analysis of this data will take place in a next step, once different options for implementing the analysis layer have been evaluated and the most suitable solution has been identified.

5 Conclusion and Further Work

This paper described the need for a secure information exchange system for companies in SCs. The current lack of suitable concepts and solutions was shown. The authors devised an architecture that protects the data ownership of the involved parties and enables information sharing. The advantages and disadvantages of the architecture have been briefly discussed and will be further evaluated within the project prototype and be addressed in future work. The concept achieves networking through an exchange platform that allows companies to exchange information based on individual contracts without having to disclose data. The concept adopts some mechanisms and ideas of existing concepts. On the one hand, this improves the quality of the architecture by using tried and tested systems and, on the other hand, it enables companies to integrate into platforms such as IDS or GAIA-X. In the future, the authors will build further implementation variants, such as a microservice architecture, and investigate them with regard to their potential. In addition, the ZLP will evaluate the prototype with support of our industrial partners. For this purpose, experimental setups with industrial partners and on-site laboratory infrastructure will be used.

References

[1] V. Alcácer and V. Cruz-Machado, "Scanning the Industry 4.0: A Literature Review on Technologies for Manufacturing Systems," *Engineering Science and Technology, an International Journal*, 22(3):899-919, 2019, doi: 10.1016/j.jestch.2019.01.006.

[2] Apache Kafka, *Apache Kafka,* [Online], Available: https://kafka.apache.org/ (accessed: Feb. 16 2023).

[3] Apache Software Foundation, *Apache Hadoop,* [Online], Available: https://hadoop.apache.org/ (accessed: Jun. 28 2022).

[4] Eugene Bagdasaryan, Andreas Veit, Yiqing Hua, Deborah Estrin, and Vitaly Shmatikov, "How to Backdoor Federated Learning," *International Conference on Artificial Intelligence and Statistics*, pp. 29382948, 2020, [Online], Available: https://proceedings.mlr.press/v108/bagdasaryan20a.html

[5] C. Colicchia, A. Creazza, C. Noè, and F. Strozzi, "Information Sharing in Supply Chains: A Review of Risks and Opportunities Using the Systematic Literature Network Analysis (SLNA)," *SCM*, 24(1):5-21, 2019, doi: 10.1108/SCM-01-2018-0003.

[6] Stefan Dumss Markus Weber, Clemens Schwaiger, Clemens Sulz, Patrick Rosenberger, Friedrich Bleicher, Manfred Grafinger, Matthias Weigold*.*, "EuProGigant – A Concept Towards an Industrial System Architecture for Data-Driven Production Systems," *Procedia CIRP*, 104:324-329, 2021, doi: 10.1016/j.procir. 2021.11.055.

[7] F. M. Elbahri, O. I. Al-Sanjary, M. Ali, Z. A. Naif, O. A. Ibrahim, and M. N. Mohammed, *Difference Comparison of SAP, Oracle, and Microsoft Solutions Based on Cloud ERP Systems: A Review*, 2019, [Online], Available: https://www.researchgate.net/profile/omar-al-sanjary/publication/333457778_difference_comparison_of_sap_oracle_and_microsoft_solutions_based_on_cloud_erp_systems_a_review/links/5ffbad6292851c13fe02ecd5/difference-com-parison-of-sap-oracle-and-microsoft-solutions-based-on-cloud-erp-systems-a-review.pdf

[8] M. Franklin, A. Halevy, and D. Maier, "From Databases to Dataspaces," *SIGMOD Rec.*, 34(4):27-33, 2005, doi: 10.1145/1107499. 1107502.

[9] Fraunhofer-Gesellschaft, *International Data Spaces,* [Online], Available: https://www.fraunhofer.de/en/research/lighthouse-projects-fraunhofer-initiatives/international-data-spaces.html (accessed: Aug. 27 2020).

[10] *GAIA-X: Technical Architecture,* [Online], Available: https://www.data-infrastructure.eu/GAIAX/Redaktion/EN/Publications/gaia-x-technical-architecture.html (accessed: Jul. 13 2021).

[11] M. Holweg, S. Disney, J. Holmström, and J. Småros,

"Supply Chain Collaboration," *European Management Journal*, 23(2):170-181, 2005, doi: 10.1016/j.emj. 2005.02.008.

[12] *Home | Catena-X,* [Online], Available: https://catena-x.net/en/ (accessed: Feb. 16 2023).

[13] International Data Space Association, *International Data Space Reference Architecture Modell,* [Online], Available: https://www.fraunhofer.de/content/dam/zv/en/fields-of-research/industrial-data-space/IDS-Reference-Architecture-Model.pdf, 2019.

[14] K. Kurbel and D. Nowak, "Customization of On-Demand ERP Software Using SAP Business ByDesign as an Example," *Innovation and Future of Enterprise Information Systems*: Springer, Berlin, Heidelberg, pp. 289-297, 2013, [Online], Available: https://link.springer.com/chapter/10.1007/978-3-642-37021-2_21

[15] W. Lehmann, "Präsentation Konzept Cantena-X," [Online], Available: https://catena-x.net/fileadmin/user_upload/Vereinsdokumente/Catena-X_UEbersicht.pdf, 2022.

[16] L. Lyu, H. Yu, J. Zhao, and Q. Yang, "Threats to Federated Learning," *Federated Learning*: Springer, Cham, pp. 3-16, 2020, [Online], Available: https://link.springer.com/chapter/10.1007/978-3-030-63076-8_1.

[17] J. A. Marmolejo-Saucedo, M. Hurtado-Hernandez, and R. Suarez-Valdes, "Digital Twins in Supply Chain Management: A Brief Literature Review," *2020 – Intelligent Computing and Optimization*, pp. 653-661, 2020.

[18] Microsoft Corporation, *Advancing Reliability through a Resilient Cloud Supply Chain | Azure Blog and Updates | Microsoft Azure,* [Online], Available: https://azure.microsoft.com/en-us/blog/advancingreliability-through-a-resilient-cloud-supply-chain/ (accessed: May 30 2022).

[19] Microsoft Corporation, *Supply Chain Management | Microsoft Dynamics 365,* [Online], Available: https://dynamics. microsoft.com/de-de/supply-chain-management /overview/ (accessed: May 30 2022).

[20] MongoDB, *MongoDB Data Lake,* [Online], Available: https://www.mongodb.com/de-de/atlas/data-lake (accessed: Feb. 16 2023).

[21] S. Oman, R. Leskovar, B. Rosi, and A. Baggia, "Integration of MES and ERP in Supply Chains: Effect Assessment in the Case of the Automotive Industry," *Teh. vjesn.*, 24(6):1889-1895, 2017, doi: 10.17559/tv-20160426094449.

[22] projects.eclipse.org, *Eclipse Dataspace Components,* [Online], Available: https://projects.eclipse.org/projects/technology.edc (accessed: Feb. 16 2023).

[23] P. Radanliev, D. C. de Roure, J. R. C. Nurse, R. M. Montalvo, and P. Burnap, "Supply Chain Design for the Industrial Internet of Things and the Industry 4.0," 2019.

[24] K. Rong, G. Hu, Y. Lin, Y. Shi, and L. Guo, "Understanding Business Ecosystem Using a 6C Framework in Internet-of-Things-Based Sectors," *International Journal of Production Economics*, 159:41-55, 2015, doi: 10.1016/j.ijpe.2014.09.003.

[25] SAP, *ERP-Software für kleine Unternehmen | SAP Business One,* [Online], Available: https://www.sap.com/germany/products/business-one.html (accessed: May 25 2022).

[26] SAP, *Übersicht Produkte | Unternehmenssoftware für KMU | SAP Mittelstand,* [Online], Available: https://www.sap.com/germany/products/sme-business-software/products.html (accessed: May 30 2022).

[27] A. D. Sarma, X. Dong, and A. Y. Halevy, "Data Modelling in Dataspace Support Platforms," *Lecture Notes in Computer Science*, *Conceptual Modeling: Foundations and Applications: Essays in Honor of John Mylopoulos*, A. T. Borgida, V. K. Chaudhri, P. Giorgini, and E. S. Yu, Eds., Berlin: Springer, 5600:122-138, 2009.

[28] K. Schneider, "SAP Business By Design – Integrated Cloud Suite," (in De;de), *HMD*, 55(1):44-61, 2018, doi: 10.1365/s40702-017-0380-8.

[29] S. Shah, M. Bolton, and S. Menon, "A Study of Internet of Things (IoT) and its Impacts on Global Supply Chains," *2020 International Conference on Computation, Automation and Knowledge Management (ICCAKM)*, Dubai, United Arab Emirates, pp. 245-250, Jan. 2020.

[30] J. Stentoft, K. W. Jensen, K. Philipsen, and A. Haug, "Drivers and Barriers for Industry 4.0 Readiness and Practice: A SME Perspective with Empirical Evidence" *Proceedings of the 52nd Annual Hawaii International Conference on System Sciences: January 8-11, 2019, Maui, Hawaii*. Honolulu, HI: University of Hawaii at Manoa Hamilton Library Scholar Space, 2019, [Online], Available: https://scholarspace.manoa.hawaii.edu/handle/10125/59952.

[31] J. Sunny, N. Undralla, and V. Madhusudanan Pillai, "Supply Chain Transparency Through Blockchain-Based Traceability: An Overview with Demonstration," *Computers & Industrial Engineering*, 150:106895, 2020, doi: 10.1016/j.cie.2020.106895.

[32] S. Wang and H. Wang, "Big Data for Small and Medium-Sized Enterprises (SME): A Knowledge Management Model," *JKM*, 24(4):881-897, 2020, doi: 10.1108/JKM-02-2020-0081.

[33] *Welcome | Superset,* [Online], Available: https://superset.apache.org/ (accessed: Feb. 16 2023).

[34] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated Machine Learning," *ACM Trans. Intell. Syst. Technol.*, 10(2):1-19, 2019, doi: 10.1145/3298981.

[35] ZDIN, *Zukunftslabor PRODUKTION,* [Online], Available: https://www.zdin.de/zukunftslabore/produktion (accessed: Jun. 27 2022).

[36] Jie Zhang Song Guo, Zhihao Qu, Deze Zeng, Haozhao Wang, Qifeng Liu, Albert Y. Zomaya, "Adaptive Vertical Federated Learning on Unbalanced Features," *IEEE Trans. Parallel Distrib. Syst.*, p. 1, 2022, doi: 10.1109/TPDS.2022.3178443.

Jonas Kallisch is a research assistant at the Emden/Leer University of Applied Sciences (Germany). He is researching the networking of manufacturing companies and their manufacturing infrastructures as part of the Future Lab Production. He holds a master's degree in industrial engineering. His research interests are data analytics, digital business ecosystems and federated data architectures.

Christoph Wunck is a full professor of Business Information System at Emden/Leer University of Applied Sciences (Germany). He serves as a member of the executive board "manufacturing" at OFFIS Institute of Information Technology in Oldenburg (Germany). He was appointed Affiliate Professor with Iowa State University, Department of Computer Science. He holds a doctoral degree from the Faculty of Mechanical Engineering and a master's degree (Diplom-Ingenieur) in Electrical Engineering, both from RWTH Aachen University. His research interests cover technical architectures for digital business ecosystems and analytics in manufacturing execution systems.

Efficient and Secured Data Lookup Protocol using Public-Key and Digital Signature Authentication in RC-Based Hierarchical Structured P2P Network

Anjila Neupane*, Reshmi Mitra *, Indranil Roy *
Southeast Missouri State University, Cape Girardeau, MO 63701, USA

Bidyut Gupta †
Southern Illinois University, Carbondale, IL 62901, USA

Narayan Debnath ‡
Eastern International University, VIETNAM

Abstract

Peer-to-peer (P2P) networks are extremely vulnerable to network security attacks due to their low budget and limited resource capabilities. In this study, we are presenting public key cryptography and digital signature-based data lookup protocol that are efficient and secured in an RC-based hierarchical structured P2P network. Utilizing RC, a modular arithmetic-based residue class, the overlay topology has been achieved. This design was chosen because it allows for minimal latency in both intra and inter-group communications. In the present study, we provide efficient schemes for public-key cryptographic security and digital signature-based authentication for existing communication protocols.

Key Words: Public-key cryptography; digital signature; P2P networks ; interest-based network formation

1 Introduction

Due to their capacity to offer computational and data resource sharing in a scalable, self-organizing, distributed manner, peer-to-peer (P2P) overlay networks are widely used in distributed systems. P2P networks are divided into two categories: unstructured networks and structured networks. Peers in unstructured systems [2] are arranged in any random topology. For data lookup, flooding is necessary. In unstructured systems, problems brought on by frequent peer joining and leaving the system, or "churn", are effectively handled. However, this compromises the effectiveness of data querying and the crucial flexibility. Lookups are not guaranteed in unstructured networks. On the other hand, structured overlay networks offer deterministic limits on data discovery. They create scalable network overlays based on a distributed data structure that truly allows deterministic data lookup behavior.

The usage of distributed hash tables (DHTs) is a recent trend in the design of structured overlay systems [5, 9, 19]. Overlay designs of this type can provide efficient, flexible, and resilient service [5, 9, 11, 19, 20]. However, addressing the churn issue while maintaining DHTs becomes expensive. It requires significant transformation in designing an effective data query service. There are numerous notable publications in this area that have examined creating hybrid systems [4, 14, 17, 21]. These works make an attempt to incorporate the benefits of both structured and unstructured structures. These works, however, have their own set of advantages and disadvantages [1].

The study and application of encrypted communication protocols is known as cryptography. It is concerned with the development and examination of protocols that prohibit harmful third parties from gaining access to information transferred between two companies, hence complying to many principles of information security. Secure communication refers to a scenario in which a message or data transmitted between two parties cannot be accessed by an adversary. An adversary in cryptography is a malevolent actor who seeks to get usable information or data by violating information security rules. Cryptographic techniques are used to ensure authentication and confidentiality stability in peer-to-peer networks. The two most common forms of cryptographic algorithms are secret key cryptographic algorithms and public key cryptographic algorithms. Secret key cryptographic algorithms are also known as symmetric key algorithms since the same key is used for encryption and decryption and is shared by all parties involved. Public key cryptography algorithms, on the other hand, are also known as asymmetric key algorithms. This version employs a pair of keys, one for encryption and the other for decryption [12].

In [8] the authors have used a well-defined approach of combining the symmetric key and the public key cryptography methods to ensure security in the architecture. The main issue with symmetric key encryption is secured key delivery among sender and receiver. This requires reliable communication network that are completely resistant to any form of privacy attacks which can be expensive and often highly unlikely in the context of P2P network. Shared symmetric key can be vulnerable to any man-in-the-middle attack as the same key is being used reused for encryption and decryption. In[16] the authors suggested robust public-key cryptography approaches for the security of current communication protocols in [8] with comparably lower costs to overcome the aforementioned issues with employing symmetric key. They have added anonymity to these strategies as well.

However, asymmetric encryption can be more expensive than symmetric encryption but it does allow us to overcome the downfall's of using symmetric encryption. Similarly the authors in [16] have considered both an intra-group and an inter-group lookup technique in an residue class(RC)-based architecture [8] with secured protocols. Asymmetric key cryptography has been used with public keys. Moreover, they have suggested secured capacity-constrained multicast algorithms for the two-level architecture as well as for use inside groups. They have also considered anonymity. We have defined the RC based architecture in section 2.

The authors of this [16] article presupposed that all of the peers can be trusted. This situation has been taken into consideration in this work. The main communication node in the RC-based system is the group-heads. The invader will try to breach the group-heads in order to destabilize the entire network. In this work, we provide secure data-lookup algorithms while considering the aforementioned threat models.

1.1 Our Contribution

In this research, interest-based P2P systems have been taken into consideration [21, 22]. We have thought about creating secure protocols for both intra and inter group lookup algorithm. Public keys for asymmetric key cryptography and digital signature authentication have been utilized. The rest of the paper is organized as follows. We give the preliminaries and the overview of the RC-based 2-level non-DHT-based structured P2P network proposed in [8] in section 2. The secured data lookup and transfer algorithms for both the inter and the intra group along with the threat model are explained in section 3. In section 4, we present the performance evaluation. Finally, we conclude in section 5.

2 Preliminaries

Here, we have taken into consideration some of the first results of an RC-based low diameter two level hierarchical structured P2P network [6, 7, 13]. We provide a structured design for an interest-based peer-to-peer system in this section. We will use the following notations and their meanings to define the architecture.

Definition 1. We define a resource as a tuple $<Res_i$, V $>$, where Res_i denotes the type of a resource and V is the value of the resource. Note that a resource can have many values.

Definition 2. Let S be the set of all peers in a peer-to-peer system. Then $S = \{P^{Ri}\}$, $0 \leq i \leq n-1$, where P^{Ri} denotes the subset consisting of all peers with the same resource type Res_i. and the number of distinct resource types present in the system is n. Also, for each subset P^{Ri}, we assume that G_i^h is the first peer among the peers in P^{Ri} to join the system. We call G_i^h as the group-head of group G_i formed by the peers in the subset P^{Ri}.

We now describe our proposed architecture suitable for interest-based peer-to-peer system. Generalization of the architecture is considered in [7].

2.1 Residue Class

Modular arithmetic has been used to define the residue class(RC)-based architecture of the P2P system. Consider the set S_n of non negative integers less than n, given as $S_n = 0,1,2,\dots(n–1)$. This is referred to as the set of residues, or residue classes (mod n). That is, each integer in S_n represents a residue class (RC). These residue classes can be labelled as [0], [1], [2], ..., [n – 1], where $[r] = \{a$: a is an integer, $a \equiv r(mod n)\}$.

For example, for n = 3, the classes are:
[0] = {..., - 6, - 3, 0, 3, 6, ...}
[1] = {..., - 5, - 2, 1, 4, 7, ...}
[2] = {..., - 4, - 1, 2, 5, 8, ...}

Each integer used to represent a residue class in the P2P architecture serves as the logical (overlay) address for the group-head of a group. For example, logical address of the first group-head is 0, for the second one it is 1, and so on. The logical (overlay) addresses of peers with a common interest (i.e., peers in the same group) are represented by integers belonging to various classes, and the number of residue classes corresponds to the number of different resource types. For the sake of simplicity, only positive integer values are used for addressing. It becomes obvious that any class may technically contain an endless number of integers, which indicates that there is no upper limit on the size of a group.

2.2 Two Level Hierarchy

It is a two-level overlay architecture and at each level structured networks of peers exist. It is explained in detail below.

1. At level-1, we have a ring network consisting of the peers G_i^h $(0 \leq i \leq n-1)$. The number of peers on the ring is n which is also the number of distinct resource types. This ring network is used for efficient data lookup and so we name it as transit ring network.
2. At level-2, there are n numbers of completely connected networks (groups) of peers. Each such group, say G_i is formed by the peers of the subset P^{Ri}, $(0 \leq i \leq n-1)$, such that all peers $(\in P^{Ri})$ are directly connected (logically) to each other, resulting in the network diameter of 1. Each G_i is connected to the transit ring network via its group-head G_i^h.
3. Based upon whether a peer in the network, a group-head or a regular peer, they maintain Information Resource Table (IRT) and Local Resource Table (LRT) that consists of n number of tuples. The group-heads maintains both Information Resource Table (IRT) and Local Resource

Table (LRT), where IRT contains list of the other group-heads, while LRT contains list of the peers present in a group. The regular peers in a group only maintain the LRT.

- The group heads will have a tuple of the form <Group Head Logical Address, Group Head IP Address, Group Head public key>for other group heads and <Peer Logical Address, Peer IP Address, peer public key>for the other peers present in the network. The Group Head Logical Address are assigned according to the proposed logical address assignment algorithm proposed in section 2.2 and the public key of the group heads or the peers are exchanged when they are joining the network and the IRT is updated and broadcasted in the network. Also, Resource Code is the same as the group head logical address.
- The peers P_i, who are not group heads but belongs to a group G_i ($P_i \in G_i$) will have the tuple of the form <Group Head Logical Address, Group Head public Key>for group head of G_i and <Peer Logical Address, Peer public Key>for the other peers present in G_i.

4. Any communication between a peer $G_{x,i} \in$ group G_x and $G_{y,j} \in$ group G_y takes place only through the corresponding group heads G_x^h and G_y^h.

The proposed architecture is shown in Figure 1.

2.3 Assignments of Overlay Addresses

Assume that in an interest-based P2P system there are n distinct resource types. Note that n can be set to an extremely large value a priori to accommodate a large number of distinct resource types. Consider the set of all peers in the system given as S = $\{P^{Ri}\}$,$(0 \leq i \leq n-1)$. Also, as mentioned earlier, for each subset P^{Ri} (i.e. group G_i) peer G_i^h is the first peer with resource type R_i to join the system. The assignment of logical addresses to the peers at the two levels and the resources happen as explained in [7-8, 12].

Remark 1. IRT remains sorted with respect to the logical addresses of the group-heads.

Definition 3. Two peers G_i^h and G_j^h on the ring network are logically linked together if (i + 1) mod n = j.

Remark 2. The last group-head H_{n-1} and the first group-head P_0 are neighbors based on Definition 3. It justifies that the transit network is a ring.

Definition 4. Two peers of a group G_r are logically linked together if their assigned logical addresses are mutually congruent.

Lemma 1. Diameter of the transit ring network is $n/2$.

Lemma 2. Each group G_r forms a complete graph.

2.4 Salient Features of Overlay Architecture

We summarize the salient features of this architecture.

1. It is a hierarchical overlay network architecture consisting of two levels; at each level the network is a structured one.
2. Use of modular arithmetic allows a group-head address to be identical to the resource type owned by the group.
3. Number of peers on the ring is equal to the number of distinct resource types, unlike in existing distributed hash table-based works some of which use a ring network at the heart of their proposed architecture [11].
4. The transit ring network has the diameter of n/2. Note that in general in any P2P network, the total number of peers N >>n.
5. Every resource type will have two group-heads, to maintain a fault-tolerance system and secure communications which are explained in the following sections.
6. Each overlay network at level 2 is completely connected. That is, in graph theoretic term it is a complete graph consisting of the peers in the group. So, its diameter is just 1. Because of this smallest possible diameter (in terms of number of overlay hops) the architecture offers minimum search latency inside a group.

3 Data Lookup Algorithms with Public Key Cryptography

This part introduces data lookup techniques [6, 13], both intra and inter, with the idea of security using public key cryptography and digital signature for authentication. Figure 2 explains how cryptographic operations are used in a two-level RC-based design.

3.1 Public key distribution among group-heads and peers

We begin with a straightforward technique that allows all group heads to be aware of one another's public keys. Suppose that there are now k groups in the network, with the group leaders' logical addresses ranging from 0 to (k-1). Specifically, the biggest resource code available right now in IRT is (k-1). The group head G_0^h initially determines if the joining peer's resource type is already available in the IRT when a peer p having instance(s) of a certain resource type requests to join. Now, one of the two subsequent circumstances may happen.

Situation 1: Resource type of peer p does not exist in IRT

1. New peer p contacts G_0^h for joining the network.
2. IRT holds resource codes from 0 up to (k-1) before peer p joins, therefore the group-head G_0^h assigns the joining peer with the next greatest available number for the code; as a result, the code for p becomes k.
3. G_0^h makes entry in the IRT for the new resource code k (which is now the logical address of the newly joining peer p) and the IP address of peer p, because now peer p becomes the group-head G_k^h.
4. G_0^h and G_k^h exchange their public keys, namely PU_0 and PU_k. Their respective private keys are PR_0 and PR_k. Group-head G_0^h updates a list T (IRT) by including PU_k

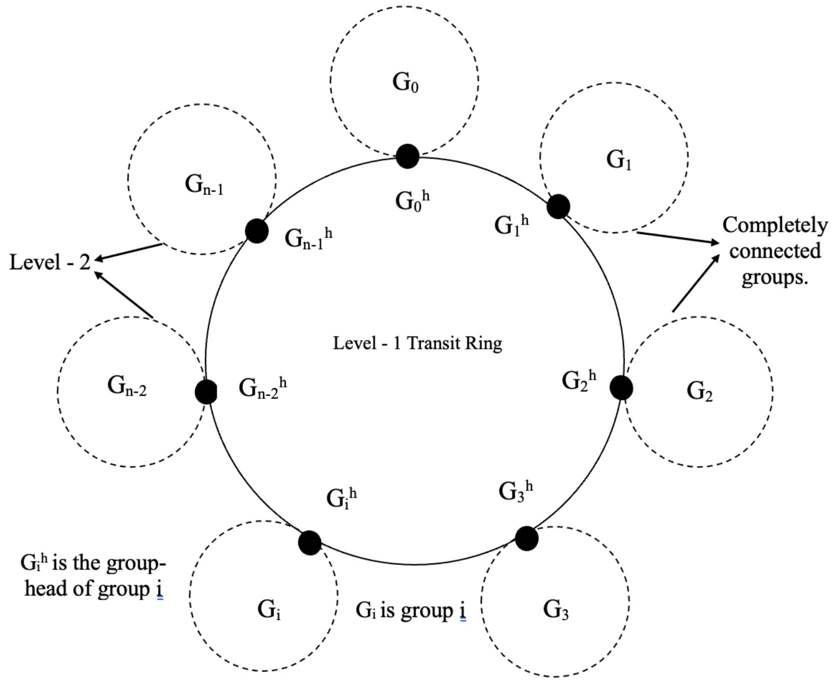


Figure 1: A two-level RC based structured P2P architecture with n distinct resource types

in it. List T now contains (k+1) different public keys corresponding to (k +1) group-heads. List T initially contains only PU_0 of G_0^h.

5. G_0^h performs E (T, PR_0k) and executes the broadcast protocol so that each node (group-head) on the transit-ring receives a copy of the encrypted IRT T; each receiving node performs D (PU_0, E (T, PR_0k)) to get the recent copy of the IRT T. Now each node has the knowledge of the public keys of all existing nodes.
6. The above five steps are executed each time a peer with a new resource type contacts G_0^h to join the network.

Situation 2: Resource type of peer p exists in IRT

The following procedures are carried out in order to add the new peer p to the network when it tries to join the system and has some instance(s) of an existing resource type with code, say, m.

1. New peer p contacts G_0^h for joining the network
2. G_0^h checks with the (m+1)th entry in IRT to get the IP address of G_m^h
3. G_0^h sends the IP address of G_m^h to p
4. Peer p contacts G_m^h
5. If peer p is the second peer to join a group of a particular resource code m, then G_m^h will assign peer p as a secondary group-head for resource code m.
   (a) G_m^h gives its public key PU_m to peer p and peer p gives it’s public key pub_i to G_m^h
   (b) G_m^h will update its IRT with peer p’s information < Peer p’s logical address, Peer p’s IP, Peer p’s public-key GS_m^h >and will broadcast it to all the other group-heads.
   (c) G_m^h update its local resource table (LRT) and broadcast it in the group
6. If peer p is not the second peer to join a group of a particular resource code m
   (a) G_m^h gives its public key PU_m to peer p and peer p gives it’s public key pub_i to G_m^h
   (b) G_m^h update its local resource table (LRT) and broadcast it in the group

3.2 Intra Group Lookup Algorithm in RC Based Architecture with Public Key Security

The resource lookup occurs within the group in this scenario, which means that the resource type is the same for both parties but the value is different. The algorithm for intra group data lookup is explained as follows.

Let us assume that in a group G_y, a peer $G_{y,i}$ with the resource $<Res_y, V_i>$is querying for a resource $<Res_y, V_j>Req_j$. The requesting peer will broadcast the request message Req_j in the group G_i using the LRT table as explained in Algorithm 1.

Observation 1: There will be two group-heads (G_i^h and GS_i^h) if there are more than two peers in a group (G_i) creating information redundancy for attack mitigation. All of the other group-heads and the group members have knowledge to this

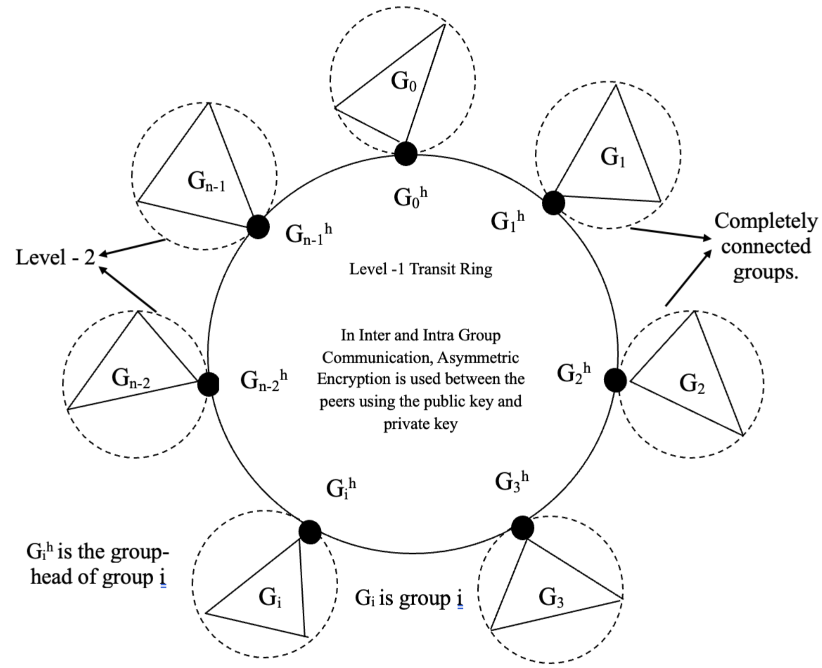


Figure 2: Inter and intra group communication using asymmetric encryption

Algorithm 1 Intra Group Lookup Algorithm in RC Based Architecture with Public Key Security

$G_{y,i}$ broadcast the request message Req_j in G_i using the IRT table
if $\ni G_{y,j} \in G_y$ with Req_j **then**
    $G_{y,j}$ encrypts Res_j: E(Res_j, $PU_{G_{y,i}}$) ▷ $G_{y,j}$ gets the $PU_{G_{y,i}}$ from IRT
    Unicast the encrypted message E(Res_j, $PU_{G_{y,i}}$)
    $G_{y,i}$ decrypts the message: D(E, PR_i)
else
    Search for $<Res_y, V_j>$fails
end if

information.

Observation 2: The maximum number of hops required to locate a resource using the Intra Group Lookup Algorithm in RC Based Architecture with Public Key Security is 2.

3.3 Inter Group Lookup Algorithm in RC Based Architecture with Public Key Security

Inter group communication takes place between nodes from two separate interest-based groups. Any communication between the peers $G_{x,i} \in$ group G_x and $G_{y,j} \in$ group G_y in our public-key based secure RC-based architecture occurs solely through the appropriate group heads G_x^h and G_y^h.

Let us assume that a peer $G_{x,i} \in$ group G_x with the resource $<Res_x, V_i>$ is querying for a resource $<Res_y, V_j>$.Both the peer $G_{x,i}$ and the group head G_x^h are aware about the fact that $Res_y \notin G_x$.

In this case 2 scenarios can happen:

1. The group-head $G_{y,j} \in$ group G_y has $<Res_y, V_j>$.
2. A peer $P_i \in$ group G_y has $<Res_y, V_j>$.

Case 1: The group-head $G_{y,j} \in$ group G_y has $<Res_y, V_j>$. The algorithm for the secured public-key based data lookup search for case 1 in RC Based Architecture is explained in Algorithm 2.

Observation 3: The maximum number of hops required to locate a resource using the Inter Group Lookup Algorithm in RC Based Architecture with Public Key Security for case 1 is n/2 + 7, where n is the total number of resource present in the network.

Case 2: The group-head $G_{y,j} \in$ group G_y has $<Res_y, V_j>$. The algorithm for the secured public-key based data lookup search for case 1 in RC Based Architecture is explained in Algorithm 3.

Observation 4: The maximum number of hops required to locate a resource using the Inter Group Lookup Algorithm in RC Based Architecture with Public Key Security for case 2 is n/2 + 11, where n is the total number of resource present in the network..

3.4 Threat model

We begin by describing the nature of attack on the peer-to-peer system along with the intruder objective and capabilities. This threat model emphasizes the vulnerability of the group heads and disruptions in inter-group communication due to insider

Algorithm 2 : Case 1: $G_{y,j} \in$ group G_y has $<Res_y, V_j>$

1: $G_{x,i}$ encrypts the request message E($<Res_y, V_j>$, $PU_{G_x^h}$) using the LRT table
2: G_x^h decrypts the request message D(E, $PR_{G_x^h}$)
3: G_x^h looks at its IRT and find $G_y^h : Res_y \in G_y^h$
4: G_x^h encrypts the request message E(Res_y, $PU_{G_y^h}$) using the IRT table and broadcast it in the transit ring
5: G_y^h decrypts the request message D(E, $PR_{G_y^h}$) ▷ $G_y^h \in$ group G_y holds the resource type of Res_y
6: **if** G_y^h itself have $<Res_y, V_j>$ **then**
7: G_y^h performs hash of its $PU_{G_y^h}$: H($PU_{G_y^h}$)
8: G_y^h unicast E($PU_{G_y^h}$ + H($PU_{G_y^h}$), $PU_{G_x^h}$) to G_x^h
9: G_x^h decrypts the message D(E, $PR_{G_x^h}$)
10: G_x^h encrypts with message E($PU_{G_{x,i}}$, $PU_{G_y^h}$ + H($PU_{G_y^h}$) and unicast it to $G_{x,i}$
11: $G_{x,i}$ decrypts the message with D(E, $PR_{G_{x,i}}$)
12: $G_{x,i}$ performs hash function on received public-key: H($PU_{G_y^h}$)
13: $G_{x,i}$ checks H($PU_{G_y^h}$) with the received hash value in step 11
14: **if** step 13 == true **then**
15: $G_{x,i}$ encrypts $PU_{G_{x,i}}$ with $PU_{G_y^h}$ and unicast it to G_x^h: E($PU_{G_{x,i}}$, $PU_{G_y^h}$)
16: G_x^h unicast it to G_y^h
17: G_y^h performs D(E, $PR_{G_y^h}$)
18: G_y^h perform E($<Res_y, V_j>$, $PU_{G_{x,i}}$) and unicast it to G_x^h
19: G_x^h forward the message to $G_{x,i}$
20: $G_{x,i}$ performs D(E, $PR_{G_{x,i}}$)
21: **else**
22: Algorithm 4
23: **end if**
24: **else**
25: Case 2
26: **end if**

or backdoor access attack. Due to the one-hop transfer to the members, the intra-group communication remains unaffected while the intruder takes covert control of the *inter-group communication in the transit ring network*.

The objective of the intruder is to cause Loss of Integrity by *compromising selective group heads that corrupts the message data*. This is an extremely realistic high-impact attack as the adversary acquires and maintains covert access on the query-and-reply process with remote administration tool or rootkit. We assume that the intruder has no control over the members. We can experiment/explore the attack impact by varying the number of compromised group heads and extent of data modifications. We want to highlight network operations and components that remain unaffected during the attack. It is assumed that the intruder does not make large data modifications in the Information Resource Table, because it can be easily verified among the neighboring group heads with fast look-ups.

In addition, the encryption at the group head and member levels should operate correctly which follows mathematically proven robust behavior of asymmetric encryption. As public key (redundancy) is maintained in all group heads and hence, such large-scale covert data modification is unlikely. In section 3.5 for both the cases mentioned in section 3.3 we present a solution to the above described difficulty of a malicious group-head by adding the notion of assigning a second group-head to each group.

If there are more than two peers in a group (G_i), there will be two group-heads (G_i^h and GS_i^h), generating information redundancy for attack mitigation. The other group-heads and members are all aware of this information. If step 13 of algorithm 2 or algorithm 3 returns false, it indicates that one or more malicious nodes (the group-heads) are present in the path between the destination node and the source node, because the hash value received from the destination node differs from the hash value obtained by the source after performing the hash on the destination node's public key. It's also possible that the destination node is malicious as well.

In this circumstance, we propose the algorithms for carrying on the communication with the second group-heads in section 3.5. We terminate communication if the aforementioned hash values are still out of sync. At this point, it is possible to recommend adding a third group head, a fourth group head, and so on. But, doing so would result in an increase in the size of the IRT table, which would slow down memory access as well

Algorithm 3 : Case 2: A peer $P_i \in$ group G_y has $<Res_y, V_j>$

1: $G_{x,i}$ encrypts the request message E($<Res_y, V_j>$, $PU_{G_x^h}$) using the LRT table
2: G_x^h decrypts the request message D(E, $PR_{G_x^h}$)
3: G_x^h looks at its IRT and find $G_y^h : Res_y \in G_y^h$
4: G_x^h encrypts the request message E(Res_y, $PU_{G_y^h}$) using the IRT table and broadcast it in the transit ring
5: G_y^h decrypts the request message D(E, $PR_{G_y^h}$) ▷ $G_y^h \in$ group G_y holds the resource type of Res_y
6: **if** G_y^h does not have $<Res_y, V_j>$ **then**
7: G_y^h broadcast $<Res_y, V_j>$in group G_y
8: **if** $P_i \in$ group G_y has $<Res_y, V_j>$ **then**
9: P_i performs hash of its PU_{P_i}: H(PU_{P_y})
10: P_i unicast E(PU_{P_i} + H(PU_{P_i}), $PU_{G_y^h}$) to G_y^h
11: G_y^h decrypts the message D(E, $PR_{G_y^h}$)
12: G_y^h encrypts: E(PU_{P_i} + H(PU_{P_i}), $PU_{G_x^h}$)
13: G_x^h decrypts the message D(E, $PR_{G_x^h}$)
14: G_x^h encrypts with message E($PU_{G_{x,i}}$, $PU_{G_y^h}$ + H($PU_{G_y^h}$) and unicast it to $G_{x,i}$
15: $G_{x,i}$ decrypts the message with D(E, $PR_{G_{x,i}}$)
16: $G_{x,i}$ performs hash function on received public-key: H($PU_{G_y^h}$)
17: $G_{x,i}$ checks H($PU_{G_y^h}$) with the received hash value in step 11
18: **if** step 13 == true **then**
19: $G_{x,i}$ encrypts $PU_{G_{x,i}}$ with $PU_{G_{y,i}}$ and unicast it to G_x^h: E($PU_{G_{x,i}}$, $PU_{G_{y,i}}$
20: G_x^h unicast it to G_y^h
21: G_y^h uincast to $G_{y,i}$
22: $G_{y,i}$ performs D(E, $PR_{G_{y,i}}$)
23: $G_{y,i}$ perform E($<Res_y, V_j>$, $PU_{G_{x,i}}$) and unicast it to G_y^h
24: G_y^h unicast message to G_x^h
25: G_x^h forward the message to $G_{x,i}$
26: $G_{x,i}$ performs D(E, $PR_{G_{x,i}}$)
27: **else**
28: Algorithm 5
29: **end if**
30: **end if**
31: **else**
32: Search for $<Res_y, V_j>$has failed
33: **end if**

as data-lookup episodes. The number of stand-by group leaders needed will therefore largely depend on the network designer, who will make this decision after considering the scenario.

4 Performance Evaluation

P2P networks offer the potential to improve network capabilities through the sharing of music, video, and other services. Nonetheless, P2P networks pose security issues since the nodes are exposed to a variety of security attacks. One of them is the man-in-the-middle attack. It is an indirect intrusion attempt in which the attacker places his computing device between two nodes. As a result, the intermediary node can intercept and change communications sent between two valid users without the knowledge of the sender and recipient. Encryption technologies can be used for data transmission as a defense measure.

The algorithms we proposed for data-lookup in section 3 are immune to such attacks because every message communicated is encrypted with the public key of the immediate destination peer, and the message can be decrypted by only the destination peer with its own private key. This prohibits anybody along the path of the message from interfering with the communication between the source and the destination.

To understand trade-off analysis between efficiency vs effectiveness of security measures, we must estimate the tri-factor performance cost related to computation, memory usage and network utilization. Cryptographic operations for signature generation and verification is inexpensive as key size is much smaller than the actual data. Our protocol generates the

Algorithm 4 : Case 1: $G_{y,j} \in$ group G_y has $<Res_y, V_j>$

1: $G_{x,i}$ encrypts the request message E($<Res_y, V_j>$, $PU_{GS^h_x}$) using the LRT table
2: GS^h_x decrypts the request message D(E, $PR_{GS^h_x}$)
3: GS^h_x looks at its IRT and find GS^h_y : $Res_y \in GS^h_y$
4: GS^h_x encrypts the request message E(Res_y, $PU_{GS^h_y}$) using the IRT table and broadcast it in the transit ring
5: GS^h_y decrypts the request message D(E, $PR_{GS^h_y}$) ▷ $GS^h_y \in$ group G_y holds the resource type of Res_y
6: **if** GS^h_y itself have $<Res_y, V_j>$ **then**
7: GS^h_y performs hash of its $PU_{GS^h_y}$: H($PU_{GS^h_y}$)
8: GS^h_y unicast E($PU_{GS^h_y}$ + H($PU_{GS^h_y}$), $PU_{GS^h_x}$) to GS^h_x
9: GS^h_x decrypts the message D(E, $PR_{GS^h_x}$)
10: GS^h_x encrypts with message E($PU_{G_{x,i}}$, $PU_{GS^h_y}$ + H($PU_{GS^h_y}$) and unicast it to $G_{x,i}$
11: $G_{x,i}$ decrypts the message with D(E, $PR_{G_{x,i}}$)
12: $G_{x,i}$ performs hash function on received public-key: H($PU_{GS^h_y}$)
13: $G_{x,i}$ checks H($PU_{GS^h_y}$) with the received hash value in step 11
14: **if** step 13 == true **then**
15: $G_{x,i}$ encrypts $PU_{G_{x,i}}$ with $PU_{GS^h_y}$ and unicast it to GS^h_x: E($PU_{G_{x,i}}$, $PU_{GS^h_y}$)
16: GS^h_x unicast it to GS^h_y
17: GS^h_y performs D(E, $PR_{GS^h_y}$)
18: GS^h_y perform E($<Res_y, V_j>$, $PU_{G_{x,i}}$) and unicast it to GS^h_x
19: GS^h_x forward the message to $G_{x,i}$
20: $G_{x,i}$ performs D(E, $PR_{G_{x,i}}$)
21: **else**
22: Algorithm 3
23: **end if**
24: **else**
25: Case 2
26: **end if**

signatures only once when we are transferring the public key. Similarly, verification also happens at the member level which democratizes the detection and further defense tasks.

A more robust solution of multi-member verification within the intra-group may incur additional processing costs leading to unnecessary detection latency. Fast comparison between calculated vs received signature at each level reduces detection delays and allows for easier mitigation and recovery. Such corrective functionality needs to be implemented at the member-level as the extent of corruption at the group heads remains unclear during the attack. Detailed exploration of the mitigation framework demands in-depth exploration and is skipped for brevity in this work.

In our signature-based authentication at the member-level, the signature values are immediately shared with the member nodes eliminating the storage needs. There is no overhead of control messages for maintaining network security since when a peer joins the network, they update the IRT with their ip_address, resource code, and public key information, which is then broadcast across the network. As a result, following the broadcast, everyone in the network will have access to the most up-to-date information. By the same reasoning, bandwidth utilization remains unaffected as no additional message bits are exchanged during the regular broadcasting

A potential vulnerability in the proposed network is the disruption of the group head, which can result in a single point of failure during a man-in-the-middle attack. The group head plays a crucial role in coordinating and facilitating communication within a specific group or subgroup of peers. If the group head is disrupted or compromised, it can severely impact inter-group communication and hinder the overall functionality of the network. As an alternative solution, redundancy measures such as multiple group heads need incur additional hardware and communication cost.

As soon as a peer enters the network, they update the IRT with their IP address, resource code, and public key information, which is then broadcast throughout the network, there is no need for control messages to ensure network security. As a result, everyone on the network will have access to the most recent information after the broadcast. The data-lookup complexity of the proposed public-key and digital signature based secured data-lookup protocols is O(n), n $<<$N, where n is the total number of resources present in the network and N is the total

Algorithm 5 : Case 2: A peer $P_i \in$ group G_y has $<Res_y, V_j>$

1: $G_{x,i}$ encrypts the request message E($<Res_y, V_j>$, $PU_{GS_x^h}$) using the LRT table
2: GS_x^h decrypts the request message D(E, $PR_{GS_x^h}$)
3: GS_x^h looks at its IRT and find $GS_y^h : Res_y \in GS_y^h$
4: GS_x^h encrypts the request message E(Res_y, $PU_{GS_y^h}$) using the IRT table and broadcast it in the transit ring
5: GS_y^h decrypts the request message D(E, $PR_{GS_y^h}$) ▷ $GS_y^h \in$ group G_y holds the resource type of Res_y
6: **if** GS_y^h does not have $<Res_y, V_j>$ **then**
7: GS_y^h broadcast $<Res_y, V_j>$in group G_y
8: **if** $P_i \in$ group G_y has $<Res_y, V_j>$ **then**
9: P_i performs hash of its PU_{P_i}: H(PU_{P_y})
10: P_i unicast E(PU_{P_i} + H(PU_{P_i}), $PU_{GS_y^h}$) to GS_y^h
11: GS_y^h decrypts the message D(E, $PR_{GS_y^h}$)
12: GS_y^h encrypts: E(PU_{P_i} + H(PU_{P_i}), $PU_{GS_x^h}$)
13: GS_x^h decrypts the message D(E, $PR_{GS_x^h}$)
14: GS_x^h encrypts with message E($PU_{G_{x,i}}$, $PU_{GS_y^h}$ + H($PU_{GS_y^h}$) and unicast it to $G_{x,i}$
15: $G_{x,i}$ decrypts the message with D(E, $PR_{G_{x,i}}$)
16: $G_{x,i}$ performs hash function on received public-key: H($PU_{GS_y^h}$)
17: $G_{x,i}$ checks H($PU_{GS_y^h}$) with the received hash value in step 11
18: **if** step 13 == true **then**
19: $G_{x,i}$ encrypts $PU_{G_{x,i}}$ with $PU_{G_{y,i}}$ and unicast it to GS_x^h: E($PU_{G_{x,i}}$, $PU_{G_{y,i}}$
20: GS_x^h unicast it to GS_y^h
21: GS_y^h uincast to $G_{y,i}$
22: $G_{y,i}$ performs D(E, $PR_{G_{y,i}}$)
23: $G_{y,i}$ perform E($<Res_y, V_j>$, $PU_{G_{x,i}}$) and unicast it to GS_y^h
24: GS_y^h unicast message to GS_x^h
25: GS_x^h forward the message to $G_{x,i}$
26: $G_{x,i}$ performs D(E, $PR_{G_{x,i}}$)
27: **else**
28: Algorithm 4
29: **end if**
30: **end if**
31: **else**
32: Search for $<Res_y, V_j>$has failed
33: **end if**

number of peers present in the network. The complexity of the data lookup is not dependent on the total number of peers in the network. This improves the robustness and scalability of our network.

5 Conclusion

In this work, a 2-level non-DHT-based P2P architecture was considered. The choice to utilize an interest-based architecture was made because:

1. We have previously shown [7] that the data lookup strategies outperform many really well-known DHT-based contributions [15, 18, 23] in terms of search latency.
2. It is beneficial over several existing interest-based systems [1, 3, 5, 9, 10, 19]. Public-key cryptography and hashing has been used to successfully integrate security into the intra-group and inter-group communication techniques described in [7] in this work.

To guarantee the security of the architecture, the writers of [8] combined the symmetric key and public key cryptography approaches using a well-defined methodology. Secured key delivery between sender and receiver is the primary drawback of symmetric key encryption. Using a single shared key exposes the network vulnerability on both sides due to attack leading to loss of integrity . In [16] the authors used asymmetric key cryptography to address the aforementioned problems and offered strong public-key cryptography methodologies for the security of contemporary communication protocols in [8] with correspondingly reduced costs.

In this paper, we have presented a public key cryptography and digital signature authentication based data lookup protocol that are efficient and secured in RC-based hierarchical structured P2P network. Adversary takes covert control of the inter-group communication in the transit ring network as part of the threat model. Our scheme compares digital signatures generated with assymetric key for authenticating the group heads. We are assuming that the peers are trusted and functioning correctly as in the earlier article [16]. The proposed public-key and digital signature based secured data-lookup protocols have a data-lookup complexity of O(n), n <<N, where n is the total number of resources in the network and N is the total number of peers in the network. The overall number of peers in the network has no impact on how costly the data lookup is. As a result, our network is more reliable and scalable. Using the algorithms presented in this research, we can detect presence of malicious nodes in the network but we cannot identify them. As a follow-up research, we are thinking of creating a trust algorithm to be used amongst group-heads in order to rule out malicious group-heads.

References

[1] Lyes Badis, Mourad Amad, Djamil Aïssani, Kahina Bedjguelal, and Aldja Benkerrou. ”ROUTIL: P2P Routing Protocol Based on Interest Links”. In *2016 International Conference on Advanced Aspects of Software Engineering (ICAASE), IEEE, pp. 1 - 5*, 2016.

[2] Yatin Chawathe, Sylvia Ratnasamy, Lee Breslau, Nick Lanham, and Scott Shenker. ”Making Gnutella-like P2P Systems Scalable”. In *Proceedings of the 2003 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, SIGCOMM ’03, New York, NY, USA. Association for Computing Machinery, pp. 407-418, 2003.

[3] Wen-Tsuen Chen, Chi-Hong Chao, and Jeng-Long Chiang. ”An Interested-based Architecture for Peer-to-Peer Network Systems”. In *20th International Conference on Advanced Information Networking and Applications-Volume 1 (AINA’06), IEEE*, 1:707-712, 2006.

[4] Prasanna Ganesan, Qixiang Sun, and Hector Garcia-Molina. ”Yappers: A Peer-to-Peer Lookup Service Over Arbitrary Topology”. In *IEEE INFOCOM 2003. Twenty-second Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE Cat. No. 03CH37428)*, 2: 1250–1260, 2003.

[5] Mo Hai and Yan Tu. ”A P2P E-Commerce Model Based on Interest Community”. In *2010 International Conference on Management of e-Commerce and e-Government, IEEE, pp. 362-365*, 2010.

[6] Swathi Kaluvakuri, Bidyut Gupta, Banafsheh Rekabdar, Koushik Maddali, and Narayan Debnath. ”Design of RC-Based Low Diameter Two-Level Hierarchical Structured P2P Network Architecture”. In Mohammed Serrhini, Carla Silva, and Sultan Aljahdali, editors, *Innovation in Information Systems and Technologies to Support Learning Research, pp. 312-320*, Springer International Publishing, pp. 312-320, 2020.

[7] Swathi Kaluvakuri, Koushik Maddali, Nick Rahimi, Bidyut Gupta, and Narayan Debnath. ”Generalization of RC-Based Low Diameter Hierarchical Structured P2P Network Architecture”. *International Journal of Computer and Their Applications pp. 74*, 2020.

[8] Swathi Kaluvakuri, Indranil Roy, Koushik Maddali, Bidyut Gupta, and Narayan Debnath. ”Efficient Secured Data Lookup and Multicast Protocols with Anonymity in RC-Based Two-level Hierarchical Structured P2P Network.”. *International Journal for Computers & Their Applications*, 28(3):140-149, 2021.

[9] Mujtaba Khambatti, Kyung Dong Ryu, and Partha Dasgupta. ”Structuring Peer-to-Peer Networks Using Interest-based Communities”. In *International Workshop On Databases, Information Systems, and Peer-to-Peer Computing, Springer, pp. 48–63*, 2003.

[10] Sardar Kashif Ashraf Khan and Laurissa N Tokarchuk. ”Interest-based Self Organization in Group-Structured P2P Networks”. In *2009 6th IEEE Consumer Communications and Networking Conference, pp. 1-5*, 2009.

[11] Dmitry Korzun and Andrei Gurtov. ”Hierarchical Architectures in Structured Peer-to-Peer Overlay Networks”. *Peer-to-Peer Networking and Applications, Springer*, 7(4):359-395, 2014.

[12] Koushik Maddali, Swathi Kaluvakuri, Nick Rahimi, Bidyut Gupta, and Narayan Debnath. ”On Designing Secured Communication Protocols along with Anonymity for CRT based Structured P2P Network Architecture”. *EPiC Series in Computing*, 75:59-68, 2021.

[13] Koushik Maddali, Banafsheh Rekabdar, Swathi Kaluvakuri, and Bidyut Gupta. ”Efficient Capacity-Constrained Multicast in RC-Based P2P Networks”. In *Proceedings of 32nd International Conference on*, 63:121-129, 2019.

[14] Zhuo Peng, Zhenhua Duan, Jian-Jun Qi, Yang Cao, and Ertao Lv. ”HP2P: A Hybrid Hierarchical P2P Network”. In *First International Conference on the Digital Society (ICDS’07), IEEE, pp. 18*, 2007.

[15] Antony Rowstron and Peter Druschel. ”Pastry: Scalable, Distributed Object Location and Routing for Large-scale Peer-to-Peer Systems”. *Distributed Systems Platforms (Middleware),In IFIP/ACM International Conference, pp. 329-350*, 2001.

[16] Indranil Roy, Nick Rahimi, Reshmi Mitra, and Swathi Kaluvakuri. ”Efficient Secured Public-Key based Data Lookup and Multicast Protocols with Anonymity in RC-Based Two-level Hierarchical Structured P2P Network”. In *Proceedings of 35th International Conference on Computer Applications in Industry and Engineering*, EPiC Series in Computing, EasyChair, 89:62-71, 2022.

[17] Kai Shuang, Peng Zhang, and Sen Su. ”Comb: A Resilient and Efficient Two-hop Lookup Service for Distributed Communication System”. *Security and Communication Networks*, 89:62-71, 2022.

[18] Ion Stoica, Robert Morris, David Liben-Nowell, David R Karger, M Frans Kaashoek, Frank Dabek, and Hari Balakrishnan. ”Chord: A Scalable Peer-to-Peer Lookup Protocol for Internet Applications”. *IEEE/ACM Transactions on networking, pp. 17–32*, 11(1):17-31, 2003.

[19] Zhiyong Tu, Wei Jiang, and Jinyuan Jia. ”Hierarchical Hybrid DVE-P2P Networking Based on Interests Clustering”. In *2017 International Conference on Virtual Reality and Visualization (ICVRV), IEEE, pp. 378–381*, 2017.

[20] Ming Xu, Shuigeng Zhou, and Jihong Guan. ”A New and Effective Hierarchical Overlay Structure for Peer-to-Peer Networks”. *Computer Communications, Elsevier*, 34(7):862-874, 2011.

[21] Min Yang and Yuanyuan Yang. ”An Efficient Hybrid Peer-to-Peer System for Distributed Data Sharing”. *IEEE Transactions on computers, IEEE*, 59(9):1158-1171, 2009.

[22] Rongmei Zhang and Y Charlie Hu. ”Assisted Peer-to-Peer Search With Partial Indexing”. *IEEE Transactions on Parallel and Distributed Systems, IEEE*, 18(8):1146-1158, 2007.

[23] Ben Y Zhao, Ling Huang, Jeremy Stribling, Sean C Rhea, Anthony D Joseph, and John D Kubiatowicz. ”Tapestry: A resilient global-scale overlay for service deployment”. *IEEE Journal on selected areas in communications, IEEE*, 22(1):41-53. 2004.

Anjila Neupane (photo not available) is pursuing her masters in Applied Computer Science from Southeast Missouri State University. She have completed her Bachelors Degree in Computer Engineering from Tribhuwan University in Kathmandu, Nepal. Her primary research interests revolve around peer-to-peer (P2P) networks and their applications. She is particularly interested in exploring the potential of P2P systems to improve communication and information sharing among individuals and organizations.

Indranil Roy (photo not available) is an Assistant Professor in the Department of Computer Science at the Southeast Missouri State University. He received his MS and Ph.D. degrees in Computer Science from Southern Illinois University, Carbondale in 2018 and 2022, respectively. His current research interest includes the design of architecture and communication protocols for structured peer-to-peer overlay networks, security in overlay networks, and Blockchain.

Reshmi Mitra (photo not available) is an Assistant Professor in the Department of Computer Science at the Southeast Missouri State University. She received her MS and Ph.D. degrees in Electrical and Computer Engineering from the University of North Carolina at Charlotte in 2007 and 2015, respectively. Previously she has worked at the National Institute of Technology India, Advanced Micro Devices Austin, and Samsung Austin R&D Center. Her research interests include Security and Performance issues in IoT, Cloud Computing, and Blockchain.

Bidyut Gupta (photo not available) received his M. Tech. degree in Electronics Engineering and Ph.D. degree in Computer Science from Calcutta University, Calcutta, India. At present, he is a professor at the School of Computing (formerly Computer Science Department), Southern Illinois University, Carbondale, Illinois, USA. His current research interest includes design of architecture and communication protocols for structured peer-to-peer overlay networks, security in overlay networks, and block chain. He is a senior member of IEEE and ISCA.

Narayan Debnath (photo not available) earned a Doctor of Science (D.Sc.) degree in Computer Science and also a Doctor of Philosophy (Ph.D.) degree in Physics. Narayan C. Debnath is currently the Founding Dean of the School of Computing and Information Technology at Eastern International University, Vietnam. He is also serving as the Head of the Department of Software Engineering at Eastern International University, Vietnam. Dr. Debnath has been the Director of the International Society for Computers and their Applications (ISCA) since 2014. Formerly, Dr. Debnath served as a Full Professor of Computer Science at Winona State University, Minnesota, USA for 28 years (1989-2017). Dr. Debnath has been an active member of the ACM, IEEE Computer Society, Arab Computer Society, and a senior member of the ISCA.

Toward Automated Feature Model Generation from UML Use Case and Class Diagrams

Tasneem Yousef* and Said Ghoul*
Philadelphia University, Amman, Jordan

Abstract

The Feature Model is one of the most important concepts in the Software Product Line development process, as it helps to represent commonalities and variabilities between software products. Several research works are currently conducted to generate feature model of a software product line from its requirements document. However, the generated feature models do not include data and actor aspects which are essential parts of any software product line. The work presented in this paper proposes an approach handling this insufficiency. This approach defines a methodology toward automated feature model generation from requirement documents limited to Unified Modelling Language use case and class diagrams. The use case is specified with a use case description template enhanced with information required for feature model. The class diagram is specified in a textual form obtained by translating the diagram into java programming language. The target generated feature model is enhanced by the introduction of data and actor concepts. An evaluating real case study (quality assurance at Philadelphia University), was used to evaluate the feasibility of the proposed approach and obtained results show its practical benefits.

Key Words: Feature model (FM), software production line (SPL), unified modelling language (UML) use case, UML class diagram, feature model generation approach.

1 Introduction

Building software from scratch is complex and expensive, and there is a high probability of mistakes. But, the concept of SPL helps to analyse a set of software that share some parts and are designed for a specific domain [7]. This approach allowed reuse of commonalities and variabilities that resulted. There is a core asset that represents the commonalities and variabilities, including the variability model such as the FM [1, 4-5]. It helped to represent the variable and common characteristics of software, to make decisions, and to avoid mistakes [2]. A FM is one of the most efficient ways to represent and manage the variance obtained from products during the SPL process. It is a visual representation of features in a hierarchy, called a feature tree, at the top of what are the main features and at the bottom are the features that branch off from the main features and the relationships between them [3].

Researchers have introduced many approaches for generating FMs. These approaches define input requirements document, generation processes, and output generated FM. Researchers have used many forms of input documents. Some of them used features set & UML-class diagram as input as in ModelVars2SPL approach [2]. Others used Hasse diagram as inputs in Equivalence Class Feature Diagram approach [3]. The feature models obtained through these approaches may not contain all the relationships and features covering all the parts of an SPL process [2, 3, 4, 6, 10, 13]. In fact, the researchers have developed automated generation methodologies, but the FM generated through these methodologies does not include all the features and relationships [2], particularly data and actor features which are important in any SPL.

To solve this insufficiency in the FM generated by current methodologies, this paper proposes an approach toward automated FM generation from the requirements specification documents: A UML [9 use case description (by semi-formal language) and a UML class diagram translated into Java language. The final FM contains traditional features, additional features (actor and data), relationships between them, and constraints.

An evaluating real case study (quality assurance at Philadelphia University), was used to evaluate the feasibility of the proposed approach, and obtained results showed its practical benefits.

2 Background

2.1 Feature Models

FMs are language, used for system visual description, which serves to determine the scope of the product line by the features of its products. It represents the features that the product may or may not have, by defining the features, relationships, and constraints between them. A FM is hierarchy called feature tree that contains the main features at the top of the tree and the features extending from them. Features can be mandatory or optional. They are linked together with relationships [3]: *OR* relation (the main feature consists of one or more features extended from it), *XOR* relation (the main feature consists of only one of the features extended from it), *Exclude* relation (The two features that found this relationship between them, cannot

*Research Laboratory on Bioinspired Software Engineering. Email: tasneem.fy18@gmail.com, shgoul@philadelphia.edu.jo.

be in the same product), *Imply* relation (the selected feature requires another feature, to be configured properly). Some researchers have represented some relationships (as exclude and imply) outside the feature model. It was represented by textual relationships called constraints, to reduce the complexity of getting a network of relationships, difficult to understand and maintain in FM [11].

2.2 Software Requirements Document

*Software r*equirements document includes its elicited functional and non-functional requirements, their evaluation, formalization, and quality [12]. The elicitation allows gathering all requirements. The evaluation allows selecting the alternative requirements according to suitable criteria (priority, feasibility, security, quality, cost, etc.). The formalization allows specifying selected requirement using formal languages (Algebraic, Z, SDL, Petri-nets, etc.) in order to facilitate their automated translation into design and code. The quality allows evaluation of the obtained software requirement according to some quality norms and indicators [4].

3 An Application Case

The Quality Assurance (QA) agenda system at Philadelphia University is used as an application case to illustrate and evaluate the feasibility of the concepts presented in this paper. This system manages 16 academic weeks, but in this study the focus is on the seventh and twelfth weeks because they are the most important and significant. The QA agenda requirements are used as inputs to the proposed approach. These inputs are specified with two UML models: use case (Figure 1a) Using these inputs, an enhanced FM [14] is generated (Figure 2). Traditional aspect of that FM includes four main features student answers, student marks, control, and store checklist. All of them are mandatory features in the system. It also contains description and Class diagram (Figure 1b) translated into java. an AND relationship that extends from some of the main features, and a group of sub-features such as the "student answers" feature, that extends from it (discuss the exam with student, store a soft copy of the exam sheets, store marking schema). The new specification of that FM includes data and actor features, as well as inherent constraints.

4 Feature Model Generation Approach

4.1 Enhanced FM with Additional Features (Data and Actor)

As consequence to current relevant research works analysis [14], some enhancements to traditional FM are proposed (Figure 3). They are related to actors, data, and constraints over them.

4.2 Generation Process

The generation approach (Figure 4), is composed of three main components: the inputs that are UML models, the generation process, which goes through a set of steps for extracting the target FM model from requirements, and finally ends with the output, which is the generated FM.

4.3 The Req-To-FM Process

The *Req-to-FM* process (Figures 5a and 5b) is built on the basis of extraction from UML models (use case description and class diagram translated in java) of the target FM. It starts with two parallel processes building partial FMs through two phase extraction and building:

- *UC-to-FM*: It extracts traditional features, actors, constraints, and inherent relationships from use case description and builds up a UC-FM.
- *CD-to-FML:* It extracts data features, inherent relationships, and constraints from class diagram in java form and builds up a CD-FM

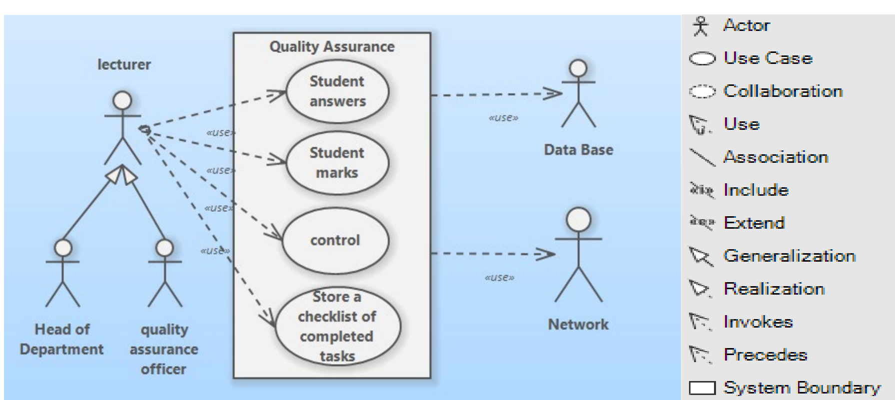


Figure 1a: QA agenda UML Use Case Diagram. This figure depicts the following:
- QA agenda actors with their relationships: QA officer, lecturer, head of department, database, and network.
- Three use cases classes: Students answers, students' marks, control and store a checklist of completed tasks

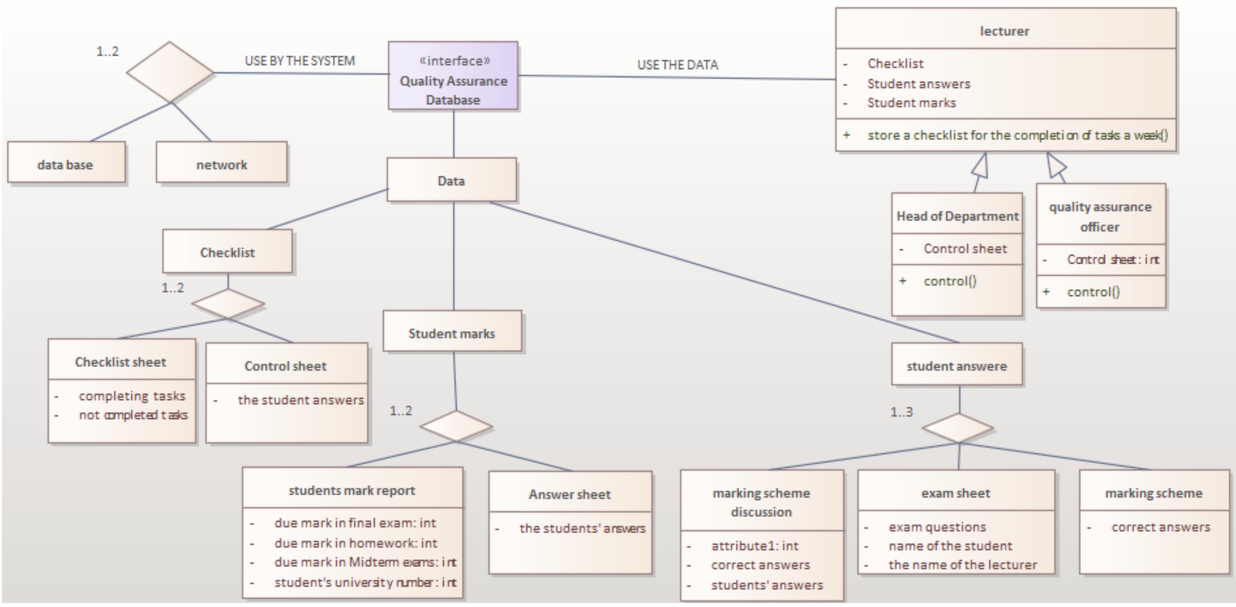


Figure 1b: QA agenda UML Use Case Diagram. This figure depicts the following:
- Actor Classes: QA officer, lecturer, head of department, database, and network.
- Data Classes: Exam sheet, answer sheet, marking scheme, marking scheme discussion, and student marks report

The process Merge-FMs ends Req-to-FM by merging the UC-FM with the CD-FM and producing the final target FM.

4.4 The UC-To-FM- Processes

This process extracts information from a UML use case description (in a logic language) and build up a UC- (FM. The extraction is carried out through four parallel processes: Features Extraction, Relation and/or/xor) Extraction, Actor Boundary Extraction, and Constraint Extraction (Figures 6a and 6b).

4.5 The CD-To-FM- Processes

This process extracts information from a UML class diagram (converted into Java language) and build up a CD-FM. The extraction is carried out through three parallel processes: Data Features Extraction, Relation and/or/xor) Extraction, and Constraints Extraction (Figures 7a and 7b).

5 Case Study

This section presents the feasibility study for the illustrative example (discussed in Sections 3) by the application of the UC-to-FM and CD-to-FM on a concrete business domain, which is the QA assurance agenda at Philadelphia University.

The Figure 8a shows partial application of the process UC-to-FM extraction phase. The outputs are the traditional part of the FM: a set of main features (student answers, student marks, control, store checklist for the completion of tasks a week), all of them are mandatory, and a set of sub-features (discuss the exam with student, store a soft copy of the exam sheets, …), and a set of relationships over them represented by the AND relationship.

The Figure 8b shows partial application of the process CD-to-FM extraction phase. The outputs are: a set of features that are divided into data and actor, and all the main features are mandatory, and a set of relations between them represented by the relationships (and, or, is-a).

After completing the components extraction step, the process moves to the rest of the process steps (depicted in Figures 5a and 5b). It ends with the FM depicted in Figure 2.

6 Comparison with Similar Works and Evaluation

Based on the current research in the field of FM generation, the following criteria are selected to evaluate the proposed approach and compare it with modern relevant studies.

(1) Improving the FM concepts to cover major aspects of SPL (data, actors, …).
(2) Using UML like -Diagrams (UC, CD, ...) as inputs to the generation approach.
(3) Formal definition of the FM generation approach.
(4) Completely automated generation approach.

In the study [2], the authors used a method to generate FMs called ModelVars 2SPL, it generates FMs from UML models

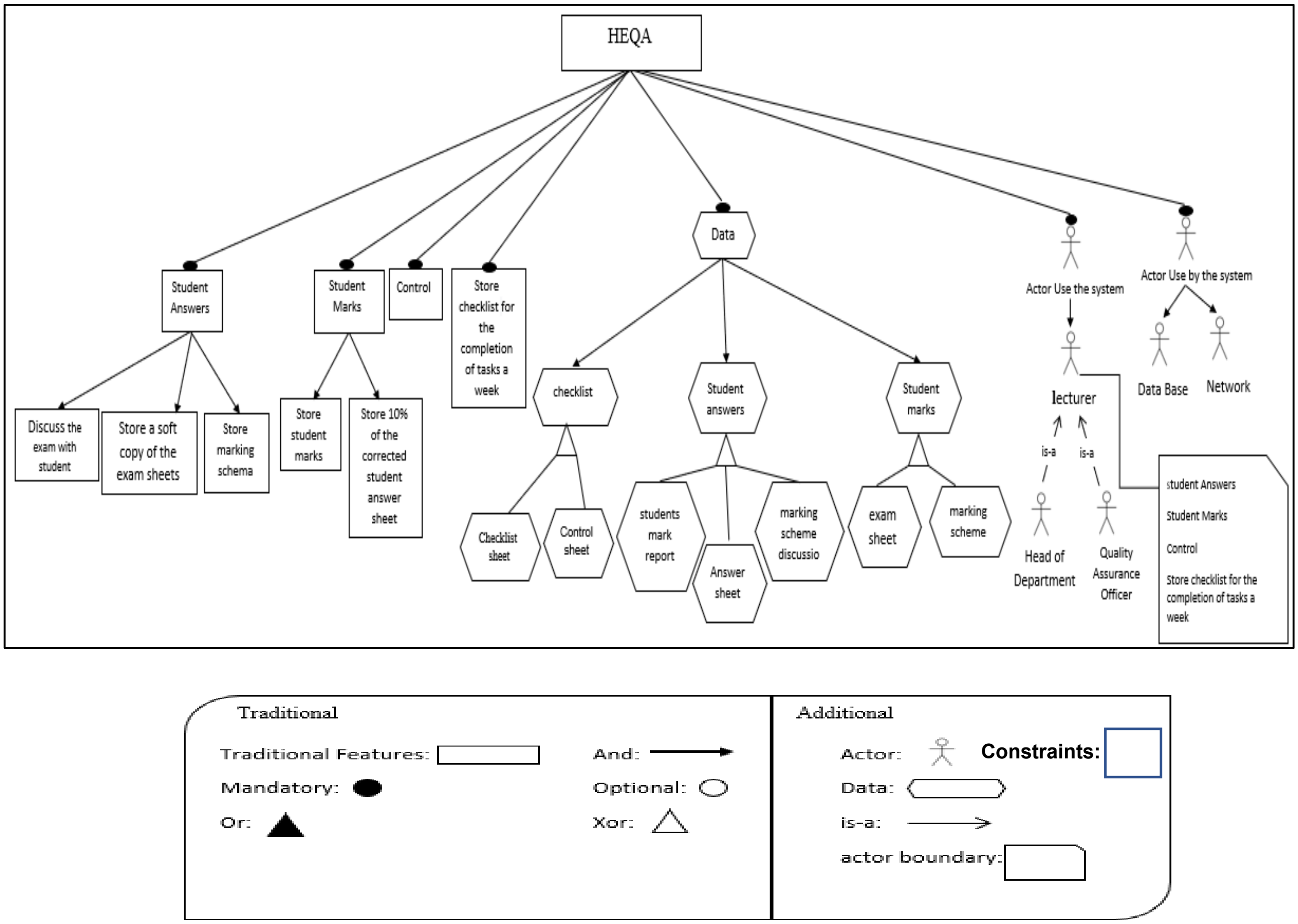


Figure 2: Generated FM from use case (Figure 1a) and class diagram (Figure 1b) for QA agenda at Philadelphia University. This FM was enhanced with additional features (data and actor), their relationships, and their inherent constraints

Constraints

Function-To-data constraints:
Discuss the exam with student **Uses** Exam Sheet ^ Marking Scheme

Data-To-data constraints:
marking scheme discussion **imply** Answer sheet ^ marking scheme

Actor-To-data constraints:
Lecturer **Has** Student answers ^ Student marks ^ Checklist

Figure 3: Generated FM uses traditional and additional notations. Additional notations specify actors, data, and constraints over them [14]. These constraints concern relationship between (function, data), (function, actor), (data, data), and (actor, data)

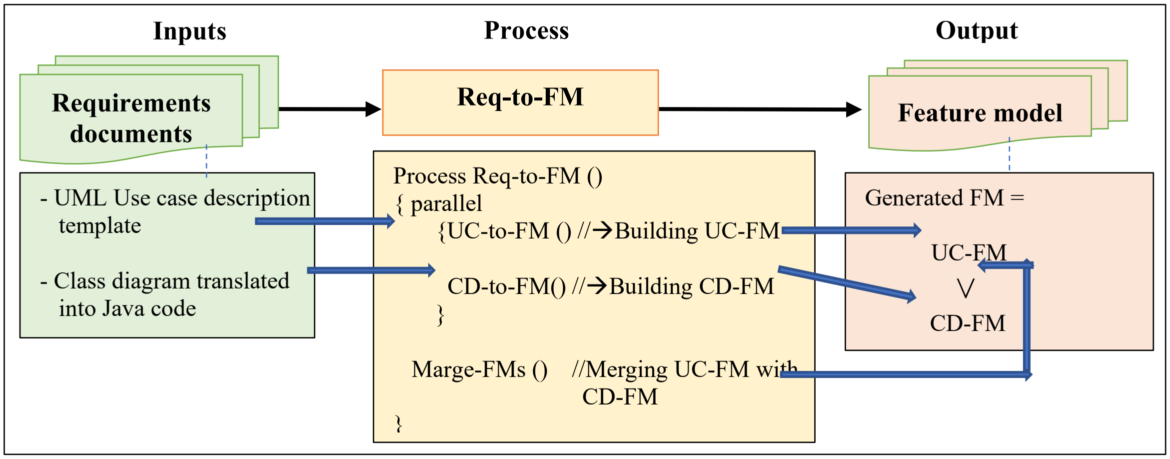


Figure 4: FM generation approach: The Req-to-FM process generates FM from UML Use and Class diagrams.

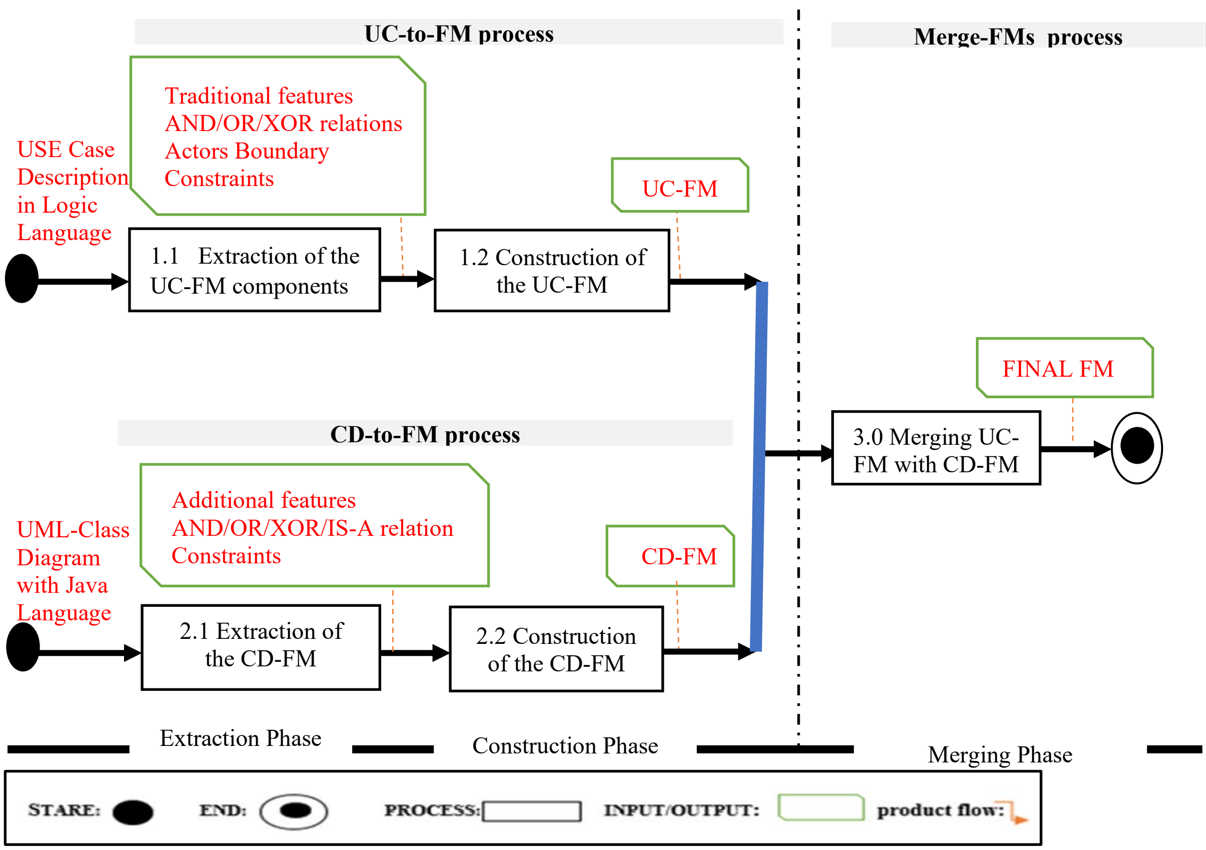


Figure 5a: The Req-to-FM process using UML notations

```
Include UC-to-FM (in UML Ucd, out FM Fm CD-to-FM (in Java Cdjava, out FM
Fmcd);
---------------------------------------------------------------------------
Process Req-to-FM (in UML UC-Description, CD-Java; out FM TargFM)
Begin
   Parallel:
    {
  UC-to-FM (UC-Description, UC-FM);
       CD-to-FM (CD-Java, CD-FM);
    }
  TargFM ← Merge-FMs (UC-FM, CD-FM)
End Req-to-FM
```

Figure 5b: The Req-to-FM process using algorithmic notations

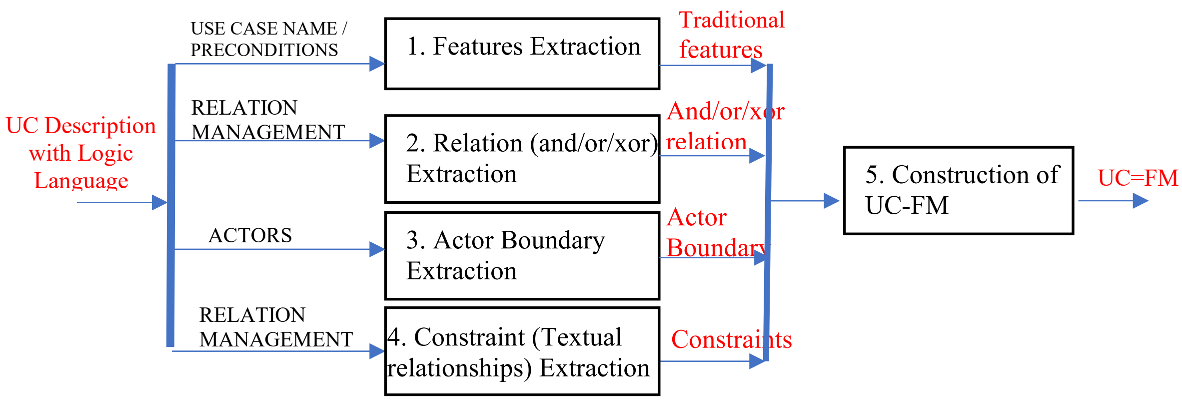


Figure 6a: The UC-to-FM process using UML notations

```
Include Features-Extraction (in UC UC-Description, out Feature Traditional-Features)
Relation-and/or/xor-Extraction (in UC UC-Description, out Relation Relations)    Actor-
Boundary-Extraction (in UC UC-Description, out Actor Actor-Boundary)      Constraints-
Extraction (in UC UC-Description, out Constraint Constraints)
---------------------------------------------------------------------------
Process UC-to-FM (in UML UC-Description, out FM UC-FM);
  Feature Traditional-Feature; Relation Relations, Actor Actor-Boundary;
  Constraint Constraints
Begin
   Parallel:
    {
      Features-Extraction (UC-Description, Traditional-Features)
      Relation-and/or/xor-Extraction (UC-Description, Relations)
      Actor-Boundary-Extraction (UC-Description, Actor-Boundary)
      Constraints-Extraction (UC-Description, Constraints)
    }
UC-FM ← Construct-FM (Traditional-Feature; Relations, Actor-Boundary, Constraints)
End UC-to-FM
```

Figure 6b: The UC-to-FM process using algorithmic notations

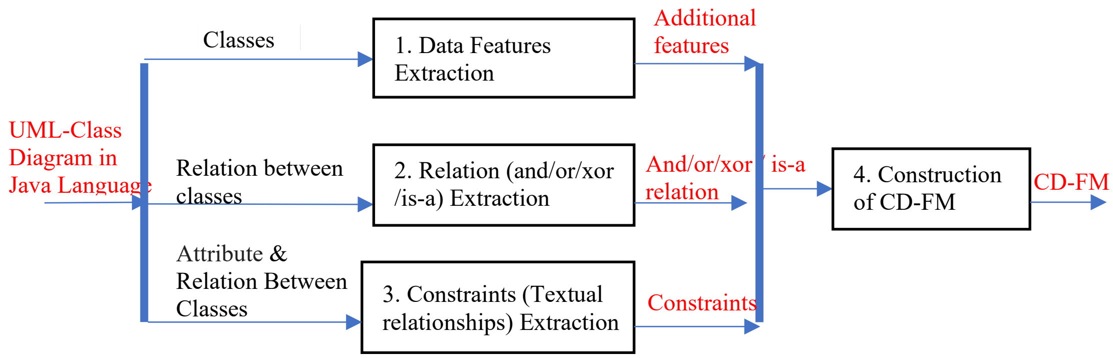


Figure 7a: The CD-to-FM process using UML notations

```
Include   Data-Features-Extraction (Java CD-Java, feature Additional-features)
          Relation-and/or/xor /is-a-Extraction (Java CD-Java, relation Additional-relationships)
          Constraints-Extraction (Java CD-Java, Constraint Additional-constraints)
-----------------------------------------------------------------------------------------------
Process CD-to-FM (in Java CD, out FM CD-FM);
  Feature Additional-Features; Relation Additional-relationships, Constraint Additional-constraints
Begin
   Parallel:
    {
     Data-Features-Extraction (CD-Java, Additional-features);
     Relation-and/or/xor /is-a-Extraction (CD-Java, Additional-relationships);
     Constraints-Extraction (CD-Java, Additional-constraints);
    }
   CD-FM ← Construct-FM (Additional-features, Additional-relationships, Additional-constraints);
End CD-to-FM
```

Figure 7b: The CD-to-FM process using algorithmic notations

and a set of enumerated features. New features may be discovered during the generation of the FM, and features that may encounter a problem during the generation process might be identified. But this technique does not detect all feature groups. Sometimes not disclosing all features can cause problems during the development stage. The aim of generating feature models is to identify all possible features for developing a product.

In another study [3], the work was done to improve relations from what was used as input. Hasse was a diagram used to describe the concept of lattice associated with the formal context. The equivalence class feature diagram technique was used to generate FMs. But this technique does not detect all feature groups.

In [4] the authors proposed a manual generation of FMs. This process was divided into several stages, including identifying features, groups of features, commonality, variability, and ending with the generation of a FM. The researchers also specified the type of input, which was in the form of an activity diagram. This method is manual and it was recommended to be automated.

The researchers, in [11], have written the CVL language, which is used to represent a FM. It contains an Imply relationship in two forms. The first one repeating the feature and linking it to the feature that brings them together. The second one represents it as constraint to avoid complexity and the occurrence of a network that makes it difficult for the reader to understand the model.

The comparison between what this study proposes and the recent works mentioned above will be summarized as it follows:

(1) Improving the FM concepts to cover major aspects of SPL (data, actors, ...)

All the previous works relied on generating the traditional aspect of the feature model only. In this paper, core feature model concepts are developed and enhanced to support more SPL aspects. In fact, data and actor features were added with

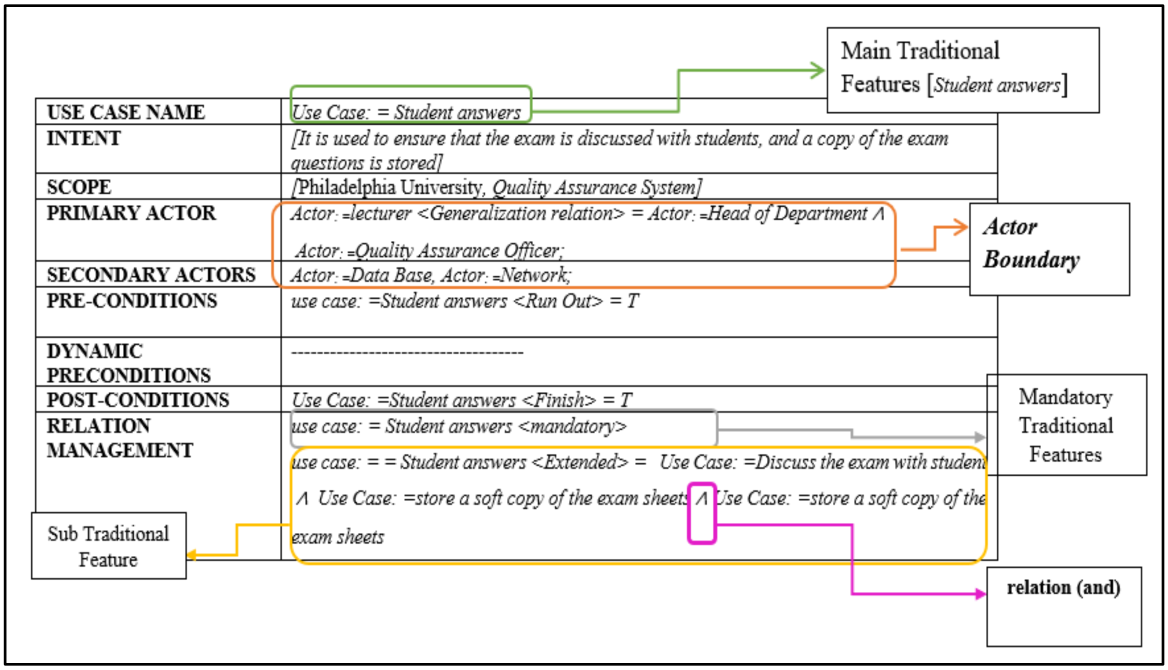


Figure 8a: An Output of UC-to-FM extraction phase, applied on QA agenda at Philadelphia University

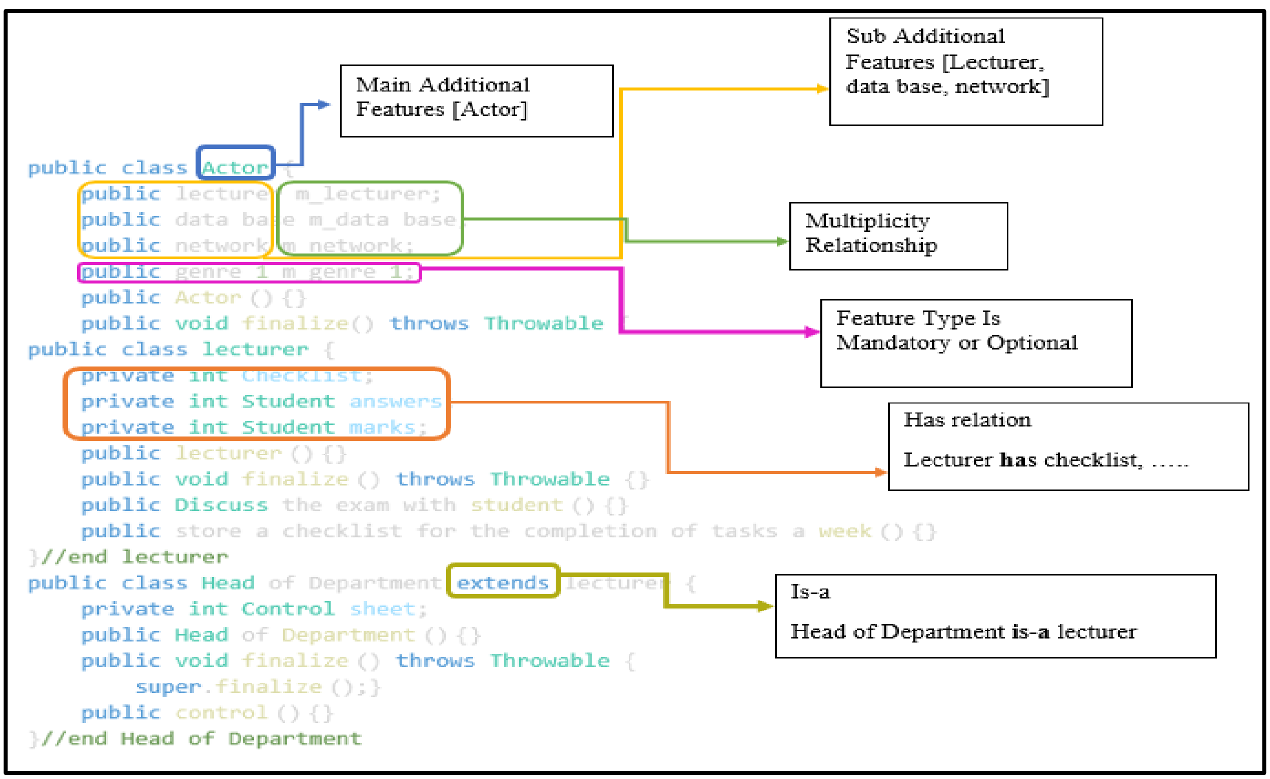


Figure 8b: An output of CD-to-FM extraction phase, applied on QA agenda at Philadelphia University

their inherent relationships.

(2) Using UML like -Diagrams (UC, CD, ...) as inputs to the generation approach

Not all of the above works used UML like diagrams to represent the requirements. In this paper, a use case description with a logic-based language was used along with class diagram translated into Java Language. The use of these diagrams reduces ambiguity and lack in requirements; as clear and understandable requirements help in the correct generation of the feature model.

(3) Formal definition of the FM generation approach

All the above works used several methods to present their work, but rare are the works that formalized their approach. In this paper, the generation approach was formalized at all levels: inputs using UML, the process using UML and algorithmic notations, and the output with an enhanced FM formalism.

(4) Completely automated generation approach

A few of the previous works above provided automatic generation for the traditional aspect of feature models, and the rest of the workers provided manual feature generation with an approach to generating. This paper presented a semi-automated approach (without a completely operational tool) for FM generation that starts from the input and ends with the target FM. The inputs and the output were formally defined using appropriated languages and tools, whereas the algorithms of the generation process were specified.

7 Conclusion

According to the previously presented study on FM generation to represent the commonalities and variabilities from SPL, it was stated that the current methods only generate traditional feature models and do not represent the whole aspects of a system. The feature model generation methods do not deal with the data and actor aspects, they only dealt with the functional aspect of a system. In this work, it is proposed that the inputs be represented by UML models. An approach has been developed consisting of two parallel processes, the first process consists of extraction of a UC-FM from use case description, and the second process consists of extraction of CD-FM from class diagrams. The approach ends with merging UC-FM with CD-FM in a target FM. The proposed approach feasibility was validated on a real business domain that is the quality assurance in higher education limited to Philadelphia University Quality Assurance Agenda system. A comparative evaluation with the closest recent works, based on some significant and usual criteria, was conducted. It worked out to clarify the value of the proposed approach regarding some relevant works. As future work, the approach could be fully automated by providing an automated supporting tool. Other input models could be investigated in order to identify the most suitable for the generation process in terms of cost and quality. The limit of the generated FM could be enlarged by rely identifying the needed and optimal information to include.

References

[1] P. Arcaini, A. Gargantini, and M. Radavelli, "Achieving Change Requirements of Feature Models by an Evolutionary Approach," *J. Syst. Softw.*, 150: 64-76, 2019.

[2] W. K. G. Assunção, S. R. Vergilio, and R. E. Lopez-Herrejon, "Automatic Extraction of Product Line Architecture and Feature Models from UML Class Diagram Variants," Information and Software Technology, 117:1-19, 2020.

[3] J. Carbonnel, M. Huchard, and C. Nebut, "Modelling Equivalence Classes of Feature Models with Concept Lattices to Assist their Extraction from Product Descriptions," *J. Syst. Softw.*, 152:1-23, 2019.

[4] O. Komarudin, D. Adianto, and A. Azurat, "Modeling Requirements of Multiple Single Products to Feature Model," *Procedia Comput. Sci.*, 161:107-114, 2019.

[5] L. Li, Y. Zheng, M. Yang, J. Leng, Z. Cheng, Y. Xiem, P. Jiang, and Y. Ma, "A Survey of Feature Modeling Methods: Historical Evolution and New Development," *Robot. Comput. Integr. Manuf.*, 61(101851):1-16, September 2020.

[6] Z. Liu, N. Japkowicz, R. Wang, Y. Cai, D. Tang, and X. Cai, "A Statistical Pattern based Feature Extraction Method on System Call Traces for Anomaly Detection," *Inf. Softw. Technol.*, 126:1-13, 2020.

[7] M. Marques, J. Simmonds, P. O. Rossel, and M. C. Bastarrica, "Software Product Line Evolution: A Systematic Literature Review," *Inf. Softw. Technol.*, 105:190-208, January 2019.

[8] L. Montgomery, D. Fucci, A. Bouraffa, L. Scholz, and W. Maalej, "Empirical Research on Requirements Quality: A Systematic Mapping Study," Requirements Engineering, 27(2):183-209, 2022.

[9] OMG, Unified Modeling Language 2.5.1, 2017

[10] D. Pantazis, P. Goodall, P. P. Conway, and A. A. West. "An Automated Feature Extraction Method with Application to Empirical Model Development from Machining Power Data," *Mech. Syst. Signal Process.*, 124:21-35, 2019.

[11] I. Reinhartz-Berger, K. Figl, and Ø. Haugen, "Investigating Styles in Variability Modeling: Hierarchical vs. Constrained Styles," *I nf. Softw. Technol.*, 87:81-102, 2017, doi: 10.1016/j.infsof.2017.01.012.

[12] M. K. Sabariah, P. I. Santosa, and R. A. Ferdiana, "Proposed User Requirements Document for Children's Learning Application,", *International Journal of Advanced Computer Science and Applications*, 11(9):317-324. 2020.

[13] A. G. Salguero and M. Espinilla, "Ontology-based Feature Generation to Improve Accuracy of Activity Recognition in Smart Environments," *Comput. Electr. Eng.*, 68:1-13,

March 2018.
[14] T. Yousef, E. Nafar, S. Ghoul, L. Quraan. "Automatically Generated Feature Model from Requirements: Toward an Enhanced Formalism," The 13th International Conference on Information and Communication Systems (ICICS2022), IEEE Xplorer, 2022

Ahlem Yousef holds a MSc in Software Engineering from Philadelphia University, Amman, Jordan.

She is a member of the research team in the Research Laboratory on Bio-inspired Software Engineering. Her research interest includes: Requirements specification, automated specifications generation, and requirements traceability.

Said Ghoul holds a PhD in Software Engineering and a Doctorate in Science in Software Processes.

He is Full Professor of Software Engineering at Philadelphia University, Jordan and Chair of the Research Laboratory on Bio-inspired Software Engineering.

His research interest includes: Bio-inspired Design and Modeling, Variability aspects, and Self-adaptive systems.

Towards Automated Goal Model Generation from UML Use Case and Swimlane Diagrams

Ahlem Yousef*, Said Ghoul*, and Mohammad Tayae*
Philadelphia University, Amman, Jordan

Abstract

The Goal Model of software is one of the important concepts in the goal-based requirements engineering. It helps in specifying the software goals and the relationships between them. Several research works were conducted to generate Goal Model of software from its requirements documents. However, the generated Goal Models merge behavior and soft goals into a single model unit. This merging leads to tangled and complex generated Goal Models. Therefore, the maintenance of these models is hard and costly. The work presented in this paper proposes an approach splitting the generated Goal Model into three separated concerns (aspects) models (behavior, soft, and constraints) that facilitate its evolution and maintenance. The proposed approach is semi-automated, taking UML use case and Swimlane diagrams as inputs and generating a separated aspects model GM as output. The separation of Goal Model aspects led to adding new required information in input requirements specification documents. The feasibility of the proposed approach was validated on a concrete business case (Philadelphia University Quality Assurance Agenda). Its implementation was demonstrated through processes programming with algorithms and UML. Its contribution was demonstrated through its comparison with similar works. According to the observed results, this approach could be valuable in any goal-oriented requirements engineering application.

Key Words: Goal model (GM), behavior goal, soft goal, unified modeling language (UML), UML use case diagram, UML swim lane diagram, goal model generation, goal model maintenance, separation of concerns, aspects programming.

1 Introduction

Software requirements analysis is an approach that allows a better understanding of the requirements collected from stakeholders [3, 17, 19] as it is stated in requirements engineering [12, 20, 22]. These requirements are often complex and extensive [10].

One of the most important concepts in the software requirements analysis and specification is the Goal Model (GM) [1, 4, 9, 11]. IT helps in the definition of a collection of software goals as well as their relationships. It is one of the most important topics in software requirements and specifications analysis [5].

The approaches to generating the GMs differ. Some of them adopted the question method [14] and some others used parsing tree [21]. They also used different forms of requirements documents (textual documents and UML diagrams). The generated goal is modeled with the basic notations of GMs. All the current approaches generate a one piece tangled model merging soft and behavior goals [7, 8, 13, 14, 15, 18, 21, 24]. This tangling leads to problems in maintenance (bad quality and high cost) [6, 8].

As a solution to the problem raised up by the one-piece tangled GM, generated by the current approaches, this paper proposes a semi-automated approach that generates a GM from the requirements specification documents: UML use case and Swimlane diagrams. The generated GM separates clearly the aspects [16, 20] of behavior goals, soft goals, and the relationships between them. So, the obtained GM is composed of separated behavior GM, soft GM, and constraints that specify the relationships between the two separated models. This separation of aspects required the addition of some basic information in the input UML use case and Swimlane.

2 Background

2.1 Goal

GMs are elements of requirements engineering that may also be used more widely in business analysis. Related elements include stakeholder analysis, context analysis, and scenarios [2] among other business and technical areas. Actors' goals are visualized within the boundaries of the actor's goals, tasks are linked through links, and many dependencies such as quality and resources can be represented [7]. GMs are based on the following concepts and relationships between them: goal, task, role, quality, resource, actor, and actor boundary. Some of these concepts concern soft goals (such as: task, actor, actor boundary, resource, quality, ...). Others concern behavior goals (such as: goal).

A goal is the result toward which effort is directed to achieve this very result or objective. Goals are most commonly expressed as imperative sentences beginning with a verb (as in the examples below). For example: "Ensure that only the

* Research Laboratory on Bioinspired Software Engineering. Email: Ahloom13@gmail.com, shgoul@philadelphia.edu.jo, mtaye@philadelphia.edu.jo

account owner can edit his account details", "Allow admin to manage all accounts' privileges".

2.2 Types of Goal

Goals are usually classified in different categories related to their function, behavior, kind, or temporal characteristics:

- *Functional / Non-functional:* Functional goals express services that a system has to deliver. All other goals are non-functional including goals related to the "-ilities" (suitability, reliability, usability, interoperability, verifiability, ...).
- *Behavioral / Developmental-quality:* Behavioral goals express what a system's behavior is to be. These goals are satisfied (or not) by what the system does when it runs. Developmental quality goals express the process by which a system is produced and evolved. These goals are satisfied (or not) by the actions of the people responsible for producing a system.
- *Hard / soft:* Some goals are either satisfied or not satisfied, there is no in-between. Other goals cannot be completely satisfied, but only satisfied to a degree; these are called soft goals. A soft goal is satisfied if it is achieved to a degree that is acceptable, with the understanding that this may cover a wide range of relative achievement, and that complete achievement is not possible. If a soft goal is not satisfied, then it is denied.
- *Achieve / Maintain / Avoid/ Optimize:* An achieve goal refers to a property that is not initially true, but that becomes true, i.e., in terms of temporal logic: *CurrentCondition => eventually TargetCondition*

A maintain goal refers to a property that starts out true and stays that way, i.e., in terms of temporal logic: *CurrentCondition => TargetCondition*

CurrentCondition => always TargetCondition unless NewSituation

An avoid goal refers to a property that is not initially false, but that becomes false, i.e., In terms of temporal logic: *CurrentCondition => not TargetCondition*

An optimize goal refers to a soft property that is to be satisfied.

2.3 Goal Relationships

Goals are related to each other by contribution relationships (partial or complete). Achieving one goal may contribute to achieving another. If the contributing goal is simpler or smaller in scope than the one to which it contributes, the contributing goal is called a sub-goal of the other. The contribution of the sub-goal may be positive or negative (conflict of goals). In the event of a conflict (negative contribution), the achievement of the sub-goal interferes with the achievement of its super-goal. Most popular sub-goal refinements are said to be AND-refinement: if the satisfaction of all of the AND sub-goals is sufficient to ensure the satisfaction of their super-goal. OR-refinement: if satisfaction of any one of the OR sub-goals is sufficient to ensure satisfaction of their super-goal.

3 An Application Case

For managing the Quality Assurance (QA) in its academic programs, Philadelphia University uses a QA agenda, planning its QA management through 16 weeks. The running example used in this study for validating the feasibility of its proposed process is limited to the 7th and 12th weeks because they are the most significant ones. The requirements of the QA agenda are the inputs to the proposed Goal Model generating process: Req-to-GM process.

They are specified using UML use case (Figure 1a) and Swimlane (Figure 1b) diagrams. Using these inputs, an enhanced GM is built up splitting it into behavior goals with their relationships, (2) soft goals with their inherent relationships, and (3) constraints defining relationship between behavior goals and soft goals. This splitting is directed by the syntax shown in Figure 2. The final generated GM is depicted in Figure 3.

4 Target Goal Model

4.1 Enhanced GM by Splitting it into Behavior Goals, Soft Goals, and Constraints

As conclusion to current relevant research works analysis [23], some enhancements to traditional GM are proposed (Figure 2). They are related to GM maintenance levering by splitting it into its three separated aspects: behavior and soft goals and constraints between them. The target GM in Figure 3 is obtained by enhancing GM notations as it is shown in Figure 2.

4.2 GMs Generation Approach – Definitions

The proposed generation approach (Figure 4), is composed of three main components: the inputs that are UML models (Use case description, swimlane diagram), the generation process, which goes through a set of steps for extracting the target GM from requirements, and finally ends with the output, which is the generated GM.

4.3 The Req-To-GM - Process

The Req-to-GM process (Figures 5a, and 5b) takes as inputs requirements specification models (UML use case description and Swimlane Diagram) and generates GM. It is divided into two parallel processes:

- UC-to-GM (Extraction of behavior, soft, and constraints GMs from use case description)
- SL-to-GM (Extraction of soft GM from Swimlane

diagram image).

Each process might do extraction of information, aggregating them into intermediate GMs, and eventually merging the intermediate GMs. In the extraction step, goals (soft, behavior), relationships, quality attributes, and constraints are extracted from the inputs. In the aggregation step, all the outputs from the first phase are aggregated and the behavior, soft, and constraints GMs are built.

Finally, in the merging step, the obtained GMs from the UC-to-GM and SL-to-GM processes are merged into a final output target GM.

4.3.1 Extraction from Use Case Specification. *UC-to-GM*. This process is performed by 6 parallel processes s it is specified in Figure 5c. The process of extracting from the use case document is composed of six parallel sub-processes, which are specified as follows:

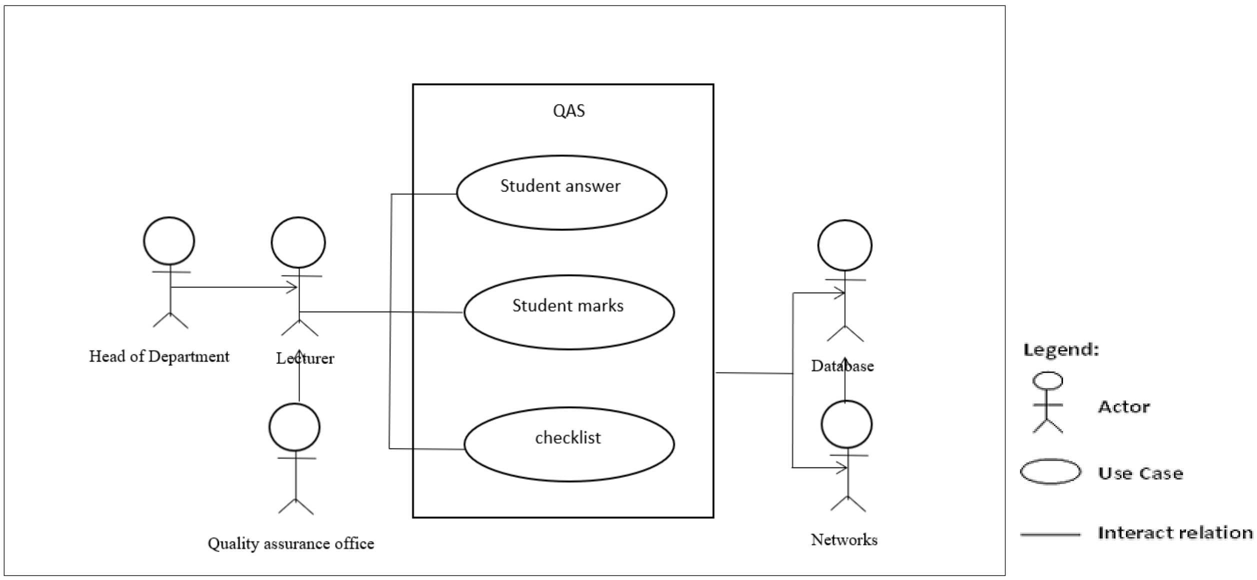


Figure 1a: QA agenda system use case model

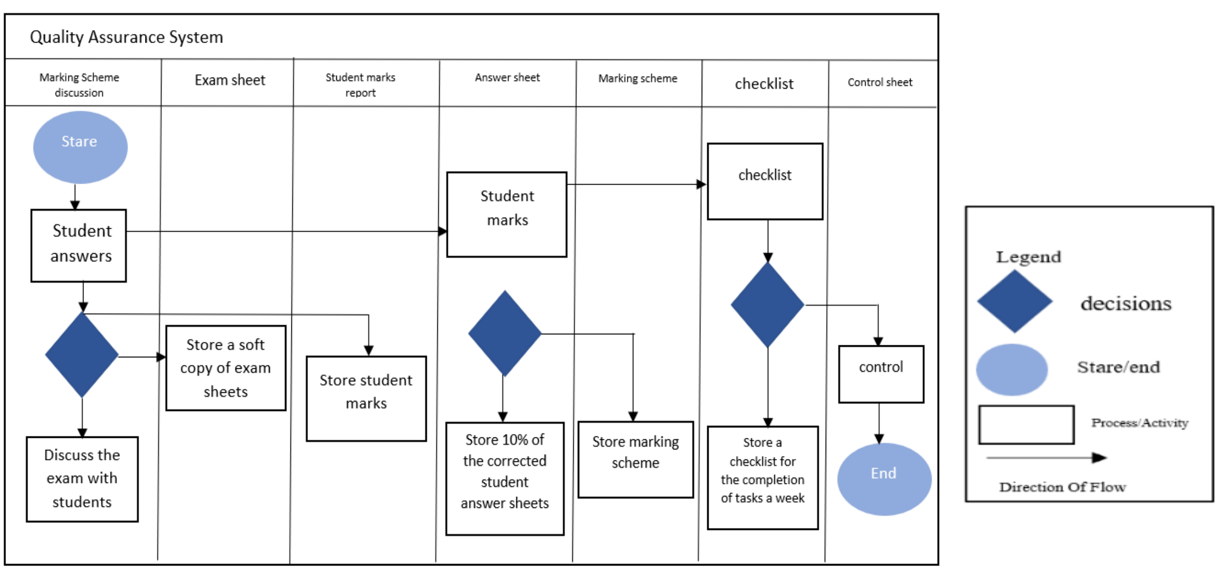
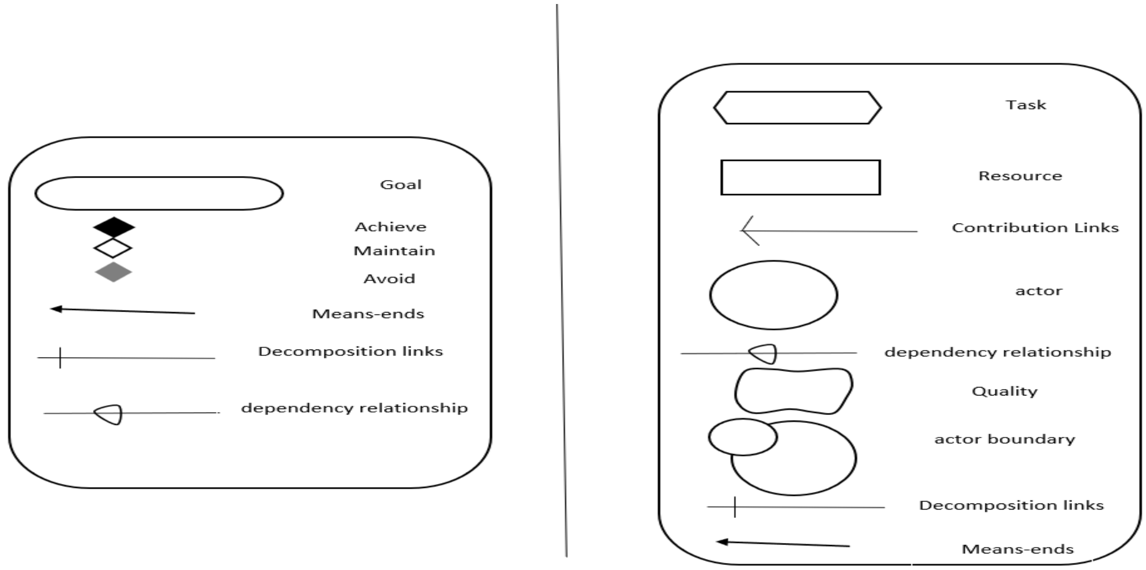


Figure 1b: QA agenda system swimlane diagram



(a) Behavior goals notations

(b) Soft goals notations

Constraints

<list-of- constraints> -- relations between behavior and soft goals

(c) Constraint's notations, defined in a pseudocode language

Figure 2: The enhanced target GM notations

Actor-Extracting: This process extracts actors by reading primary and secondary actor from the use case description (Figure 5d. The actor is represented in the soft goal part.

Actor Boundary Extracting: reading the primary and secondary actor helps identify all parts associated with it in the use case description (Figure 5e). The extracted information is represented in the soft part of the goal model.

Goal Extracting: This process extracts goals behavior from reading the use case name in use case description (Figure 5f). These goals are represented in the behavioral part of the model**.**

Task Extracting: this process reads the use case description and the task component is obtained, and represented in the soft part of the model (Figure 5g).

Relation (AND / IS-A) Extracting: The use case model contains finite relationships (AND relation / IS-A), specified in the N**ote** attribute that are extracted to represent the relationships between goals and between tasks and also between the actors (Figure 5h).

OR Relation /Goal Type /Quality Extraction: This process extracts *OR Relation /Goal Type /Quality* from note (constraints) notations (Figure 5i).

Constraints Extraction: This process extracts *Constraints* from note (constraints) notations (Figure 5j).

Extraction from swim: lane: This extraction is carried out through *SL-to-GM* process which is performed by single process:

Data Extraction (Figure 5k): This task extracts data from the Swimlane as soft goals.

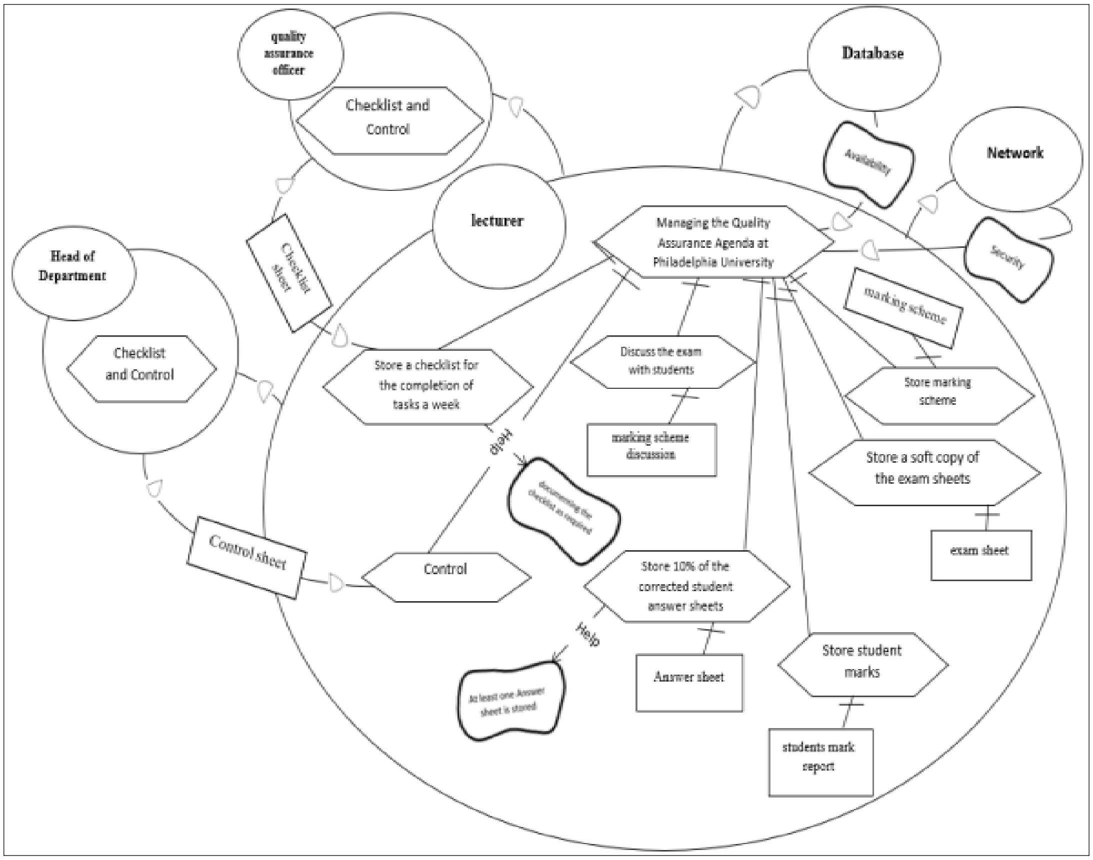
5 Case Study

The feasibility evaluation of the proposed approach was validated by application of the processes Req-to-GM on QA agenda system (paragraph 3). The Figure 6a shows the outputs from the extraction process in UC-to-GM process: actors, actor's boundaries, goals, goals type, relationships between goals, tasks, quality, and constraints. The Figure 6b depicts the soft GM, generated by SL-to-GM process which extracts the data artifacts from the swim lane. The final GM which is the outputs of the Req-to-GM process is depicted in Figure 3.

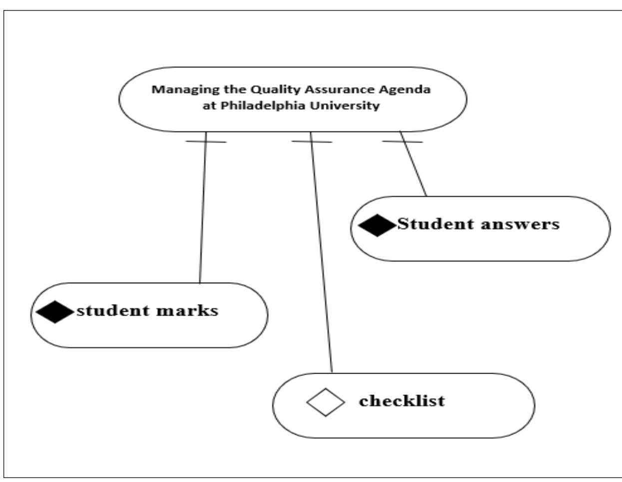
6 Comparison with Similar Works and Evaluation

This section presents a comparison of the proposed approach with some similar works according to the generating process, its inputs and its outputs.

In the work [24], the input is a natural language document which makes it complicated to process as there may be some informal writing leading to lingual mistakes. It requires a further processing to guess the relation between nouns and verbs



(a) Extracted soft GM for QA agenda case



1. Student answers ∧ Store marking scheme ∧ Discuss the **exam** with students ∧Store a soft copy of the exam sheets.
2. Student marks ∧ Store 10% of the corrected student answer sheets ∧ Store student marks.
3. Checklist ∧ Store a checklist for the completion of tasks a week ∧ Control.

(b) Extracted behavior GM for QA agenda case

(c) Extracted constraints GM for QA agenda case

Figure 3: Final generated GM for the case study: QA agenda

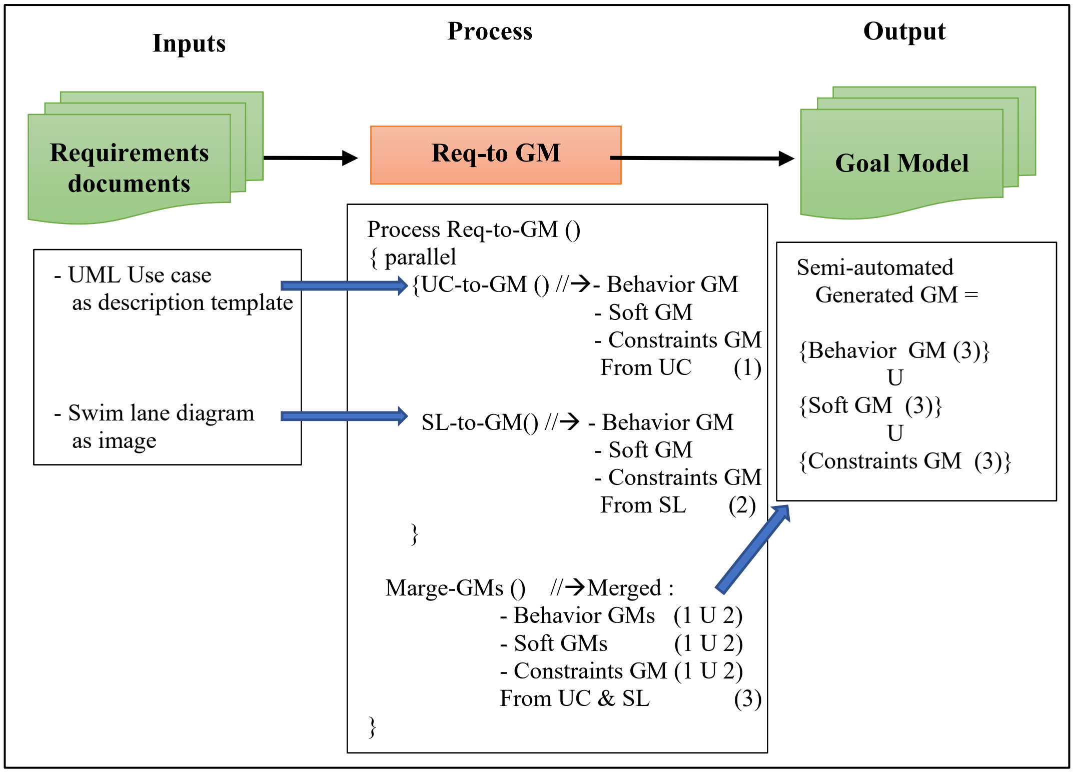


Figure 4: GM generation approach: The Req-to-GM process generates GM from UML use and swim lane diagrams

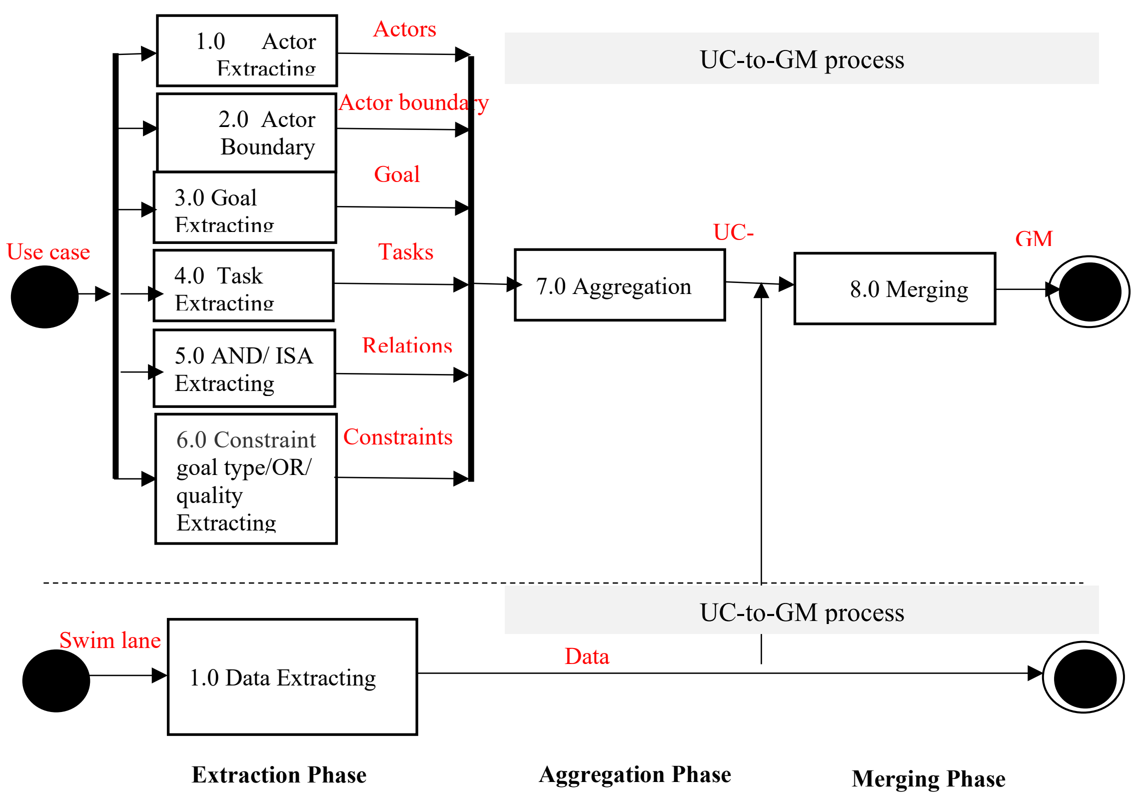


Figure 5a: The Req-to-GM process using UML notations

```
Include goal-model, framework, image //ADT
          UC-to-GM (), Aggregation (), SL-to=GM (), _U_;// functions
---------------------------------------------------------------------
Process Req-to-GM (in framework UC-Description, in image SL-image, out goal-model GM)
  goal-model Goal-Model-Elemts, Behavior-GM, Constraints-GM; GMUC, GMSL;
Begin
   Parallel
   {
      ( //UC-to-GM process
        UC-to-GM (UC-Description, Goal-Model-Elemts)
         GMUC ← Aggregation (Goal-Model-Elemts, Behavior-GM) U
                    Aggregation (Goal-Model-Elemts, Soft-GM) U
                    Aggregation (Goal-Model-Elemts, Constraints-GM))
      )
      (//SL-to-GM process
        SL-to-GM (SL-Image, Goal-Model-Elemts )
        GMSL ←  Aggregation (Goal-Model-Elemts, Soft-GM)
      )
    }// end parallel

    // merging
      GM ← GMUC U GMSL;
   }
End Req-to-GM
```

Figure 5b: The Req-to-GM process using algorithmic notations

```
Include goal, framework, actor, task, relation, constraint //ADT
      Actor-Extracting()  ,Actor-Boundary-Extracting();  Goal-Extracting(),  Task-Extracting(),  Relation-
    and/isa-Extraction(), Constraints-Extraction(); //functions
---------------------------------------------------------------------
Process UC-to-GM (in framework UC-Description, out goal-model Goal-Model-Elemts )
     actor Actors, Actor-Boundary; goal Goals; task Tasks; relation Relations; constraint Constraints;
Begin
  Parallel
   {
     Actor-Extracting ( UC-Description, Actors)
     Actor-Boundary-Extracting ( UC-Description,  Actor-Boundary)
     Goal-Extracting ( UC-Description,  Goals)
     Task-Extracting ( UC-Description,  Tasks)
     Relation-and/isa-Extraction (UC-Description, Relations)
     Constraints-Extraction (UC-Description, Constraints)
   }
   Goal-Model-Elemts ← {Actors, Actor-Boundary, Goals, Tasks, Relations, Constraints)
End UC-to-GM
```

Figure 5c: The UC-to-Gm process specification using algorithmic notations

```
Include framework, actor, //ADT
      Read(), Extract() //functions
```

Process Actor-Extracting (in framework UC-Description, out actor Actors)
actor Actors, Primary-actors, Secondary-actors
Begin
Read (UC-Description.Actors)
Actors ← Extract (Primary-actors) U Extract (Secondary-actors)
End **Actor-Extracting**

Figure 5d: The actor-extracting process specification using algorithmic notations

Include framework, actor //ADT
Read(), Extract-Boundary() //functions

Process Actor-Boundary-Extracting (in framework UC-Description, out actor Actor-boundary)
Begin
Read (UC-Description.Primary-actors, UC-Description.Secondary-actors)
Actors-boundary ← Extract-Boundary (Primary-actors U Secondary-actors)
End **Actor-Boundary-Extraction**

Figure 5e: The actor-boundary-extracting process specification using algorithmic notations

Include framework, goal //ADT
Read(), Extract-Goal() //functions

Process Goal-Extracting (in framework UC-Description, out goal Goals)
Begin
Read (UC-Description.Use cases)
Goals ← Extract-Goal (Use cases)
End **Goal-Extracting**

Figure 5f: The Goal-Extracting process specification using algorithmic notations

Include framework, task //ADT
Read(), Extract-task() //functions

Process Task-Extracting (in framework UC-Description, out task Tasks)
Begin
Read (UC-Description.Use cases)
Tasks ← Extract-task (Use cases)
End **Task-Extracting**

Figure 5g: The task-extracting process specification using algorithmic notations

```
Include framework, relation //ADT
        Read(), Extract-Relation () //functions
```

```
Process Relation-and/Isa-Extraction (in framework UC-Description, out relation Relations)
  Begin
     Read (UC-Description. Note)
     Relations ← Extract-Relation (relation And) U Extract-Relation (relation Isa)
  End Relation-and/Isa-Extraction
```

Figure 5h: The relation (AND / IS-A)-extracting process specification using algorithmic notations

```
Include framework, attribute //ADT
        Read(), Extract-Attributes () //functions
Process ORRelation-GoalType-Quality-Extraction (in framework UC-Description, out attribute ORTypeQuality)
Begin
   Read (UC-Description. Note)
   ORTypeQuality ← Extract-Attributes (OR, GoalType, Quality)
End ORRelation-GoalType-Quality-Extraction
```

Figure 5i: The ORRelation-GoalType-Quality-Extraction process specification using algorithmic notations

```
Include framework, relation //ADT
        Read(), Extract-Constraints () //functions
Process Constraints-Extraction (in framework UC-Description, out constraint Constraints)
Begin
   Read (UC-Description. Note)
   Constraints← Extract-Constraints (Note)
End Constraints-Extraction
```

Figure 5j: The constraints-extraction process specification using algorithmic notations

```
Include image, data //ADT
        Scan(), Data-Pattern-Match () //functions
Process Data-Extraction (in  image SL-Image, out data Data)
Begin
   Scan (SL-Image)
   Data ← Data-Pattern-Match (SL-Image)
End Constraints-Extraction
```

Figure 5k: The Data-Extraction process specification using algorithmic notations

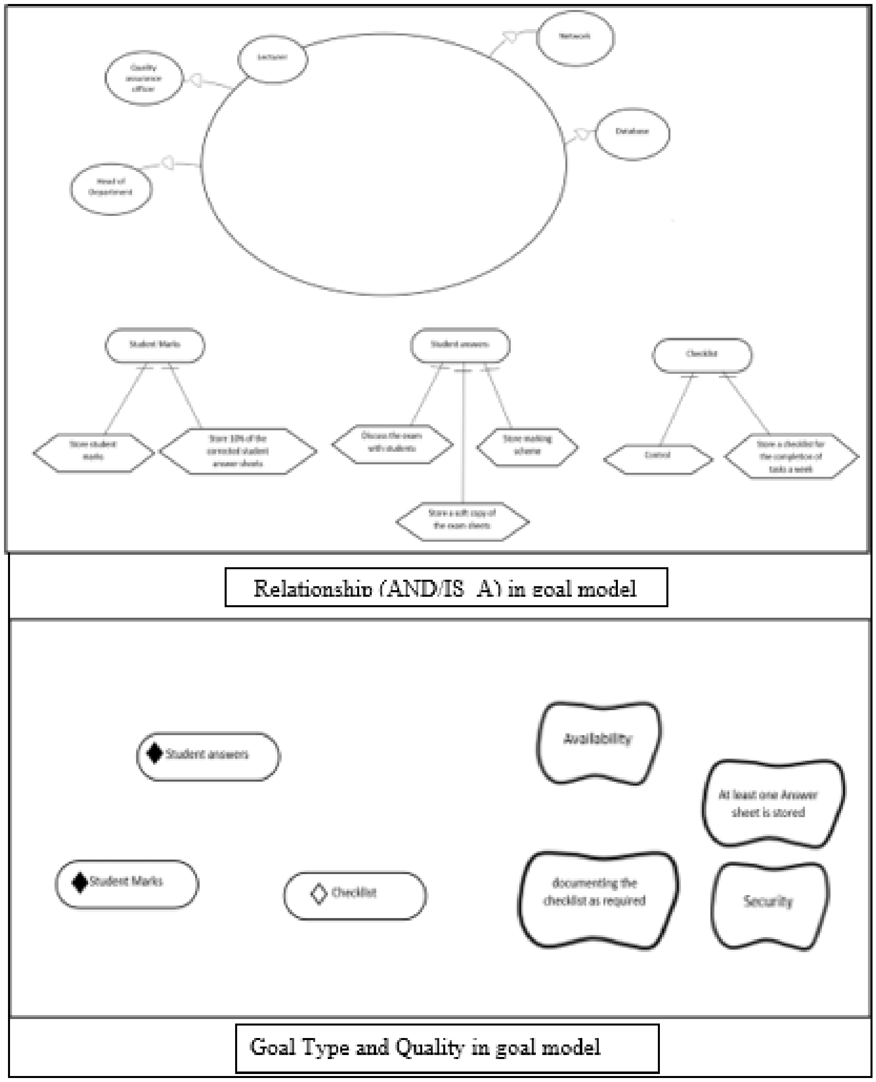
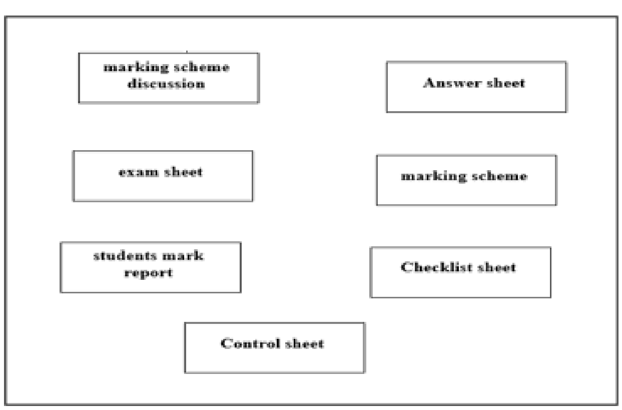


Figure 6a: Generated output by the process UC-to-GM for the QA agenda system



- Student answers ∧ Store marking scheme ∧ Discuss the exam with students ∧Store a soft copy of the exam sheets.
- Student marks ∧ Store 10% of the corrected student answer sheets ∧ Store student marks.
- Checklist ∧ Store a checklist for the completion of tasks a week ∧ Control.

Figure 6b: Generated output by the process SLC-to-GM for the QA agenda system to finally extract the main goals. Consequently, it takes a considerable time. They do not take into account various types of goals and their automatically generated model was only a single unit.

In [14], the authors used a semi-automatic method that relies on a tool that enables the analyst to ask questions and receive answers, and through these answers, goals and relationships can be determined. However, they have noticed that there is a weakness in this method as it defines the relationships between the objectives as "And relationship". Thus, this method as in the previous one takes time for processing natural language phrases but it's simpler. On the other hand, extracted goals were without types, and the generated models were large single units.

Through [21], again the authors have dealt with the requirements document directly (NLP). Thus, generating a parsing tree through which the analyst can determine the goals then draw the goal model. The intentional tree is often complex and difficult to understand. Finally, there is no distinction between target types.

In [7], once again, entries are written in the natural language. The pipeline technology is used to deal with language keys and the NLP is used to analyze sentences to define goals and relationships. The generated model is based on heuristics which may not generate the optimal model. Their work can't identify the (Or) relation but only the (And).

As a conclusion, the above-mentioned works are NLP-based, they do not take into consideration slicing the generated GMs on behavioral, soft, and constraints aspects leading to difficulty in understanding and maintaining the generated models. This paper proposes a solution to this insufficiency by providing a target cleaned GM separating behavior, soft, and constraints goals, which necessitated addition of some information in the input UML diagrams. The application of the generation process, Req-to-GM, on a real application case revealed its feasibility and effectiveness.

7 Conclusion

The study of similar works revealed that the current methods only generate the basic goal models, which are usually huge, complex containing a lot of goals, tasks and other components. It has been observed that the generation methodologies are not formalized which leads to misunderstanding. Therefore, automated generation methodologies and their input and output remain challenging and need enhancements. This paper proposes the inputs specified semi-formally by UML models. The goal is represented by the use case in first level diagram, and the tasks by the use case in the other levels of the use case diagram. The notes help in representing the rest of the requirements of the components and parts of the goal model (the type of the goal and relationships), it is possible through the swim lanes diagram to represent the data (resources). A formal methodology was developed consisting of two parallel processes, the first process performing extraction from use case description, and the second process performing extraction from swim lane diagram image. The methodology consists of three steps: the extraction, the aggregation, and the merging the proposed methodology feasibility was validated on a real business domain running example. The comparative evaluation with the closest recent works led to clarify the conceptual and practical value of the proposed methodology. This study used UML specification diagrams as input (use case description and swim lane diagram). It could be valuable, in the future, to try other specification models and to compare between the obtained results. The variability and meta modeling of input, process, and output is an important issue.

References

[1] S. Abrahão, E. Insfran, F. González-Ladrón-de-Guevara, M. Fernández-Diego, C. Cano-Genoves, and R. Pereira de Oliveira. "Assessing the Effectiveness of Goal-Oriented Modeling Languages: A Family of Experiments," *Inf. Softw. Technol.*, 116:1-24, 2019.

[2] I. Alexander and L. Beus-Dukic, *Discovering Requirements: How to Specify Products and Services*, John Wiley & Sons, 2009

[3] F. Bozyiğit, Ö. Aktaş, and D. Kılınç, "Linking Software Requirements and Conceptual Models: A Systematic Literature Review," *Eng. Sci. Technol. an Int. J.*, 24(1):71-82, 2021

[4] M. B. Duran and G. Mussbacher, *Reusability in Goal Modeling: A Systematic Literature Review,* Information and Software Techonology, 110:156-173, 2019.

[5] I. A. ElSayed, Z. Ezz, and E. Nasr, "Goal Modeling Techniques in Requirements Engineering: A Systematic Literature Review," *Journal of Computer Science*, Science Publications, 13(9):430-439, 2017.

[6] V. Etemadi, O. Bushehrian, and G. Robles, "Task Assignment to Counter the Effect of Developer Turnover in Software Maintenance: A Knowledge Diffusion Model," *Inf. Softw. Technol.*, 143:106786, 2022.

[7] T. Gunes and F. B. Aydemir. "Automated Goal Model Extraction from User Stories Using NLP", *Proc. IEEE Int. Conf. Requir. Eng.*, 2020:382-387,August 2020.

[8] C. Gupta, P. Inácio, and M. Freire, "Improving Software Maintenance with Improved Bug Triaging," *Journal of King Saud University - Computer and Information Science*s, Part A, 34(10):8757-8764, 2022,

[9] J. Horkoff, N. A. Maiden, and D. Asboth, "Creative Goal Modeling for Innovative Requirements," *Inf. Softw. Technol.*, 106:85-100, 2019.

[10] D. Inkermann, T. Huth, T. Vietor, A. Grewe, C. Knieke, and A. Rausch, "Model-Based Requirement Engineering to Support Development of Complex Systems," *Procedia CIRP*, 84:239-244, 2019.

[11] H. Kaiya, K. Hayashi, and Y. Sato, "Tools for Logging and Analyzing Goal Dependency Modeling," *Procedia Comput. Sci.*, 192:1639-1648, 2021.

[12] R. Kasauli, E. Knauss, J. Horkoff, G. Liebel, and F. G. de Oliveira Neto, "Requirements Engineering Challenges and Practices in Large-Scale Agile System Development," *J. Syst. Softw.*, 172:8757-8764, 2021.

[13] Y. Lai, A. Gupta, and Y. Zhang, "Goal-Embedded Dual Hierarchical Model for Task-Oriented Dialogue

Generation," *Proceedings of the 23rd Conference on Computational Natural Language Learning*, pp.798-81, 2019

[14] H. Nakagawa, H. Shimada, and T. Tsuchiya, "Interactive Goal Model Construction Based on a Flow of Questions," Transactions on Information and Systems, 103(6):1309-1318, 2020.

[15] A. Pozanco, S. Fern, and D. Borrajo, "Learning-Driven Goal Generation," *AI Communications*, 3(2):137-150, 2018.

[16] S. R. Raheman, H. B. Maringanti, and A. K. Rath, "Aspect Oriented Programs: Issues and Perspective," *J. Electr. Syst. Inf. Technol.*, 5(3):562-575, 2018.

[17] M. R. R. Ramesh and C. S. Reddy, "Metrics for Software Requirements Specification Quality Quantification," *Comput. Electr. Eng.*, 96:107445, 2021.

[18] Z. Ren and K. Dong. "Exploration via Hindsight Goal Generation," 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), 2019.

[19] M. Sadiq and V. S. Devi. "Fuzzy-Soft Set Approach for Ranking the Functional Requirements of Software," *Expert Syst. Appl.*, 193:11-45, 2021.

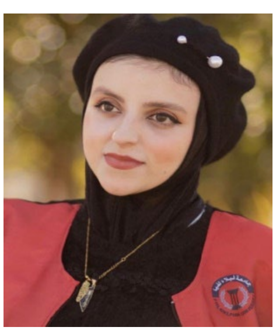
[20] V. Santos, H. Mamede, C. Silveira, and L. Reis, "Methodology for Introducing Creativity in Requirements Engineering," *Procedia Comput. Sci.*, 196:27-35, 2022.

[21] H. Shimada, H. Nakagawa, and T. Tsuchiya, "Constructing a Goal Model from Requirements Descriptions Based on Extraction Rules," *Commun. Comput. Inf. Sci.*, 809(no. June):175-188, 2019

[22] L. Yin, Q. Sun, D. Tang, Y. Xu, and L. Shao, "Requirement-Driven Engineering Change Management in Product Design," *Comput. Ind. Eng.*, 168:1-13, 2022.

[23] A. Yousef, S. Ghoul, and M. Taye, "Automatically Generated Goal Model from Requirements: Toward an Enhanced Formalism," Proceedings of ICSIC2022, pp. 1-6, 2022.

[24] C. Zhang Mentor, A. M. Grubb, and M. Chechik, "Toward Automatic Generation of Goal Models using Natural Language Processing," 1:1-6, 2019, [Online], Available: https://www.graphviz.org.



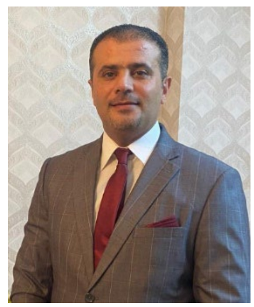
Ahlem Yousef holds a MSc in Software Engineering from Philadelphia University, Amman, Jordan.

She is a member of the research team in the Research Laboratory on Bio-inspired Software Engineering. Her research interest includes: requirements specification, automated specifications generation, and requirements traceability.

Said Ghoul holds a PhD in Software Engineering and a Doctorate es Science in Software Processes.

He is Full Professor of Software Engineering at Philadelphia University, Jordan and Chair of the Research Laboratory on Bio-inspired Software Engineering.

His research interest includes: Bio-inspired Design and Modeling, Variability aspects, and Self-adaptive systems.



Mohammad Mustafa Taye is an Assistant Professor in Software Engineering department at Philadelphia University, Amman, Jordan. He received his Ph.D. degree in Computer Science from De Montfort University, Leicester, UK in 2009, for his dissertation on "Ontology Alignment Mechanisms for Improving Web-based Searching". He also holds MSc. degrees in Computer Science from Amman Arab University for Graduate Study, Amman, Jordan in 2004. Also, holds a bachelor's degree in Computer Science from Irbid National University, Irbid, Jordan, in 2002. Mohammad has been with Philadelphia University since 2009. His research interests include semantic web, ontology, ontologies alignment and matching, ontology languages, artificial intelligence, web service, software requirement, and network.

Effect of Big Data and Analytics on Managing Projects

Mounir Elkhatib
Hamdan Bin Mohamad Smart University, Dubai, UAE

Raed Abu Zitar*,
Sorbonne University, Abu Dhabi, UAE

Khadeej A. Alnaqbi, Wesal Alnaqbi, and Shaima Alharmoodi
Hamdan Bin Mohamad Smart University, Dubai, UAE

Ali Baydoun†
St. George's University, Grenada, West Indies

Abstract

The rapid pace of technological advancement necessitates constant adaptation. As a rapidly evolving field, project management has little choice but to take use of technological breakthroughs to stay relevant and fresh. Using big data analytics, businesses and project managers alike can reap the benefits of this technology. Big data analytics is definitely useful for influencing the future of project management, as outlined in this article's preliminary comments. A new era of 21st-century living appears to be upon us, and project management as a profession appears to be ready to embrace it. Big data collects and stores enormous amounts of data that are becoming increasingly difficult to manage and analyze. The potential benefits and competitive edge of this new technology are motivating the majority of businesses to invest in big data analytics. Structured or unstructured, large amounts of heterogeneous data are processed and managed as "Big Data" in the enterprise. This includes both structured and unstructured data. Analytic methods and technologies are heavily employed in the management and analysis of large and complex data sets for use in a wide range of applications that enhance the performance of a business. This paper analyzes the impact of big data and business analysis on project management. A literature review is followed by primary and secondary data analysis, which includes interviews and surveys for architectural analysis, in an exploratory study. Qualitative and quantitative data are the norm in all studies in this study. The present study is descriptive in nature, as the goal is to examine the impact of big data and business analysis on project management. Below mentioned impacts have been analyzed and concluded that big data analytics helps to reduce the project complexities, reduces the project cost and enhances the project risk management.

* Email: raed.zitar@sorbonne.ae.
† School of Medicine.

Key Words: Big data; analytics, project management; digital transformation; digital disruption; and disruptive technologies.

1 Introduction

There are a wide range of business analytics skills, technologies and apps that may be used to continuously iteratively explore and investigate previous performance to gain actionable insights. Based on data and statistical approaches, business analytics aims to uncover new insights and better understand business performance. Data sets that are huge, diversified, and fast changing are referred regarded as "big data" by many businesses. System capabilities beyond those of ordinary SQL-based databases are required for handling big data. More than 1.5 million managers and analysts are expected to be needed in the United States to cope with big data analytics [15]. The estimated demand for deep business analytical roles could exceed supply produced by the current trend by 140,000 to 190,000. There is a growing consensus among prominent academics and business leaders that big data and analytical techniques will have a significant impact on our economy and society over the next decade or two [13]. Business analysis and big data are examined to ensure that firms will remain relevant in the future. In addition, the potential for use in a variety of fields is also highlighted.

Big data captures and retains enormous amounts of data that are becoming increasingly difficult to manage and analyze. The potential benefits and competitive edge of this new technology motivate most businesses to invest in big data analytics. Large amounts of data, whether structured or unstructured, that are used in project management can be processed and managed using big data techniques. The administration and analysis of massive volumes of complex data sets for use by numerous applications that improve the operation of a business is significantly reliant on analytical methods and technologies in big data analytics.

As defined by [19], a "project" is a one-time assignment with a defined start and end date that involves the allocation of resources in order to achieve predefined goals. Project managers are responsible for the sound management necessary for a project's success. Therefore, the initiatives must be assigned to the proper managers. The assignment of a project manager is a crucial project choice that influences the success of the project and the performance of the organization [20]. Stakeholder expectations and needs are met or exceeded through the application of project management knowledge, skills, tools, and technology [11]. Project management is the systematic control of a project that begins with the project planning phase and ends with the project's conclusion

The term "big data" refers to data sets that are larger and more diverse than those typically handled by relational database management systems [22]. The term "big data" refers to data that has a high volume, a high rate of change, and a high variety; it is characterized by the use of new technologies and techniques to acquire and analyze the data; and it enhances decision-making capabilities, and provides more insight, discovery, and support for process optimization [2]. Social media, movies, digital photos, and sensors are just a few of the many places where big data may be found. By utilizing data analytics, a company can gain a competitive edge, improve efficiency, and discover new revenue streams. Data analytics can be used for a variety of purposes, including credit risk assessment, marketing, and detecting fraud [25]. Analytical techniques can be divided into descriptive, predictive, diagnostic and prescriptive analytics [10]. Analyzing large amounts of data with the goal of discovering new patterns and gaining new insights [26]. There are a lot of variables at play when it comes to project management, both inside and externally. When risks become 'Issues', they must be managed in order to minimize their negative impact on the project's outcome. Prudence dictates that project teams identify and manage risks on a continuous basis by taking an active role in doing so. That would necessitate the recording of all risk incidents. Additionally, when risks actually occur and turn into issues, the firefighting and troubleshooting efforts necessary to deal with them are documented thoroughly. This method generates a great deal of data that may be used to better manage risks and difficulties. Assessing risk and issues-related data using big data analytics can lead to new approaches and procedures for discovering, analyzing, prioritizing, monitoring, and establishing risk response strategies. Risk management may not be given the attention it deserves because of this, but in order to make it more widely accepted by those involved in projects and use the newly developed procedures to better address risks and issues. It is imperative that new methods, procedures, and techniques be developed. These goals can be met with the help of big data analytics.

2 Literature Review

The Literature review covers the four elements of digital disruption (Business – Technology – Industry and Society) with their respective 3 parameters as follows:

2.1 Effect of Business Analysis and Big Data Project Management from Business Side

2.1.1 Market. Marketing campaigns can be tailored to specific client needs thanks to the use of big data. Predicting what clients will require in the future can actually be done with the help of a good analysis. With this information, it is much easier to create customized adverts for certain products. Also, it encourages corporations to pay greater attention to what their customers want. Predicting future trends can be done by collecting and analyzing data regarding client wants. Insights gleaned from big data analytics can be used to create new goods and services for businesses. As a result, they are better able to anticipate the demands of their clients. The corporation can provide data- driven evidence for product creation by taking into account client demand, interest, or product popularity. Now you don't even have to wait for your clients to tell you what they want; you can meet their demands faster than ever before. In addition, being more creative than your competitors is a bonus [24].

2.1.2 Pricing. Big data gives information that can be used to cut costs and better allocate resources. Data gathered may be utilized to alter and improve corporate operations, which in turn reduces costs and increases profits. In addition, big data analysis makes it easier to save costs and increase efficiency. Many companies are saving a great deal of money by utilizing big data's operational cost-cutting capabilities. It is also possible to use big data to develop new goods. With the information obtained, companies can either help their clients solve problems related to big data or form partnerships with third parties that share aggregated data. Telcos, banks, and insurance providers now have new avenues for generating cash. In order to provide their customers with more relevant offers, certain merchants may be interested in using telecommunications companies' location-based data. For all sides, this is a win-win.

2.1.3 Delivery. With the help of big data, businesses can better understand their goods and make better decisions resulting in increased operational efficiency. This can be done by analyzing operational and historical data acquired from a variety of sources such as machine logs and social media, as well as the web or mobile apps [23]. There are many chances to consolidate and analyze information linked to planning and delivery operations because they are frequently documented. The collection and analysis of big data related to project planning and execution has never been easier because of the rising use of technology in projects. This data volume insights into planning and delivery-related big data could be gained by consolidating it across various business segments (e.g., information technology (IT), human resources (HR), manufacturing, logistics (logistics), and inventory management (inventory)), industries (e.g., construction and IT), economic sectors (e.g., mining and agriculture), and geographic regions (e.g., Asia-Pacific, Americas, Europe, etc.).

Project and delivery activities may benefit from big data analytics in the future [21].

2.2 Effect of Business Analysis and Big Data Project Management from the Technology Side

2.2.1 Invention. Organizations now rely on data as the foundation for their success. Massive storage capacities and data collection methods make it possible to access large amounts of data. It plays a key role in generating an enormous amount of data that must be processed in order to extract value from it. Furthermore, storing traditional data is far less expensive. This information is always changing, necessitating an innovative data analytics system and storage and analysis methods. As a result, they require thorough examination in order to yield the necessary data [17].

2.2.2 Design. For businesses, the rise of social media and weblogs led to the rise of big data. As a result, additional data sources have been added to the fundamental analytics and business intelligence (BI) activities, allowing for real-time, in-depth analytics and BI with operational integration. As the amount of digitally generated data increases at an exponential rate, data warehouse technology is having difficulty keeping up. Several studies have lately reported on the enormous amounts of raw data created by diverse data sources, requiring big data technology for analysis [16].

2.2.3 Usage. For more than a decade, researchers and businesses have been studying and implementing big data analytics. Since big data may be used in a wide range of industries and applications (such as healthcare and commercial decision-making), this is a major factor in the rise in popularity of big data. There have been numerous research and reviews published recently on big data analytics, implementations, and related technologies as a result. [12].

2.3 Effect of Business Analysis and Big Data Project Management from Industry Side

2.3.1 Standard. Using data, statistical and quantitative analysis, and explanatory and predictive modelling, business analytics helps make meaningful decisions and enhance corporate processes. Real-time and non-real-time analytics, strategic and tactic, planned and unplanned, and organized and unstructured are only few of the various types of business analytics [4]. For a long time, managers have relied on business data to guide their decisions. With business analytics, not only do they look back at the past, but they are also looking ahead to see where they can make improvements [14]. As stated by [6], business analytics is a collection of unstructured data that can't be evaluated using relational database management technologies such as big data, test analytics, web analytics, network analytics and mobile analytics. Graduates with a major in big data business analytics can pursue three distinct job paths, according to [9]: management consulting at the top tier, financial and risk analysis, or work as a data scientist. Because healthcare managers and organizations can better translate data and have actionable intelligence based on present infrastructure, [27] recommends that the healthcare industry may start with small data analytics before using big data analytics.

2.3.2 Method. These approaches have been used to assess a significant volume of data generated by different businesses. As a result, every organization must be able to have access to increasing amounts of transactional data more quickly. Real-time data analysis aids companies in gaining a better understanding of the past and predicting the future. Therein lies the allure of real-time analytics, which gives us the ability to see what has occurred, why it has occurred, what might happen in the future, and, most importantly, what we can do about it (prescriptive) [8].

2.4 Effect of Business Analysis and Big Data Project Management from the Society Side

2.4.1 Culture. Over the past few years there has been a lot of attention paid to big data, business analytics, and "smart" environments in driving organizational decision making, as organizations are working on how to give purpose to the data, and get value-driven answers that will increase their performance [5]. Among the most major recent technology shifts in industry and academia, big data is likely to be one of the most crucial [1]. Massive amounts of information are being generated and made available online and in digital media ecosystems under the umbrella term "big data." Big data can be generated from a variety of sources, including transactions, social media posts, and sensors put in a wide variety of things (e.g., mobile phones, home appliances, cars, etc.). For example, it is possible to forecast future occurrences with the use of big data analytics and artificial intelligence, as well as to automate processes, transform businesses, and establish new sorts of businesses as it can do now [18].

2.4.2 Habits. Using big data to improve corporate operations isn't the only benefit. Opportunities for economic progress and a better standard of living exist for the society as a whole. Many industries and businesses can benefit from the use of big data analytics. These include a rise in the quality of health care and education, as well as increased safety and security for the country [7].

2.4.3 Movement. With this information, it can assist policymakers in creating rules that will allow investors to have safe playgrounds, help waste managers locate the type of garbage that is generated more frequently in a certain area, and provide insight into how to better share waste collection materials. Big data and business analytics can also be used to evaluate instructors' performance and enhance their attitude to work. It is also possible to effectively plan public transportation in large cities by utilizing mobile network location data [3].

3 SWOT Analysis

Table 1: SWOT analysis

Strengths	**Allocation of Resources:** Big data provides information that may be utilized to save expenses and distribute resources more effectively. The information acquired may be used to change and enhance company processes, lowering costs and increasing profits
	Production of goods and services: Big data analytics may be used by businesses to create new goods and services. As a result, they are more equipped to address the demands of their customers. The corporation may provide data-driven proof for product creation by evaluating client demand, interest, or product popularity.
Weakness	**High Implementation Cost:** The costs of keeping and processing massive volumes of data to understand it might be prohibitive. Traditional clinical research approaches will continue to be used since big data lacks the scientific rigor associated with RCTs.
	Benefits limited audience Furthermore, big data is meant to benefit only
	Personal data security and privacy issues Privacy and security of personal data are important problems for big data platforms. As a result, individuals may decline to engage in studies utilizing large data platforms.
Opportunities	**Marketing Campaigns** Due to the usage of big data, marketing efforts may be adapted to individual customer demands. With the aid of a good study, it is possible to predict what clients will demand in the future. It is considerably easier to build tailored advertisements for certain items using this information.
	Creates insights in internal planning Due to the increasing usage of technology in projects, using big data in project planning and execution has never been easier. This quantity and diversity of data might be utilized to perform research and get insight into how to rethink internal planning procedures and parameters in order to apply new and innovative ways.
Threats	**Ethical Issues and Genetic Screening** Excessive genetic screening, a feature of big data platforms, creates ethical concerns. Despite the multiple potentials and directions that big data may go in the future, such systems are associated with numerous risks. To begin with, recognizing risk factors might cause people anxiety, especially if the under-lying disease is incurable or has no known solution. Insurance companies utilize this knowledge to boost premiums for those who are genetically predisposed to particular diseases, increasing the cost of care.
	Threat of data loss Data loss and third-party access are other threats to health big data systems. Finally, institutions may join together to benefit from bio-monitoring and other data gathering technologies at the expense of patients.

4 Methodology

4.1 Research Approach

The goal of this study is to examine the impact of big data and business analysis on project management, an organization that aims to be completely data-driven in all of its decision-making and quantitative data are the norm in all studies in this study. Project Managers/Specialists from companies in UAE were the study's unit of analysis. The managers were basically from different known companies of United Arab Emirates that involve Intertec Systems, Ewaan Tech, LRB infotech, DataQraft, and Tech Falcon.

The present study is descriptive in nature, as the goal is to examine the impact of big data and business analysis on project management. In this case, questionnaire was viewed the appropriate method for data collecting since it enabled project managers to respond whenever they had spare time and without the researchers' personal intervention.

The research method is based on the Simplilearn approach for digital disruption and strategies for digital transformation. This approach is based on analyzing key digital transformations across different dimensions – business, technology, industry, and society. Then they create their own digital transformation prioritization matrix to identify the digital transformation initiatives most relevant to your industry and organization. The Simplilearn investigation questionnaire is structured in 9 steps as follows:

1. List of top 10 digital transformations in the organization and a SWOT analysis.
2. A list of top 10 digital initiatives by digital native disruptors most relevant to the organization. Using a what, how, and outcome framework to brainstorm the information.
3. A list of top three digital disruptions across all four elements – business, technology, industry, and society most relevant to the organization.
4. A list of top three digital initiatives by the competitors.
5. A list of top three digital disruptions across all the five key areas of disruption that are most relevant to the organization. We are making progress with our understanding of digital initiatives. Now how about getting some ideas going? Let's look at each of the key

areas of digital disruption one by one using the trend-benefit framework

6. A list of top three transformations across the following key technologies that are most relevant to the industry.
7. Based on the previous steps, a list of all key transformations compiled.
8. A list of transformations made across the impact/difficulty matrix.
9. The most relevant digital transformation plans identified.

The 9 steps gathered based on 30 interviews conducted with key leaders, project managers, technology specialists, and digital transformation specialists.

Details about the above nine steps are provided in appendix A.

5 Analysis

As previously said, data collection for the study took a long time; as a consequence, the researcher wanted to double-check the data's correctness before going on to data analysis. The data analysis phase of this research is critical since the findings will be utilized to draw inferences from the investigation's overall outcomes. Data analysis, in this context, refers to the process of verifying, purifying, altering, and changing data after it has been received in order to arrive at a valid result. In order to meet the research objectives, the researcher was charged with selecting the most appropriate data analysis technique for this study, taking into account the nature of the data obtained. In this study, the researcher employed certain methods to investigate quantitative data, as outlined in the previous chapter. The goal of finding the best technique is to exclude data that isn't relevant to the study or to spot problems in the research. As previously stated, this study gathered data in quantitative format. The goal of finding the optimal approach is to exclude data that isn't relevant to the study or to spot weaknesses in the research.

Several interviews were conducted based on secondary data, and project managers were asked a variety of questions about Big data and its use. Several structured questions were asked from the project managers in their interviews and these inquiries came in the form of a questionnaire. The following is a list of the answers to these questions:

5.1 "Can Big Data Handle Large-Scale Real-Time Applications?"

5.1.1 Interpretation. In question 1, we asked our respondents can big data handle large-scale real-time applications while the Likert scale for this question was 1 to 5 (Yes, More likely, Neutral, No, Not at all). As Figure 1 shows that almost 25% of people rated it "Yes" and suggested the idea that yes, big data can handle large-scale real-time applications, while 32% rated it "More Likely to" and suggested the idea that Big data can handle large-scale real-time applications to a great extent, while 30% had biased views and they opted neutral, 8% disagreed with the idea, and remaining 5% opted "Not at all", meaning they totally disagreed with the idea. Also, the total number of project managers that answered our question was 15, while this question was asked from all 20 managers, which means 5 people skipped it intentionally. In a nutshell, our analysis shows that most rated their response as "More likely to" and least rated this idea as "Not at all," i.e., 5, which is 32% and 5%, respectively.

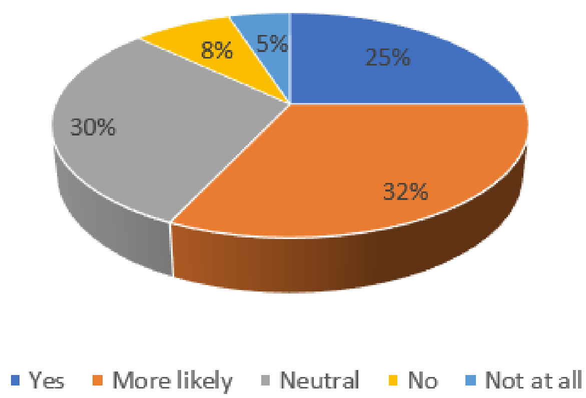


Figure 1: Scaling about big data

5.2 "Does Big Data Help to Improve Costs?"

5.2.1 Interpretation. In question 2, we asked our respondents does big data help to improve costs while the Likert scale for this question was 1 to 5 (Yes, More likely, Neutral, No, Not at all). As Figure 2 shows that almost 44% of people rated it "Yes" and suggested the idea that yes, big data helps to improve costs, while 47% rated it "More Likely to" and suggested the idea that big data helps to improve costs to a great extent, while 2% had biased views and they opted neutral, 2% disagreed with the idea and remaining 4% opted "Not at all", meaning they totally disagreed with the idea. Also, the total number of project managers that answered our question was 14, while this question was asked from all 20 managers, which means 6 people skipped it intentionally. In a nutshell, our analysis shows that most rated their response as

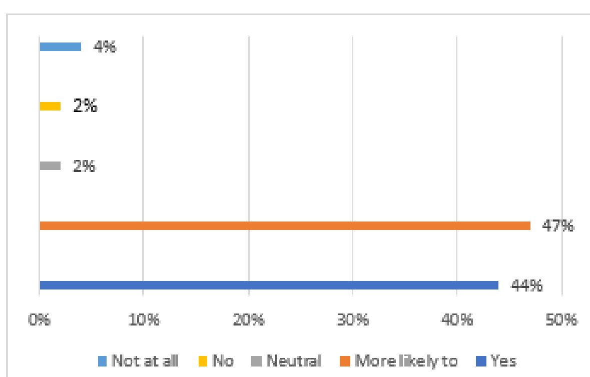


Figure 2: Scaling about big data

"More likely to" and least rated this idea as "Not at all and Neutral," i.e., 5, which is 47% and 2%, 2% respectively.

5.3 Does Big Data Help to Form New Partnerships?"

5.3.1 Interpretation. In question 3, we asked our respondents does big data help to form new partnerships while the Likert scale for this question was 1 to 4 (Extremely likely, Quite likely, Slightly likely, Not at all). As Figure 3 shows that almost 38% of people rated it "Extremely likely" and suggested the idea that yes, big data helps to form new partnerships, while 16% rated it "Quite likely" and suggested the idea that Big data helps to form new partnerships to a great extent, while 10% had biased views and they opted slightly likely, 36% disagreed with the idea and opted "Not at all", meaning they totally disagreed with the idea. Also, the total number of project managers that answered our question was 16, while this question was asked from all 20 managers, which means 4 people skipped it intentionally. In a nutshell, our analysis shows that most rated their response as "Extremely likely to" and least rated this idea as "Slightly likely," i.e., 3, which is 38% and 10%, respectively.

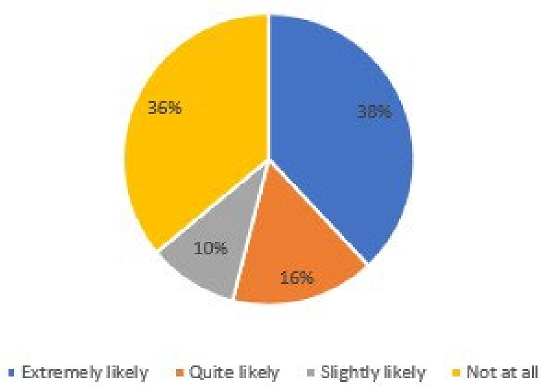


Figure 3: Scaling about big data

5.4 "Can we Build a Library to do an Auto Conversion of Standard Algorithms to Support MapReduce?"

5.4.1 Interpretation. In question 4, we asked our respondents can we build a library to do an auto conversion of standard algorithms to support MapReduce? We asked our respondents to rate their opinion on a scale of 1 to 5. As Figure 4 shows that almost 19% of project managers rated it as 1 and suggested the idea that No, we cannot build a library to auto-convert standard algorithms, while 11% rated it "2" and suggested the idea that it possibly cannot happen to some extent, while 30% had biased views and they opted the option 3, 20% agreed with the idea and opted "4," and 20% chose "5" as they fully agreed with the idea. Also, the total number of project managers that answered our question was 15, while this question was asked from all 20 managers, which means 5 people skipped it intentionally. In a nutshell, our analysis shows that most rated their response as "3" and least rated this idea as "2".

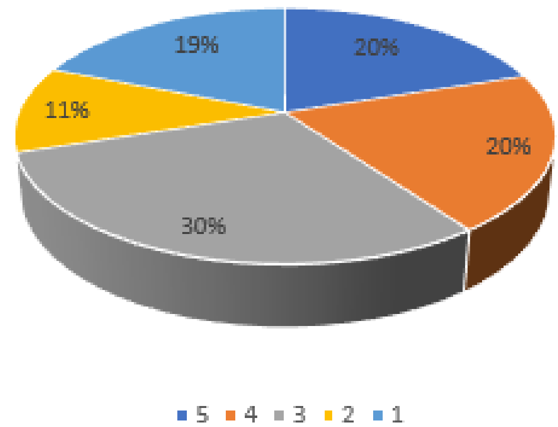


Figure 4: Scaling about big data

5.5 "Can the Existing Systems be Enhanced with Low Latency and More Accuracy?"

5.5.1 Interpretation. In question 5, we asked our respondents can the existing systems be enhanced with low latency and more accuracy?, while the Likert scale for this question was 1 to 5 (Yes, More likely, Neutral, No, Not at all). As Figure 5 shows that almost 34% of people rated it "Yes" and suggested the idea that yes, while 45% rated it "More Likely to", suggesting that the existing systems can be enhanced with more accuracy and low latency to a great extent, while 12% had biased views and they opted neutral, 4% disagreed with the idea and remaining 4% opted "Not at all", meaning they totally disagreed with the idea. Also, the total number of project managers that answered our question was 15, while this question was asked from all 20 managers, which means 5 people skipped it intentionally. In a nutshell, our analysis shows that most rated their response as "More likely to" and least rated this idea as "Not at all and Neutral," i.e., 5, which is 45% and 4%, 4% respectively.

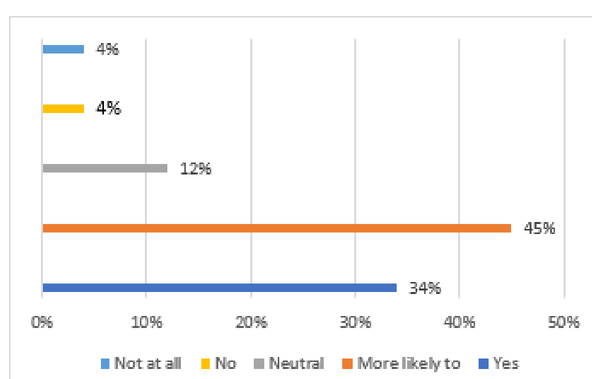


Figure 5: Scaling about big data

5.6 "Which is More Important (1) Client Demand, (2) Interest, or (3) Product Popularity in Terms of Data-Driven Popularity?"

5.6.1 Interpretation. In question 6, we asked our respondents which is more important (1) client demand, (2) interest, or (3) product popularity in terms of data-driven popularity? While the Likert scale for this question was 1 to 3. As Figure 6 shows that almost 40% of people opted for "client demand", while 35% opted for "product popularity," while 25% chose "interest". Also, the total number of project managers that answered our question was 15, while this question was asked from all 20 managers, which means 5 people skipped it intentionally. In a nutshell, our analysis shows that most opted for "client demand" while least opted for "interest."

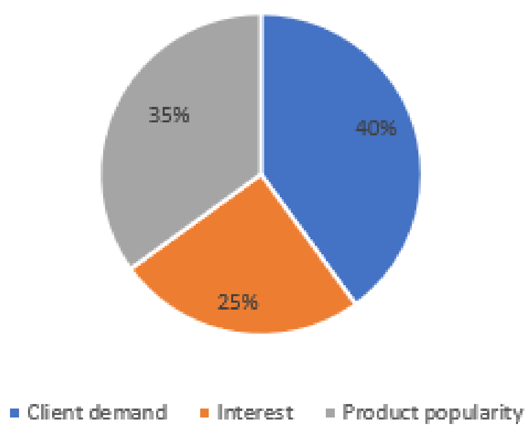


Figure 6: Scaling about big data

5.7 "On a Scale of 1-5, to What Extent the Rise of Social Media and Weblogs have Led to the Rise of Big Data?"

5.7.1 Interpretation. In question 7, we asked our respondents on a scale of 1-5, to what extent the rise of social media and weblogs has led to the rise of big data? While the Likert scale for this question was 1 to 5. As Figure 7 shows that almost 44% of people rated it "1" and suggested the idea

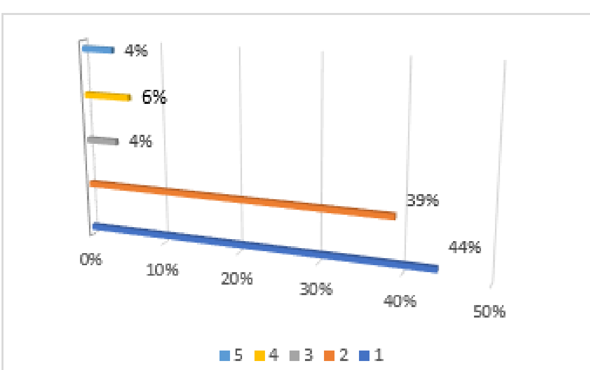


Figure 7: Scaling about big data

to a very lesser extent, while 39% rated it "2," i.e., More Likely to, and suggested the idea that the rise of social media and weblogs have led to the rise of big data, while 4% had biased views and the "3," i.e., neutral, 6% agreed with the idea and remaining 4% opted "5", meaning they totally agreed with the idea. Also, the total number of project managers that answered our question was 16, while this question was asked from all 20 managers, which means 4 people skipped it intentionally. In a nutshell, our analysis shows that most rated their response as "1" and least rated this idea as "5", which is 44% and 4%, respectively.

5.8 "Can Business Analytics be Evaluated Using Relational Database Management Technologies? If yes, which is More Valid? (Big Data, Test Analytics, Web Analytics, Network Analytics, or Mobile Analytics)"

5.8.1 Interpretation. In question 8, we asked our respondents whether business analytics can be evaluated using relational database management technologies. If yes, which is more valid? (big data, test analytics, web analytics, network analytics, or mobile analytics) while the Likert scale for this question was 1 to 5. As Figure 8 shows that almost 40% of people chose "big vata" and suggested the idea that yes, big data is more valid, while 25% chose "web analytics" and suggested the idea that it can be relatively more valid than other sources, while 20% chose test analytics, 10% chose "mobile analytics," and the least of all was considered to be "network analytics". Also, the total number of project managers that answered our question was 14, while this question was asked from all 20 managers, which means 6 people skipped it intentionally. In a nutshell, our analysis shows that most managers considered "big data" to be valid while least considered "networking analytics" as valid.

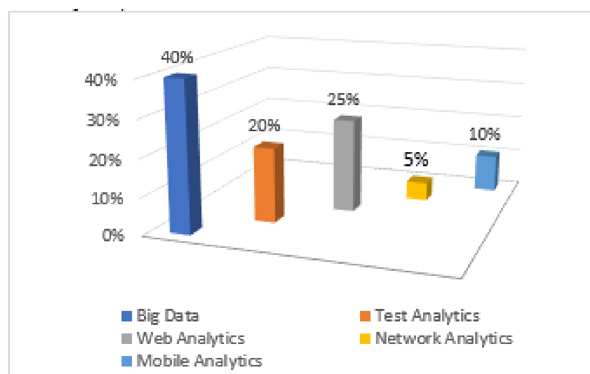


Figure 8: Scaling about big data

5.9 "Can we Give Purpose to Data and Get Value-Driven Answers that will Increase Their Performance?"

5.9.1 Interpretation. In question 9, we asked our respondents can we give purpose to data and get value-driven answers that would increase their performance? We asked our respondents to rate their opinion on a scale of 1 to 5 (Yes, more likely to, Neutral, Less likely to, and Not at all). As Figure 9 shows that almost 30% of project managers rated it as 1 and suggested the idea that Yes, we can give purpose to data and get value-driven answers that will increase their performance, while 25% rated it "2", and suggested the idea that it is more likely to happen, while 20% had biased views and they opted the option 3 "neutral", 10% disagreed with the idea and opted "4," and 15% chose "5" as they didn't agree with the idea at all. Also, the total number of project managers that answered our question was 15, while this question was asked from all 20 managers, which means 5 people skipped it intentionally. In a nutshell, our analysis shows that most rated their response as "1" and least rated this idea as "4".

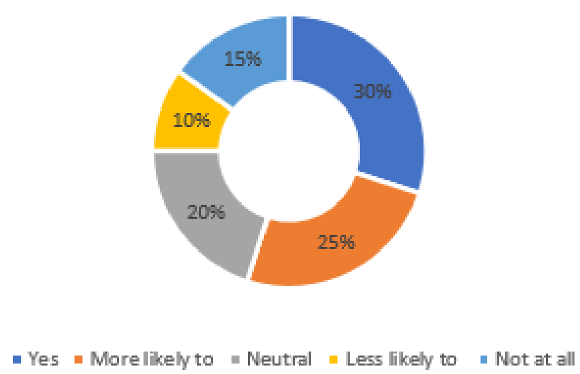


Figure 9: Scaling about big data

5.10 "Can We Expect Opportunities for Economic Progress and a Better Standard of Living for the Society as a Whole Through Big data?"

5.10.1 Interpretation. In question 10, we asked our respondents can we expect opportunities for economic progress and a better standard of living for society as a whole through big data? And we asked our respondents to rate their opinion on a scale of 1 to 5. As Figure 10 shows that almost 30% of project managers rated it as "Yes" and suggested the idea

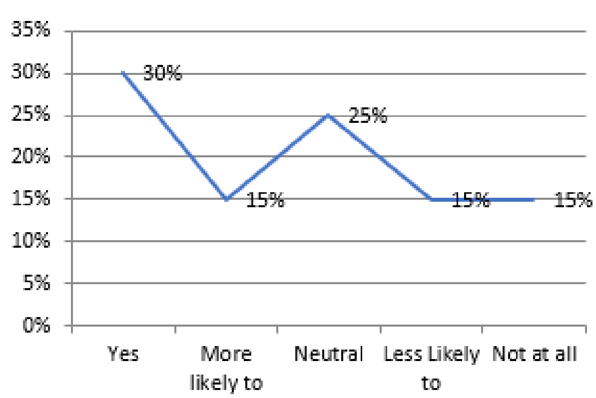


Figure 10: Scaling about big data

that yes, we can expect opportunities for economic progress and a better standard of living for society as a whole through big data, while 15% opted more "More Likely to", 25% had biased views, and they opted the option 3 "Neutral", 15% opted "Less likely to", while 15% opted "No, not at all". Also, the total number of project managers that answered our question was 15, while this question was asked from all 20 managers, which means 5 people skipped it intentionally. In a nutshell, our analysis shows that most rated their response as "Yes" and least rated this idea as "Not at all."

6 Conclusion

Working with data is one of the most crucial subjects for today's company leaders. Data, in particular, has the power to change project management practices, according to nearly nine out of ten financial professionals. Big data and analytics are among the top ten company priorities for slightly more than half of CEOs, according to recent study. All of this indicates that, after the recent IT revolution, project management is moving toward a data-driven future. Big data might have a favorable impact on how the project management office performs its tasks. Big data has a big influence on project and resource management, from cutting project costs to enhancing project efficiency to resource allocation optimization. As this report was specifically made to analyze the effect of business analysis and big data on project management, we conducted structured interviews from our project managers and came to know that big data software is typically identified by its ability to store, access, and manage datasets in a reasonable length of time. It's vital to keep in mind that big data tools can potentially gather and generate incomplete data. Different questions were asked from project managers, and they answered us and helped us to comprehend business, big data and its significance in project management.

Below mentioned impacts have been analyzed in this project.

- **Helps to reduce the complexities of projects**
  Improper methodologies, a lack of expertise, or a misinterpretation of the information provided on crucial project areas can all contribute to the project's complexity. Big data analysis may help to improve the projects by helping us to rapidly discover project challenges and problems, which can help to reduce the project complexity.
- **Helps to reduce the cost of project**
  Project managers will be able to better foresee patterns and occurrences in their business by acquiring more data. Big data analysis is the process of gathering a large quantity of data in order to foresee events and trends in one's sector. This improves the efficiency of the resource forecast and planning process by providing a library of data that is required for a successful budget, calendar, assessments, and other components of project delivery.
- **Helps to enhance the project risk management**
  Big data provides important information on a company's

operations, goods, and services. Engines and other project difficulties can be identified and fixed promptly, allowing each project and the company to perform at its best. Because project management is dynamic and influenced by a number of internal and external circumstances, it is susceptible to a variety of risks that might jeopardize your success. Project management risks must be actively and continuously identified and controlled, as all-risk occurrences are documented and fire-fighting efforts are recognized and addressed. Big data analytics allows for a more accurate evaluation of project challenges and risks, as well as a decrease in their influence on operations and outcomes.

6.1 Concluding Thoughts

Big data is just as savvy as the tools you use to analyze it. Big data offers several chances to improve the competencies of your team and the project management process. It enables users to find information that effects the outcomes of their initiatives, independent of their intended aim. As a thriving profession, project management is entrusted with making technology developments relevant and original. Big data analytics is a technology that may be utilized for both business and project management. Big data is used to assess previous and current information, as well as future data, to anticipate the possibility that your project will provide results, as well as to make data-driven decisions and increase efficiency.

7 Recommendations

Handling with data is one of the most crucial subjects for today's company leaders. Data has the capacity to change organizational practices, according to nearly nine out of ten financial experts. Big data and analytics are among the top ten company priorities for slightly more than half of CEOs. All of this shows that, following the recent IT revolution, industry is entering a data-driven era.

- The implementation of a big data management system is a costly undertaking that, if done incorrectly, may result in significant losses. It's better not to overlook this factor.
- Again, for informed resource allocation, planning and calculating at least a preliminary scale of activities with big data is required to obtain the desired results.
- Organizations are changing in the digitization and Industry 4.0 period. Companies aiming to compete must adapt to a changing world by adopting activities to satisfy changing client and market expectations. Project managers should accept change as the global trend shifts. Companies all across the world are undergoing a process known as "digital transformation."

References

[1] Ritu Agarwal and Vasant Dhar, "Editorial—Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research," *Information Systems Research*, 25(3):443-448, 2014, https://doi.org/10.1287/isre.2014.0546.

[2] I. A. Ajah and H. F. Nweke, "Big Data and Business Analytics: Trends, Platforms, Success Factors, and Applications," *Big Data and Cognitive Computing*, 3(2):32, 2019.

[3] J. Bizon-Górecka and J. Górecki, "Risk of Construction Investment Project in Perspective of Construction Cost," Organization Review, pp. 6:36-44, 2013.

[4] J. O. Chan, "A Conceptual Model for Operations-Analytics Convergence," *The Journal of American Business Review*, 1(1):223-229, 2012.

[5] Hsinchun Chen, Roger H. Chiang, and Veda C. Story, "Business Intelligence and Analytics: From Big Data to Big Impact" *MIS Quarterly*, 1165-1188, 2012.

[6] H. Demirkan and B. Dal, "Big Data, Big Opportunities, Big Decisions," Harvard Business Review Turkish Edition, (Published in Turkish), pp. 28-30, March 2014.

[7] P. Drucker, Innovation and Entrepreneurship, Routledge, New York, 2014.

[8] Michael Fitzgerald, "Training the Next Generation of Business Analytics Professionals," MIT Sloan Management Review, 2014.

[9] C. Hemlata and P. Gulia, "Big Data Analytics," *Research Journal of Computer and Information Technology Sciences*, 4(2):1-4, 2016.

[10] P. Henderson, "Why Large IT Projects Fail," *ACM Trans. Program. Lang. Syst.,* 15(5):795-825, 2006.

[11] A. J. Hoffman, "The Next Phase of Business Sustainability," *Stanford Social Innovation Review*, 16(2):34-39, 2018.

[12] D. Kiron, R. B. Ferguson, and P. K. Prentice, "From Value to Vision: Reimagining the Possible with Data Analytics," *MIT Sloan Management Review*, 54(3):1, 2013.

[13] S. LaValle, E. Lesser, R. Shockley, M. S. Hopkins, and N. Kruschwitz, "Big Data, Analytics and the Path from Insights to Value," *MIT Sloan Management Review*, 52(2):21-32, 2010.

[14] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. Hung Byers, *Big Data: The Next Frontier for Innovation, Competition, and Productivity*, McKinsey Global Institute, 2011.

[15] A. McAfee, E. Brynjolfsson, T. H. Davenport, D. J. Patil, and D. Barton, "Big Data: The Management Revolution," *Harvard Business Review*, 90(10):60-68, 2012.

[16] J. F. Moore, *The Death of Competition: Leadership and Strategy in the Age of Business Ecosystems*, New York: Wiley Harper Business, 1996.

[17] A. Ng, *How Artificial Intelligence and Data Add Value to Businesses*, McKinsey Global Institute, NY, 2018.

[18] A. R. Nikumbh and S. S. Pimplikar, "Role of Project Management Consultancy in Construction Project," *IOSR Journal of Mechanical and Civil Engineering*, 10(6):14-

19, 2014.

[19] P. Patanakul, "Project Manager Assignment: A Strategic Perspective," *Open Economics and Management Journal*, 2(1):21-28, 2015.

[20] S. Ransbotham, D. Kiron, and P. K. Prentice, "Minding the Analytics Gap," *MIT Sloan Management Rev*, 56(3):63, 2015.

[21] J. Su, *Enabling Risk Analytics for Project Success*, Ontario: Deloitte and Touche, 2013.

[22] S. Thomke and E. Von Hippel, "Customers as Innovators: A New Way to Create Value," *Harvard Business Review*, pp. 74-85, April 2005.

[23] R. Vidgen, S. Shaw, and D. B. Grant "Management Challenges in Creating Value from Business Analytics," *European Journal of Operational Research*, 261(2):626-639, 2017.

[24] H. J. Watson, " Tutorial: Big Data Analytics: Concepts, Technologies, and Applications," *Communications of the Association for Information Systems*, 34(1):65, 2014.

[25] J. Weibl and T. Hess, "Success or Failure of Big Data: Insights of Managerial Challenges from a Technology Assimilation Perspective," *Proceedings of the Multikonferenz Wirtschaftsinformatik (MKWI)*, pp. 12-59, 2018.

[26] M. J. Wills, "Decisions Through Data: Analytics in Healthcare," *Journal of Healthcare Management*, 59(4):254-262, 2014.

Mounir El khatib has a long and varied experience in Digital Government and Digital Transformation in the UK, Middle East and UAE. He has worked as a practitioner and an academic in focused areas like Project and Program Management, Information Systems, Strategy and Operations, Business Process and Service Management, providing teaching, training, consulting, advising, research and professional services

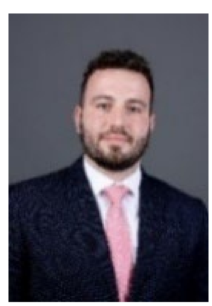
Raed Abu Zitar (BS 88, MS 90, and PhD 93) is a senior research scientist in the context of Thales Endowed Chair of Excellence, Sorbonne Center of Artificial Intelligence (SCAI), Sorbonne University - Abu Dhabi (SUAD), UAE. Prof. Abu Zitar has joined SCAI-SUAD in Feb-2021. Prof. Abu Zitar is directing the research at Thales/SCAI in the fields of Intelligent detection and tracking of drones in cooperation with industrial experts from Thales and Aveillant (France and the UK). Prior to that date, he was the founding coordinator for the Master of Artificial Intelligence Program (one of the first in UAE), Head of the Center of Artificial Intelligence, and Manager for the Teaching and Learning Center at Ajman University, UAE. He has more than 30 years of experience in teaching, research, and administration in the USA, Jordan, and the UAE. He has published more than 100 Journal and conference papers.

Khadeeja Al Naqbi (photo not available)is holding a master degree in project management from Hamdan bin Mohammed Smart University in Dubai, United Arab Emirates.

Wesal Al Naqbi (photo not available) is holding a master degree in project management from Hamdan bin Mohammed Smart University in Dubai, United Arab Emirates.

Shaima AlHarmoodi (photo not available) is an Assistant Professor at Hamdan bin Mohamed Smart University. She has a PhD in the field of project management with a focus on innovation projects. She is also an innovation and entrepreneurship ecosystem developer.

She also actively runs workshops, masterclasses, and seminars in areas related to innovation, entrepreneurship, and project management in addition to her interest in research that resulted in a number of publications in international journals and conferences.



Ali Baydoun is a Medical Student completing his studies at St. George's University. He graduated with a B.S in Biology with a minor in Leadership from the University of Minnesota in 2018. During his bachelors, he volunteered at the Adamala/Engelhart lab where he was involved in a number of projects. Currently, he is in the process of collaborating on clinicoeconomic research with distinguished professors in the Middle East in addition to joining an urology research lab at the University of Miami at the Desai-Sethi Urology Institute.

A Simulation Study of Auction Based Pricing Strategies in Grid Computing

Kalim Qureshi* and Noor Alsaeed
Kuwait University, Shadadiya, KUWAIT

Abstract

The optimal utilization of assets for cloud/grid resources is a strategy that may be employed to increase the profit of service providers. In this paper, we empirically investigate the Continuous Double Auction (CDA), the Stable Continuous Double Auction (SCDA), and the Threshold Price Double Auction (TPDA) strategies, as well as the Preston McAfee Double Auction (PMDA) strategies to demonstrate how the use of grid/cloud computing might have a highly positive impact on the allocation of computing resources. The CDA, SCDA, PMDA, and TPDA strategies are analyzed in terms of their suitability for their end-users while taking into account; prospective resources, deadline consumption, budget spending, resource-derived profits, and immediate resource allocation. According to the measured results, SCDA performs efficient budget spending, CDA is good in resource allocation, with TPDA and PMDA demonstrating the highest performance with respect to deadline consumption.

Keywords: Cloud/grid computing; market-based computing resource economic management; CDA; SCDA; PMDA; TPDA.

1 Introduction

Better resource utilization is equally important for fog, cloud, and grid computing applications [8]. The terms 'cloud' and 'grid' are used interchangeably in this paper. Over the past several years, the consumption of information technology resources by the general public and businesses has increased exponentially. The advancement of service-oriented computing, defined as cloud and grid computing systems, has been led by utility computing, SaaS (Software as a Service), as well as the convergence of the cloud. The central instigator behind this progress is the external deployment of computing power, storage, or applications as services on a self-appeal basis [5]. The long-term ambition of utility computing has been realized [4, 7] as a consequence of the grid/cloud paradigm. The contributions of several companies, namely, Amazon, Microsoft, and IBM, allow consumers to utilize resources and amenities in a pay-as-you-go manner. The prospect of a positive net revenue via leveraging their available data center resources to serve conceivably thousands of consumers is one of the fundamental motivators for Infrastructure as a Service (IaaS).Cloud computing providers aim to maximize the number of accepted new requests to increase profit; however, the Quality of Service (QoS) per the agreed-upon Service Level Agreement (SLA) must not be compromised. The establishment of some efficient resource management is vital to achieving this goal [9]. Resource management is a fundamental challenge in grid/cloud computing systems. It has inspired researchers to identify features that facilitate economizing the grid/cloud environment, with one domain exemplar of such attempts being the ones applied in finance, trading, and pricing. Defining a computational market, in which grid providers and users interrelate, is compulsory in the economics-based models of grid resource allocation. Establishing price levels is an integral factor of the interaction and it must be concurrent with the client's assessment of the commodities traded [14]. This category of mechanisms is based on brokerage policies and trading between the resource owners (resource providers) and users (service consumers).

Several articles of research conducted on economic-based resource management systems have identified distributed resource management challenges as well as the requirements of economy-based grid systems. Moreover, they have deliberated on many diversified representative economy-based systems, emergent or otherwise, for both sympathetic and antagonistic trading of assets, including CPU cycles, storage, and bandwidth. The following emphasizes the reasons why market-based allotment mechanisms are preferable for grid resource allocation [2]:

- Market-based allocation systems eradicate the necessity of a central control point and accommodate the decentralized constitution of a computational grid. All users and providers are involved in the decision-making process. Self-interested participants may enact effective decisions once associated trading rules and an exchange protocol have been established.
- The market's introduction promotes the informed use of resources and attempts to encourage general users' decisions that would maximize overall value. Markets form a competitive environment that stabilizes any conflicts of interest between the parties.
- High prices incentivize providers to offer resources; however, users may withdraw from using them.
- Low prices may bankrupt the providers but attracts users.
- Complex combinatorial resource requests may be conducted in a market environment. Users can designedly

* Department of Information Science, College of Life Sciences, Email: kalimuddin.qureshi@ku.edu.kw and noor.alsaeed@ku.edu.kw.

obtain assets in a market whenever they require, provided that these resources have been sufficiently paid.

Market-based allocation mechanisms are attractive for grid resource allocation. Markets constitute a competitive environment that naturally balances the conflicts of interest between parties. We studied the auction models for resource management using the GridSim simulator. We compared CDA, SCDA, PMDA, and TPDA strategies using deadline consumption, budget spend resource profit, and immediate resource allocation.

This paper is organized as follows: In Section 2 we discuss pertinent studies: The Auction Allocation Model and the model participant's roles are explained in Section 3; Section 4 describes the CDA, SCDA, PMDA, and TPDA strategies. In Section 5 we demonstrate the experimental setup and measured results. We reach a conclusion for this paper in Section 6.

2 Related Work

"A taxonomy is proposed to characterize and categorize various market-based distributed systems (RMSs) that can support utility-driven cluster computing in practice. The taxonomy emphasizes five different perspectives: (i) the Market Model, (ii) the Resource Model, (iii) the Job Model, (iv) the Resource Allocation Model, and (v) the Evaluation Model. A survey was also conducted wherein the taxonomy is mapped to selected market-based RMSs designed for both cluster and other computing platforms" [13]. "Economic models for cloud service markets framework have been developed based on inter-organizational economic models for pricing cloud network services when several cloud providers co-exist in a market, servicing a single application type" [12]. "The proposed work in [10] investigates three types of auction allocation protocols: (i) *First-Price Auction*, (ii) *Vickrey Auction,* and (iii) *Double Auction*. The goal was to find which one is best suitable for the grid environment from the users' perspective as well as from the resources' perspective. They studied these protocols in terms of economic efficiency and system performance. The results showed that the First-Price Auction is better from the resource's perspective while the Vickrey Auction is more suitable from the user's perspective" [1]. "This implements a dynamic pricing model by using advanced reservations as a substrate resource allocation model. Unfortunately, there is one aspect that these works did not consider: as the price is set at the beginning of the execution, applications coming into the system in a low-demand period will be charged a low price. While urgent applications arriving later, may not get all the resources they require for their execution" [6].

In this paper, we empirically evaluate the performance of four double auction-based pricing strategies in a grid/cloud environment. We evaluate which strategy is suitable in terms of user perspective and which one is suitable from a resource perspective. For this purpose, different parameters are used, such as the consumed budget by a certain user, the deadline for each user to complete a job, how much profit is gained by a certain resource provider, and how fast a resource is allocated to a user.

3 Auction Allocation Model

Three parties constitute our market model: resource providers, users, and the market itself. Providers sell resources according to a selling strategy, whilst consumers buy resources following a buying strategy, alongside which, the market determines an appropriate price. Below, we explain each role in terms of the model. The players' interaction in the auction-based model in terms of resource allocation is shown in Figure 1.

3.1 Users

Users express their willingness to buy a resource from a resource provider to do their jobs. Each user has a limited budget, which is assigned to tasks, and each user has a limited time or deadline [5]. The user is also accountable for: bid submission to the auctioneer present at the local market, collecting results, sending user jobs to resources, and providing a uniform view of grid resources to the user. We infer that there are 'Ui' users in our model, all of whom have an executable task ready for submission. The two types of users are defined by their affinity to risk. The first category, risk-averse users, are expected to be in a winner's curse situation, in which the winner pays a surplus for an item more than its actual value. On the other hand, risk-neutral users are less expected to be in this situation.

3.2 Resource Provider

Providers host a configurable number of resources with different processing rates; they contribute to the grid with their resources, charging users for the services from which they benefit [10]. An internal auctioneer prepares every resource placed in the auction process. The service providers publish their resource requests (ask) in the central market. In our model, we assume that we have 'RPi' resource providers that have an 'N' number of resources. A resource (Ri) is characterized by:

- Reserved price: price reserved for auction participation
- Processing rate: in MIPS
- Cost: the cost incurred by a resource provider to execute a job

3.3 Central Market for Auction

The central market contains complete information about every consumer and resource provider's current auction offer. It also provides a set of external auctioneers (EA), which will be responsible for running two-sided auctions (e.g., continuous double auctions). The central market takes offers (bids and asks) from consumers and providers, decides on the winning bid and ask, and then informs both the consumer and provider about the price.

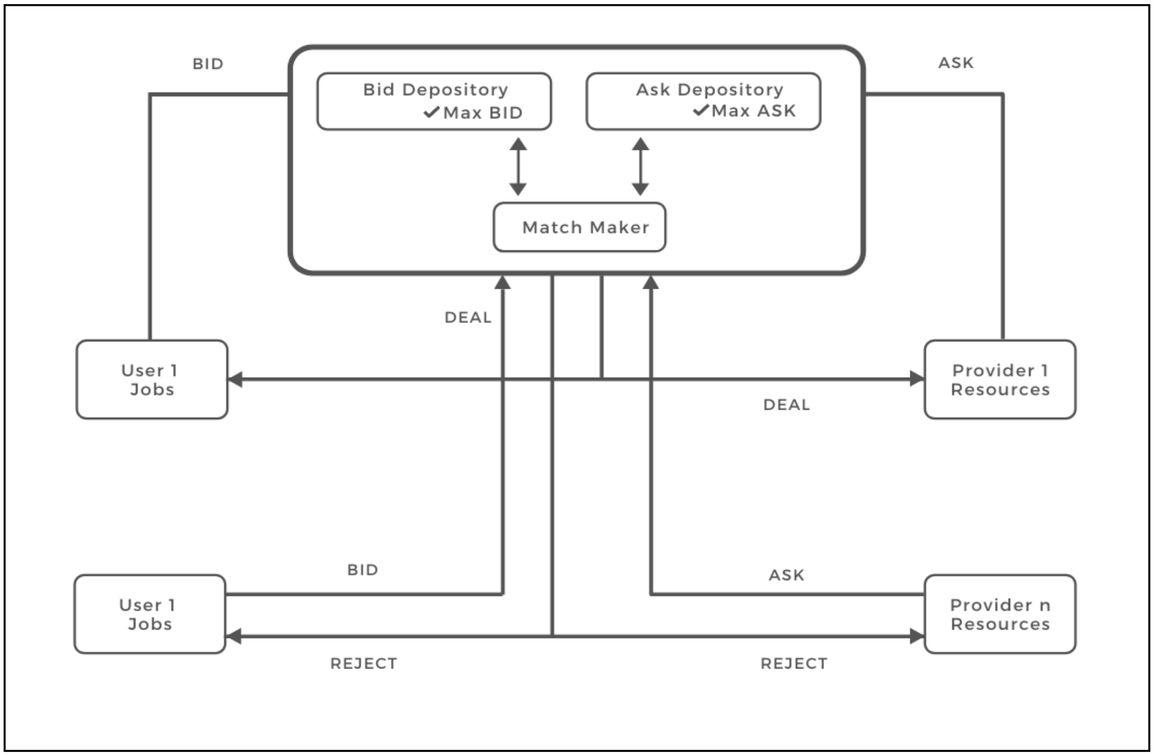


Figure 1: Player's interaction in auction-based model for resource allocation

4 Auction-based Resource Allocation Strategies

In this section, we will describe the CDA, SCDA, PMDA, and TPDA strategies.

4.1 Continuous Double Auction (CDA) Strategy [1]

The most popular form of the CDA strategy is open-cry with an order queue. In this strategy the bids and asks prices that are unsuccessful stay open until they are accepted in a transaction or are changed by their owner(s). There are no trading phases in the CDA, and Grid Services Providers (GSPs), as well as users, may yield asks prices and bids to the EA. The EA organizes lists of the current ask price and bids and matches the two offers when the ask price is lower or equal to the highest bid. At the average of the matching bid and asks prices, the trade occurs. The EA executes this strategy, which is part of local markets for auctions (LMA). The EA posts the auction description on LMA [1] once a set of GSPs decides to engage in a CDA. The pseudo-code of the CDA strategy is shown in Figure 2.

4.2 Stable Continuous Double Auction (SCDA) Strategy [13]

This strategy is designed to reduce unnecessary price volatility contributed by the insensitive or impatient behavior of the bidders [14]. To construct an SCDA, a Compulsory Bidding Adjustment Layer (CBAL) is added around a CDA. All orders first have to go through the CBAL before reaching the CDA. A heuristic mechanism is used to identify unreasonable orders for the current market conditions in the CBAL and are then adjusted appropriately before submission to the core CDA. There are two intuitions that the CBAL is based on (1) the Kaplan strategy that it is a very good time to trade when the ask-bid spread is small, and (2) the reference price r (the median value of the history prices Hl) which is an important indicator of the market condition notably so after taking r, a min, and b max into account [xx]. Intuition (1) reflects changes in market conditions more efficiently, while (2) and stabilizes the market [2]. The pseudo-code of the SCDA Strategy is shown in Figure 3.

4.3 Preston McAfee Double Auction (PMDA) Strategy [9]

In this type of auction, the users submit bids and GSPs submit asks to an EA. The equilibrium price is determined by matching asks (starting from the lowest price to the highest) with demand bids (starting from the highest price to the lowest). Once a set of GSPs decided to participate in a double auction, EA posts the auction description on LMA. We assume that only GSPs having resources of the same type participate in one double auction strategy. The pseudo-code of the CDA strategy is shown in Figure 4.

4.4 Threshold Price Double Auction Strategy [11]

The possibility of cheating by false-name bids is accounted

Three players of CDA interact with each other

1. User Side:

Ui, i=1,2,3.. M send the bid bi and mi to Central Auctioneer (CA)

2. Resource provider side:

RPn, n= 1,2, 3 ..k send asks price ak to CA

3. Central Auctioneers (CA):

After receiving bids and asks price

Sorts bids in ascending order

Sorts asks price in descending order

Compares incoming bid with minimum ask price

Compares incoming ask with maximum bid

Satisfies conditions

Sends winner bid and ask price with final price to broker

Figure 2: Pseudocode of CDA strategy

Three players of SCDA interact with each other

1. User Side:

Ui , $i = 1,2,3....,M$ *sends bid* $\rightarrow$ *CBAL*

2. Resource provider side:

RP^n, $n = 1,2,3....k$ *sends ask* $a_k \rightarrow$ *CBAL*

3. CBAL Side:

Receives Bid and asks

Sorts bids ascending order

Sorts asks descending order

Uses mundane fuzzy rules to stabilize the price

Sends modify bid and ask $\rightarrow$ *CA*

Compares incoming ask with maximum bid

Satisfies conditions

Sends winner bid and ask with final price to broker

Figure 3: Pseudocode of SCDA strategy

1. ***User Side:***

U_i , i = 1,2,3….,M sends bid → CBAL

2. ***Resource provider side:***

RP^n , n= 1, 2, ..k sends ask a_k → CBAL

3. ***Central Auctioneers (CA):***

After receiving bids and asks

Stores bids in ascending order

Stores asks in descending order

Receives the asks and bids for specific time interval once the interval clear

Finds trade price

Satisfies condition on the basis of trade price

Sends winner bid and ask with final price to broker

Figure 4: Pseudocode of PMDA strategy

Three players of CDA interact with each other

1. User Side:

U_i , i = 1,2,3….,M sends bid → CBAL

2. Resource provider side:

RP^n , n = 1,2,3....k sends ask a_k → CBAL

3. Central Auctioneers (CA):

After receiving bids and asks

Stores bids in ascending order

Stores asks in descending order

Finds the Threshold price

Satisfies condition on the basis of Threshold price

Sends winner bid and ask with final price to broker

Figure 5: Pseudocode of TPDA Strategy *1 User Side*

for in this strategy. A GSP may try to cheat by pretending to be a user (a potential buyer) and submitting a false-named bid. To counteract. this, the TPDA Strategy uses a threshold price that is determined by the EA without knowing the evaluations of GSPs and users. The number of trades and the trading price will be controlled by this threshold price. As in PMDA, the GSPs submit asks to an EA and the users submit bids. Thus, the trading price is determined by matching asks (in ascending order) with demand bids (in descending order) considering the threshold price. This strategy is performed by the EA (part of LMA). The pseudo-code of this strategy is shown in Figure 5.

5 Results and Discussion

Many experiments were conducted to evaluate the performances of CDA, SCDA, TPDA, and PMDA strategies, which consisted of the interactivity of three agents: providers, consumers, as well as the market itself. At this stage, an auctioneer from the local market becomes the central agent who carries out all experiments, while consumers and providers liaise with the auctioneer agent through their asks and bids respectively.

5.1 Experiment Environment and Setup

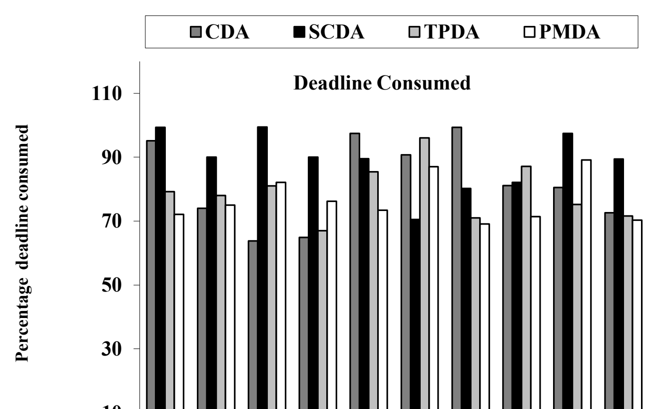
The simulated grid environments consider 16 resources {R0, R1, …, R15} divided into two categories: slow resources (Intel core i3 3.4 GHz – 4.2 GHz) and fast resources (Intel core i52.4 GHz -3.8GHz). Slow resources have lower reservation prices than fast resources. We will use the GridSim toolkit [9] also, regarding its resource allocation. The simulation also consists of 10 users {User 0, User 1, …, User 9}, whom have a number of computational jobs that are achieved to get the resource. Each user wants to execute its jobs in terms of its preferences. There are 30 jobs simulated in the grid system. Each user has a maximum of 3 jobs. Each user has a budget and deadline which are distributed in these jobs according to their preferences.

5.1.1 Deadline Consumed. We are referring here to the deadline consumed by each user who has a job; each job has its deadline for allocating a resource. In Figure 6, we can see that stable continuous double auctions have a much longer deadline than other strategies due to the higher number of extra components within the bidder, asker, and central auctioneer. This strategy establishes the market price but is more time-consuming in terms of completing a job. It has shown that a continuous double auction had a lower time consumption than a stable continuous double auction.

5.1.2 Budget Spent. In terms of budget, every user has a certain budget that is attached to their own job depending on their priority. The aim of this study, for each of the four auction strategies, is to determine which auction strategy is more suitable for the user in terms of budget. The auction strategy defines the budget spent by a winner user for a resource. Figure *7* shows the budget spent by the users in each of the four auction strategies. In Preston McAfee, the spent budget by a certain user is variable but higher than the budget spent in other strategies as users are sending random sealed bids for resources while in Stable Continuous Double Auction (SCDA) budget spent by users is more stable than in other strategies.

In CDA, the budget spent by users is lower than in PMDA because of its open double auction and the price paid by the user is the median value of the winning bid and ask. The budget spent by users in threshold price double auction is higher than the stable continuous double auction and continuous double auction due to sealed random bid but its budget is lower than Preston McAfee Double Auction strategy due to threshold price.

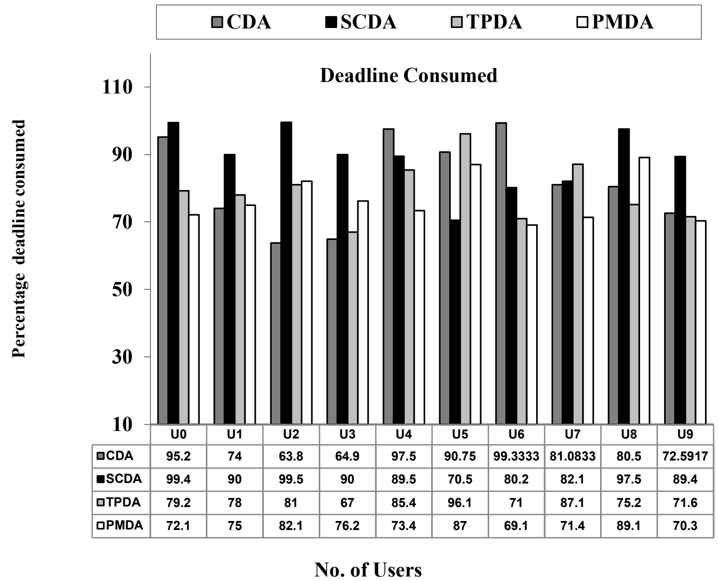
5.1.3 Resource Profit. It is the difference between the payment received from the winning users and the cost of the resource assigned to it. Figure 8, demonstrates that resource profit in Preston McAfee double auction varied from resource to resource, some resources have higher resource profit and some of them have a lower profit. In a Stable Continuous Double Auction, the resource profit for every resource from R0 – R11 is consistent. In Preston Continuous Double Auction, the resource profit is lower than Preston McAfee Double Auction because of using median value trade price of bids and asks are sent. Threshold Price Double Auction has variability in the budget as compared with other strategies due to threshold price



	U0	U1	U2	U3	U4	U5	U6	U7	U8	U9
CDA	95.2	74	63.8	64.9	97.5	90.75	99.3333	81.0833	80.5	72.5917
SCDA	99.4	90	99.5	90	89.5	70.5	80.2	82.1	97.5	89.4
TPDA	79.2	78	81	67	85.4	96.1	71	87.1	75.2	71.6
PMDA	72.1	75	82.1	76.2	73.4	87	69.1	71.4	89.1	70.3

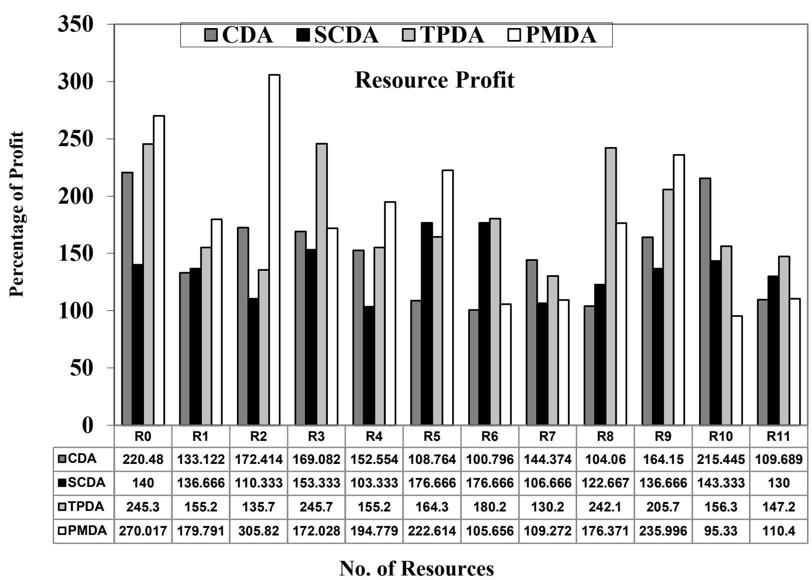
No. of Users

Figure 6: Percentage of deadline consumed by users



	U0	U1	U2	U3	U4	U5	U6	U7	U8	U9
CDA	95.2	74	63.8	64.9	97.5	90.75	99.3333	81.0833	80.5	72.5917
SCDA	99.4	90	99.5	90	89.5	70.5	80.2	82.1	97.5	89.4
TPDA	79.2	78	81	67	85.4	96.1	71	87.1	75.2	71.6
PMDA	72.1	75	82.1	76.2	73.4	87	69.1	71.4	89.1	70.3

Figure7: Budget spends by each user



	R0	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11
CDA	220.48	133.122	172.414	169.082	152.554	108.764	100.796	144.374	104.06	164.15	215.445	109.689
SCDA	140	136.666	110.333	153.333	103.333	176.666	176.666	106.666	122.667	136.666	143.333	130
TPDA	245.3	155.2	135.7	245.7	155.2	164.3	180.2	130.2	242.1	205.7	156.3	147.2
PMDA	270.017	179.791	305.82	172.028	194.779	222.614	105.656	109.272	176.371	235.996	95.33	110.4

Figure 8: Percentage of resource profit

because trading is done based on threshold price.

5.1.4 Immediate Resource Allocation. Figure 5 shows that CDA has a higher rate of resource access than other strategies due to continuously sending bids for resources. Here, users have more chances to complete their jobs. Users 0, 3, 5, 8, and 9 have 50 % of resources and the other users have a higher rate of resource allocation. The SCDA also has a higher rate of resource allocation due to its continuously sending bids and asks the central auctioneer, but resource allocation is lower than

CDA because of performing the extra activity to stabilize the price. In TPDA and PMDA resources allocated by users are lower because of their discrete time. The users' bids and resources ask will stay there for a specific time interval. Figure 9 shows that both TDPA and PMDA have a lower rate of resource allocation. Users 6, 7, and 8 have a higher rate of resource allocation due to their budget spending and deadline consumption.

6 Conclusion

This paper considered the issues concerning the management of CDA, PMDA, TPDA, and SCDA computing systems in terms of deadline consumption, budget spending, resource profit, and immediate resource allocation. The results of this empirical investigation demonstrate conclusively that whereas the SCDA framework would perform more efficiently in terms of budget spending, the CDA may be seen as highly efficient concerning immediate resource allocation. The efficiency of the CDA may be attributed to such factors as the lack of specific trading phases and the direct transmission of the relevant information to the EA responsible for evaluating the bids in question. On the other hand, the TPDA and PMDA frameworks have demonstrated the highest performance in concordance with deadline consumption.

References

[1] R. Ananthakumar and K. Kartheeban, "Effectual Resource Allocation using Auction Mechanism in Cloud Computing," *International Journal of Innovative Technology and Exploring Engineering*, 9(2S2):2278-3075, December 2019.

[2] Amina Chikhaoui, Kamel Boukhalfa, and Jalil Boukhobza, "A Cost Model for Hybrid Storage Systems in a Cloud Federations," 2018 Federated Conference on Computer Science and Information Systems (FedCSIS), pp. 220-232, 2018.

[3] Se-Hak Chun and Byong-Sam Choi, "Service Models and Pricing Schemes for Cloud Computing," Journal of *Cluster Computing,* 17:529-535, 2014.

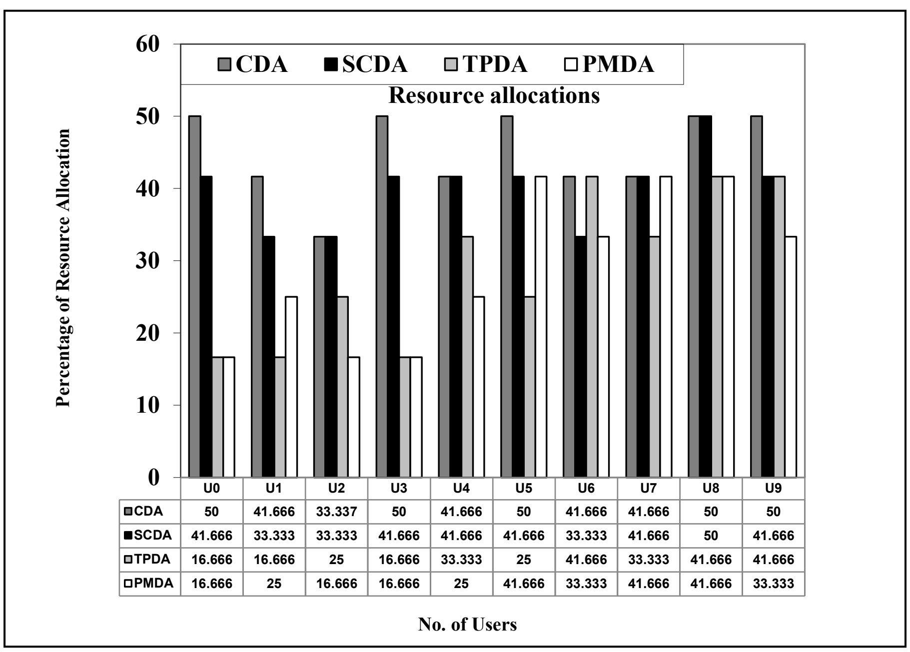
[4] Thibaud Ecarot, Djamal Zeghlache, and Cedric Brandily, "Consumer-and-Provider-Oriented Efficient IaaS Resource Allocation," *Parallel and Distributed Processing Symposium Workshops (IPDPSW) 2017 IEEE International*, pp. 77-85, 2017.

[5] Yuan Gao, Yaoxue Zhang, and Yuezhi Zhou, "Performance Analysis and Improvement of Storage Virtualization in an OS-Online System," *Journal of Software Engineering and Applications*, 5(11A):912-922, November 30, 2012.

[6] Sukhpal Singh Gill, Peter Garraghan, Vlado Stankovski, and Guiliano Casale, "Hholistic Resource Management for Sustainable and Reliable Cloud Computing: An Innovative Solution to Global Challenge", *Journal of Systems and Software*, 55:104-129, 2019.

[7] Barka Javad, Peter Bloodsworth, and Raihan Ur Rasool, "CloudMarket Maker: An Automated Dynamic Pricing Marketplace for Cloud Users," *Journal Futire Generation Computer Systems*, Elsevier, 54:52-67 2015.

[8] Kevin Lee, Georg Buss, and Daniel Veit, "A Heuristic Approach for the Allocation of Resources in Large-Scale Computing Infrastructures," *Concurrency and Computation: Practice and Experience*, 28:1527-1540, 2016.



	U0	U1	U2	U3	U4	U5	U6	U7	U8	U9
CDA	50	41.666	33.337	50	41.666	50	41.666	41.666	50	50
SCDA	41.666	33.333	33.333	41.666	41.666	41.666	33.333	41.666	50	41.666
TPDA	16.666	16.666	25	16.666	33.333	25	41.666	33.333	41.666	41.666
PMDA	16.666	25	16.666	16.666	25	41.666	33.333	41.666	41.666	33.333

Figure 9: Immediate resource allocations

[9] Sung-Han Lin, Ranjan Pal, Marco Paolieri, and Leana Golubchik, "Performance Driven Resource Sharing Markets for the Small Cloud," *Distributed Computing Systems (ICDCS) 2017 IEEE 37th International Conference,* pp. 241-251, 2017.

[10] Nguyen Cong Luong, Ping Wang, Dusit Niyato, Yonggang Wen, and Zhu Han, "Resource Management in Cloud Networking Using Economic Analysis and Pricing Models: A Survey," IEEE Communications Surveys & Tutorials, 19(2):954-1001, 2017.

[11] Panel ParniaSamimi, Youness Teimouri, and Muriati Mukhtar, "A Combinatorial Double Auction Resource Allocation Model in Cloud Computing," *Journal of Information Sciences*, 357: 201-216, 2016.

[12] Muhammad Bilal Qureshi, Maryam Mehri Dehnavi, Nasro Min-Allah, Muhammad Shuaib Qureshi, Hameed Hussain, Ilias Rentifis, Nikos Tziritas, Thanasis Loukopoulos, Samee U. Khan, Cheng-Zhong Xu, and Albert Y. Zomaya, "Survey on Grid Resource Allocation Mechanisms," Journal of Grid Computing, Springer, pp. 399-441, 2014.

[13] Xingwei Wang, Xueyi Wang, Hao Che, Keqin Li, Min Huang, and Chengxi Gao, "An Intelligent Economic Approach for Dynamic Resource Allocation in Cloud Services," *Cloud Computing IEEE Transactions on*, 3(3): 275-289, 2015.

[14] Muhammad Zakarya and Lee Gillam, "Modeling Resource Heterogeneities in Cloud Simulation and Quantifying Their Accuracy," Journal of Simulation Modeling Practice and Theory, 94:43-65, 2019.

Kalim Qureshi is an Associate Professor of Information Science Department, Kuwait University, Kuwait. His research interests include network parallel distributed computing, thread programming, concurrent algorithms designing, task scheduling, performance measurement and medical imaging. Dr. Qureshi receive his Ph.D and MS degrees from Muroran Institute of Technology, Hokkaido, Japan in (2000, 1997). He published more than 70 journal papers in reputed journals. His email address: kalimuddin.qureshi@ku.edu.kw.

Noor-Alsaeed (photo not available) received her Bachelor's degree in Information Science Department, Kuwait University, Kuwait in the College of Life Science. She during Master's degree in Computing Information Systems, from the same department.

Implementing IOT in Effective Project Management

Mounir El Khatib*
Hamdan Bin Mohammed Smart University Dubai Academic City, Dubai, United Arab Emirates

Raed Abu Zitar
Sorbonne University Abu Dhabi, United Arab Emirates

Ahmed Alnaqbi†, Abdulla Alnaqbi†, Hamad Alsuwaidi†, Meera Al Marri‡, Ahmed Ankit§
Hamdan Bin Mohammed Smart University Dubai Academic City, Dubai, United Arab Emirates

Abstract

The Internet of Things (IoT) can revolutionize project management by allowing devices to communicate and share data, automate tasks, and improve communication and coordination among team members. However, this comes with challenges, such as data security and privacy, interoperability, and managing the complexities of IoT systems. This research proposes identifying the benefits and shortcomings of using IoT in project management and identifying solutions and best practices for implementing and managing IoT systems in this context. The research was conducted through semi-structured interviews on questionnaires. The research found that IoT can offer significant benefits in project management, including cost reduction, increased efficiency, and improved communication and coordination. However, these benefits can only be realized if data security and privacy challenges, interoperability, and system complexity are addressed through careful planning, design, and management. Key lessons include considering the end-to-end user experience, carefully managing data and devices, and properly testing the system.

Key Words: Internet of things (IoT), project management, technology, implementation.

1 Introduction

The Internet of Things (IoT) is the hook via the internet of computing appliances fixed in daily items, allowing them to dispatch and obtain data (Brous et al., [7]). This technology has the potential to connect everything from cell phones, washing machines, and cars to roadways, healthcare facilities, and power grids. IoT is a transformational technology with the potential to change the way we live, work, and play. It can connect people, devices, and data in impossible ways. IoT can help organizations improve efficiency, optimize resources, and create new revenue streams when correctly leveraged. IoT is already being used in several industries, including manufacturing, healthcare, transportation, and retail. In the coming years, IoT will become increasingly prevalent in our everyday lives (Brous, et al., [7]). As IoT becomes more widespread, organizations must consider leveraging this technology to achieve their business goals. Project managers ensure that IoT projects are appropriately planned, executed, and monitored. Project managers must deeply understand IoT technology and its potential applications (Brous et al., [7]). They must also communicate effectively with stakeholders, sponsors, and team members.

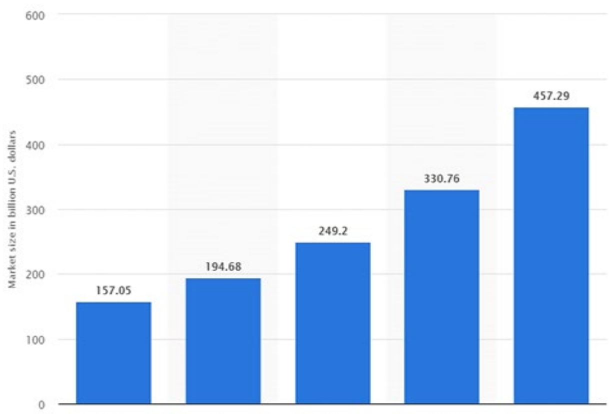


Figure 1: Size of the IoT market worldwide from 2016 to 2020 (in billion U.S. dollars)

1.1 Background Information

Project management is persistently changing to adapt to the more efficient way to control organizations. Although project management has been excellent in recent years, the introduction of IoT by the ever-changing or adapting technology has brought a new breath in management. Therefore, IoT has transformed modern business – project management can utilize the practical application of IoT to achieve overall goals. The IoT has positive ramifications on project management. In the Internet of Things (IoT) and

* Associate Professor, School of Business & Quality Management, P.O.Box 71400, Email: m.elkhatib@hbmsu.ac.ae.

† Graduate Project Management, Email: (200118800@hbmsu.ac.ae; 200118764@hbmsu.ac.ae; 200121128@hbmsu.ac.ae.

‡ Assistant Professor, School of Business & Quality Management, P.O.Box 71400,Email: M.Alreyaysa@hbmsu.ac.ae.

§ Professor, Dean of Scientific research and Doctoral Studies, Email: a.ankit@hbmsu.ac.ae.

shifting the current prospect of Project Management dissertation, PrasherVikram gave a synthetic scope of project management based on the IoT.

The Internet of Things (IoT) is a chain of concrete items ingrained with sensors, programs, and other mechanization that enables them to acquire and exchange information (Durmic, [8]), please see Figure 2. The IoT has the potential to revolutionize daily activities by providing individuals with real-time data that can be used to improve efficiency, safety, and productivity. Project management is the process of monitoring, executing the progress, and planning of a project (Khalafi et al., 18). Project managers use various tools and techniques to plan and execute projects, and they are in charge of ensuring the project is finalized on time, within budget, and within scope (Khalafi et al., 18). The IoT field is rapidly growing, and project managers must stay updated on the latest developments. IoT project management requires a unique set of skills and knowledge, and it is crucial to partner with an experienced IoT service provider who can help you navigate the challenges of this dynamic field.

IoT has been identified as a game-changer in project management, as it can revolutionize how work is carried out and monitored. IoT-enabled project management systems can provide real-time data and insights to help project managers make better decisions, optimize resources, and improve project delivery. IoT can also help project managers better understand work progress and identify potential issues and risks. By connecting devices and sensors to project management software, project managers can receive alerts about problems as they happen and take corrective action to prevent them from becoming more significant issues. In addition, IoT can be used to automate project tasks. For example, IoT-enabled sensors can track the location of assets and equipment and trigger alerts when moving to unauthorized areas; This can help to prevent theft and loss and to ensure that assets are being used as intended (Brous et al., [7]). IoT can also monitor susceptibleness, such as humidity and temperature. This information can be used to adjust the work environment to improve comfort levels and prevent equipment damage.

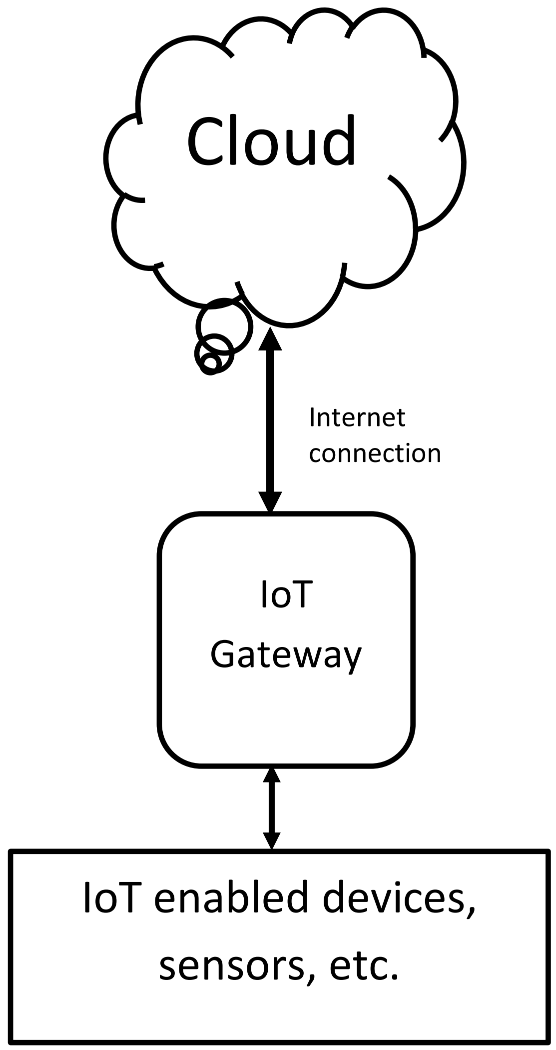


Figure 2: Architecture of IoT with clouds

1.2 Problem Statement

Although there has been tremendous advancement in technology application in business, there is a problem with how IoT is applied in project management. In 2017, a Cisco survey revealed that almost 75% of the project using IoT fail– this shows a significant challenge (Rivett, [29]). With the growing technological advancement, some industries have been scared of incorporating IoT in project management. The failure of IoT in management is making some industries lose confidence in investing in the growing IoT technology. There are several factors contributing to the delinquency of those projects.

One major cause of failure of project failure using IoT is human factors ranging from managerial and technical skills to general culture. The unavailability of universally recognized organizational techniques is causing significant challenges in implementing IoT in project management. Since current projects incorporate many systems, there is the potential for huge risks. The prevailing traditional information technology is straightforward to handle project management across all phases. Therefore, project managers should be integral in researching and developing IoT in significant organizations' projects.

1.3 Hypothesis

- The Internet of Things presents numerous pros and cons that affect the overall effectiveness of project management.
- IoT could help managers make better-informed decisions about their projects by providing real-time data.
- IoT can help automate tasks and processes, making projects more efficient and reducing manual labor.
- Moreover, IoT can improve communication and collaboration between team members, managers, and clients.
- IoT could help track, and monitor project progress, making it easier to identify and resolve issues as they arise. (Deleted, it is not covered in the literature review, questionnaire, or analysis)

- IoT can give managers insights into how their projects are being used and accessed, allowing them to optimize and improve their offerings.

1.4 Research Question

This study purposely explores the function of implanting IoT in the effective management of a project. Additionally, it emphasizes the requisite positive and negative ramifications of implementing IoT in modern project management. Therefore, the research questions will be:

- What is the relationship between IoT and project management, and how to implement IoT technology in effective project management?
- What are the advantages and disadvantages of using IoT in project management?
- How can data from IoT be used to monitor and track project progress effectively?

1.5 Research Objectives

This research aims to effectively develop a framework for managing Internet of Things (IoT) projects. Therefore, this study will focus on the following objectives.

- To understand how IoT can be used in effective project management.
- To investigate the advantages of using IoT in project management.
- To identify the disadvantages of using IoT in project management.
- To explore the feasibility of using IoT in project management in a range of different contexts

2 Literature Review

2.1 The Role of IoT in Changing Project Management

Internet of Things (IoT) and changing face of project management research by PrasherVikram [26] is an excellent article discussing IoT in business management. The study revolves around IoT in varying project management. Prasher [26] explained that regardless of the non-tech or tech, IoT is a huge technological advancement that covers all business domains. In addition, the research pinpoints that IoT is the foundation of the hardware and software combination of components that disrupts traditional perceptions in manufacturing. Therefore, this article aims to explore the role of the Internet of Things in project management's changing face. The article's central idea rests on the importance of self-efficacy when discussing the Internet of Things (IoT), as this can influence many aspects of our lives (Prasher, [26]). For example, a person's self-efficacy may influence their decision to purchase a smart home device or whether they believe they can successfully use it. Self-efficacy may also affect how people interact with and use IoT devices (Prasher, [26]). The author struggles to express the research scope, but this article provides a roadmap toward realizing the goal of my paper.

2.2 Impacts of IoT on an Organization's Project Management

The impact of Internet of Things unification with project management disciplines in project-based organizations research using concepts of project management in IoT - oriented organizations instill the prevalent practices of IoT. Percudani & Batrawi [25] explained that IoT and project management are two close disciplines in a business environment. In addition, the research pinpoints four motivational concepts of IoT in managing projects. Percudani & Batrawi [25] explained that IoT could work independently and make decisions without human intervention. IoT motivational concept does not end there; instead, it is a profound concept that includes connectedness – the ability of devices to connect and share data with other systems. In addition, the research identified that IoT is proactive – the power of machines to take action in anticipation of user needs or changes in the environment. Later, the study narrowed down the capability of IoT as context-awareness – the ability of devices to understand and respond to the user's current situation and context. For instance, IoT can monitor the progress of different project segments and give real-time updates to requisite authorities. This research relies on Percudani and Batrawi's multi-facet ideas in IoT and explains the deep-lying concept of IoT in the business project management sector.

The Internet of Things provides numerous advantages and benefits for all sectors. It ushered the next industrial revolution characterized by radical changes, disruptions, and new paradigms focusing on the environment Babun, et al. [6] argued that IoT extends current connections between users and computer devices. IoT can help share and transfer data from simple numbers to complex details shared using multiple data streams and sensors. IoT facilitates machine-to-machine communication, enabling physical devices to remain connected. This aspect increases overall transparency, efficiency, and quality (Babun, et al., [6]). The connected infrastructure of digital devices can be controlled remotely and centrally, increasing the potential for control and ensuring faster and more timely activity completion; this enables individuals and organizations to save money, time, and other resources. IoT increases information access, allowing individuals to enhance their knowledge and make better decisions.

2.3 Concepts of Technology in Project Management

The article discusses the various aspects of IoT project management and how it can help organizations manage their IoT projects. It also highlights the importance of clearly understanding the multiple stakeholders involved in the project and the need for effective communication and coordination among them (Hurtoi & Avadanei, [16]). Moreover, the article emphasizes the importance of setting clear objectives and milestones for the project and regularly monitoring and assessing the project's progress. In addition, the report covers the basics of IoT project management and details the critical components of a successful IoT project (Hurtoi & Avadanei, [16]). It offers tips for managing an IoT project from start to finish, including selecting the right platform and tools, setting up a development environment, managing devices and data, and troubleshooting and debugging issues (Hurtoi & Avadanei, [16]). The research also sheds light on the role of IoT in project management. For example, it can lead to improved project outcomes as managers become more confident in their ability to complete tasks and make decisions. In addition, IOT self-efficacy increases job satisfaction as managers feel more capable and empowered in their roles. IoT's self-efficacy is multi-dimensional; thus, it can lead to better team dynamics as managers are more likely to delegate and trust team members. Lastly, the article pinpoints that IoT can reduce stress and anxiety as managers feel more capable of handling challenges and setbacks. This article is integral to my research since it summarizes the importance of IoT in project management. It sheds light on what project managers should expect when incorporating IoT in a business environment.

2.4 Disadvantages of IoT on an Organization's Project Management

Patel et al. [24] argued that IoT delivers multiple benefits but presents numerous challenges. This study identified security as a top concern as IoT creates an environment of constantly connected devices and continuously communicating over networks. IoT's sophistication allows devices to share a substantial amount of personal data without the owner's consent or participation. Internet systems may not provide adequate control and protection due to the need to upgrade security measures continuously. This aspect exposes users to malware and intrusion that could result in data breaches and substantial losses. IoT uses complex designs and deployment and maintenance techniques that make it increasingly sophisticated. Some individuals may find it challenging to use IoT systems due to the numerous technologies and enabling structures embedded within such tools.

Parteek [23] evaluated IoT's potential drawbacks, including increased unemployment with increased efficiency and automation. The author highlighted how IoT promotes unemployment, including developments in robotics, smart surveillance, automated machines, and artificial intelligence. He argued that the development and maintenance of such systems require substantial finances. IoT is a high-risk system due to the high chance of being corrupted. One factor that maintains this issue is the lack of international standardizations, which makes it challenging to satisfy compatibility requirements, especially for devices from different manufacturers communicating with each other.

3 Methodology

3.1 Research Design

Based on the research method description in Percudani & Batrawi [25], it is an approach that can be divided into two categories – quantitative and qualitative. Qualitative research relies on studying and observing a particular phenomenon before making an interpretation and final judgment on the specific matter they are learning. Scholars can use several qualitative research methodologies to study a phenomenon (Goyal et al., [15]). Some standard qualitative research methodologies include ethnography, case study research, grounded theory, and phenomenology (Goyal et al., [15]). On the other hand, quantitative research methodologies allow for the collection and analysis of numerical data. This type of research is often used in the social sciences and can be used to study various topics. Some of the most common quantitative research methodologies include surveys, experiments, and observation.

In this research, the methodology used to achieve the outlined objectives is quantitative research, namely semi-structured interviews. Semi-structured interviews on questionnaires were used to collect qualitative data concerning the use of IoT in project management. One advantage of semi-structured interviews is that they can be utilized to collect open-ended data. Another benefit of using this methodology is that it is effective in capturing the thoughts and perceptions of participants regarding the topic of interest. The research paper aims at establishing the feasibility of IoT in project management, and this can be determined by collecting feedback from relevant people concerning their experiences in using IoT and their reaction toward its usability. Since the data collected is open-ended, participants are granted flexibility in answering the questions.

3.2 Data Collection and Analysis Tools

The author uses qualitative data collection methodologies to obtain primary data in this research. For instance, in this research, the author used questionnaires to get data from interviews. The participants in answering questionnaires were based on various factors, such as having at least five years of experience and a professional background in project management. During the collection of data, this research used semi-structured. In a semi-structured interview, the interviewer has a general guide of what topics will be covered but allows the conversation to flow somewhat naturally rather than sticking strictly to a predetermined set of questions (Rebelo, et al., [28]). It will enable the interviewer to explore areas of interest that may arise during the conversation while still covering the main topics that were initially planned. The semi-structured approach allows a more in-depth exploration of the interviewee's thoughts and experiences on a given topic (Prasher & Onu, [27]). Therefore, this research used it to dig deeper into understanding the underlying knowledge in this field.

3.3 Research Approach

The study utilized two ways of establishing incorrect and correct data – deduction and induction, as explained in Ghimire et al. [14]. Since the research was more probing and emerging, it often used an inductive approach – which is suitable for this situation, according to Nord, et al. [22]. In addition, the study will uncover data unavailable in the literature review. Later, the research will establish the connection between IoT and project management following Alazzawi & Alotaibi [2] approach. Lastly, the study will incorporate the finding with the existing knowledge and improve the overall strategies.

4 Result Analysis

The IoT research shows that the internet of things can revolutionize how we live and work. It can connect physical objects and devices to the internet, enabling them to liaise with each other and share information. This data can facilitate the efficiency of processes and help make better decisions. In terms of the development of IoT technology, it is mainly based on applying various sensors. IoT presents diverse advantages, including enhancing customer engagement and addressing the current blind spots in organizational analytics. It would help optimize technology to enhance customer experience and improve device use. This aspect would reduce wastage and provide users with real-time information to enhance the management of their resources. Despite the benefits, IoT presents various cons associated with security and privacy concerns. IoT also presents other challenges resulting from its complexity and low flexibility, making it challenging to integrate easily. A multi-layer security approach would help protect users when exposed to diverse degrees of security and privacy threats (Tukur, et al., [33]). A multi-layer approach would secure the entire architecture, ensuring the system is secure. This aspect would help to increase the advantages and minimize the threats from the cons.

4.1 Interview Analysis

The interview was conducted among 20 people; only 75% cooperated throughout the interview and answered all questions. All fifteen respondents agreed that the project is divided into five phases. All respondents agreed that the project reaps enormous benefits when technology is implemented in all stages. In addition, all respondents believed introducing IoT in the project would ease the operation and boost team workings. Additionally, 60% out of those who cooperated agreed IoT still needs to be developed in some organizations and maybe their framework is limiting them from realizing IoT's full potential. In addition, one of three interviewees indicated a change in approaches and paradigms in the initial project phase when the project manager implemented IoT. In this idea, the respondents supported their claim by explaining that IoT gave the project flexibility that enabled it to perfect the outcomes. On top of that, 50% of respondents claimed IoT deeply impacted the project's bidding phase.

In the project planning phase, 73.3% of respondents agreed that IoT enabled efficient and effective communication with multiple clients and contractors. IoT implementation in this phase helped staff and was a guide throughout the project. During planning, the communication between stakeholders and project managers was efficient since the manager gave a real-time update on the progress (Liu, [20]). In the execution phase, 60% of respondents agreed that the implementation of IoT had a significant impact. For instance, they all decided that IoT improved effectiveness, efficiency, and productivity. IoT is a strategy feature for the project manager since they are responsible for aligning the project to achieve the organization's goals. In the closing phase of the project, all interviewees agreed that collecting and storing data in this stage is essential since it will act as a reference in the future. Therefore, they all decided that the implementation of IoT was crucial since it would ease collecting future referencing data.

Table 1: Participants

Number of participants	20
Participants that cooperated	15 (75%)
Participants that did not cooperate	1 (25%)

4.1.1 Benefits of Implementing IoT in Project Management. IoT gives a project innovation in organizational systems. It allows digital transformation that drives change and enables efficient product and project scope control (Respondent 1;4;10).

Examining the depth of IoT, it is the heart of product and service delivery. It is the most impacted sector by IT and engineering (Gal et al., [13]). 40% of the interviewees agreed that the primary advantages of IoT are protecting assets information and availing the data for analysis in real-time (2, 4, 5, 7, 8, 9, 10, 15). In addition, 40% of the respondents agreed that implementing IoT technologies positively impacted the organization's flow and improved processes (Xie & Yang, [37]).

In addition, IoT provides a high degree of transparency and data sharing (3, 7; 12). The IoT also forms the foundation for leveraging the value of products and allows project managers to examine any requisite information (7, 9, 12, 13, 14, 15). Information recognition from the entire project process is more accessible when the organization has implemented it than when it uses traditional information technology (1; 3; 4; 6; 11; 15). One of the most significant benefits is that it can help improve project management efficiency because IoT can provide real-time data and information about project status. Project managers can make decisions based on accurate and up-to-date information rather than relying on guesswork. According to 100% of respondents, it can lead to projects being completed more quickly and efficiently and can also help reduce project management costs.

5 Discussion

5.1 Impacts of IoT on Project Management

IoT has a significant impact on project management, as discussed by Saariko, et al., Percudani and Batrawi. [25]. Implementation of IoT in project management improves it in all phases. This statement is correlated with the interviewee's acceptance that IoT benefits the whole project phase. There is an unprecedented increase in production, increased awareness and efficient decision-making, and sensible improvement in project management (Tang et al., [32]). Since IoT can send real-time information to users, there is an increased flow of information in the entire project management and process (1, 2, 5, 6, 8, 10, 13, 14). Most organizations that have implemented IoT in their projects have reported a decrease in cost and time and making controlling operations easier (Sanchez, [30]). In addition, implementing IoT in project management reduces the wastage of resources. When an organization formulates a proper tracking technique of goods and resources, whether in project operation or storage ensures that resources are not misused, according to Percudani & Batrawi [25].

5.2 Barriers to Implementing IoT in Project Management

Adopting new technology is always challenging due to two distinct barriers to implementing IoT in project management. Technical and cultural issues are the main challenges facing organizations implementing IoT in their projects (Martens et al., [21]). Implementing new technology affects radical change and influences all organizational operations (Kozak-Holland & Procter, [19]). For instance, the IoT in a project has no guarantee of compatibility with stakeholders (4, 8; 9, 15). As discussed in previous chapters, IoT is the interconnection of various technological devices; therefore, these connections can be exposed to potential failures, which can stop key project sectors (Martens et al., [21]). In addition, implementing IoT in the project requires huge capital, according to all interviewees. For instance, continuous information collection needs massive cloud storage, which is costly.

6 Research Limitations

One of the critical limitations of IoT implementation research is the need for more standardization around IoT technologies, protocols, and data formats. This lack of standardization makes it difficult to compare and contrast different IoT implementations and to identify best practices. In addition, many IoT implementations are still in the early stages of development and must be rigorously tested or evaluated. As a result, there is a lack of empirical evidence to support the claims made about the benefits of IoT implementations. For instance, all fifteen interview respondents were from different organizations using various distinct standardization, which makes it difficult to relate. The research limitations of this study include the need for a comprehensive understanding of the role of IoT in effective project management. Furthermore, the study should have included a detailed analysis of the data collected from the survey.

7 Lessons Learned

One lesson is that it is essential to have a clear understanding of the goals and objectives of the project before implementation begins. Another takeaway is that it is crucial to clearly understand IoT technology's capabilities and limitations before the performance starts. Additionally, it is vital to consider IoT technology's security and privacy implications before the implementation begins. Finally, it is essential to have a clear plan for collecting and managing data before the implementation begins (Wang et al., [35]).

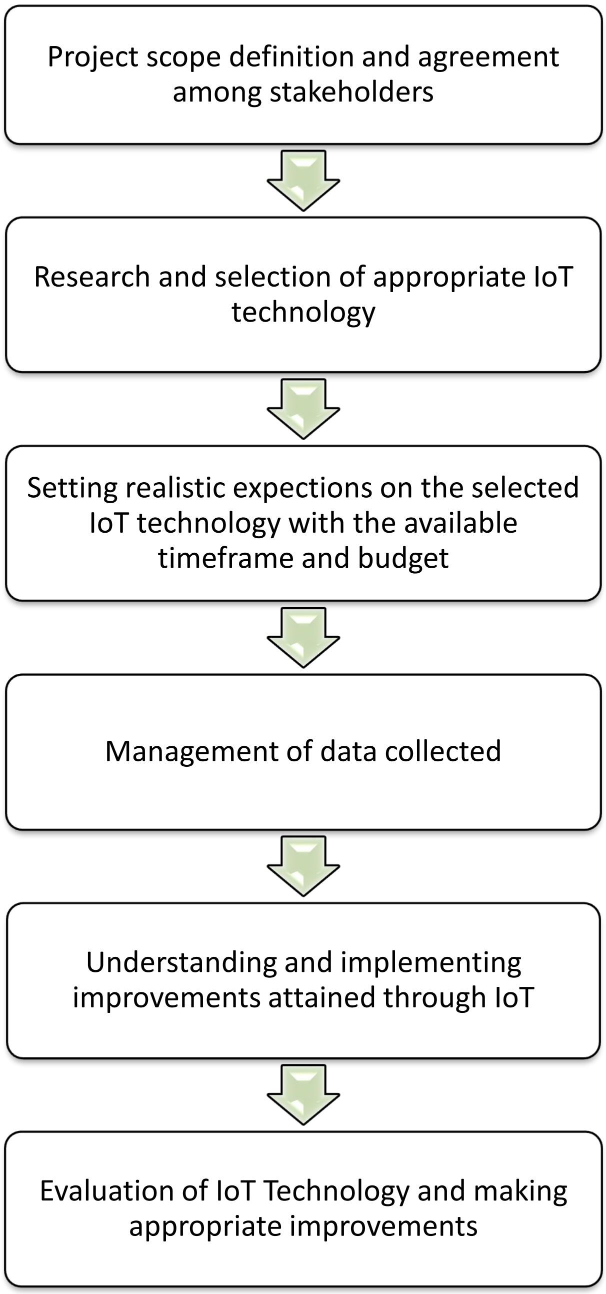


Figure 1: Recommendation steps

8 Recommendations

The following are research recommendations for implementing IoT in effective project management:

- From the outset, clearly define the project scope and objectives and ensure that all stakeholders agree (Prasher & Onu, [27]).
- Research and select the most appropriate IoT technology for the project requirements (Arnesen, [5]).
- Manage expectations around IoT's potential benefits and capabilities, and be realistic about what can be achieved within the timeframe and budget (Iriarte & Bayona, [17]).

- Manage the data collected by IoT devices carefully, considering data security, privacy, and compliance issues from the outset (Ancarani et a., [4]).
- Understand how IoT can improve project management processes and workflows and implement them accordingly (Adzmi & Hassan, [1]).
- Regularly review and assess the performance of the IoT implementation, and make necessary adjustments to ensure continued success (Shokouhyar et al., [31]).

9 Conclusion

The implementation of IoT in effective project management can help to improve the quality of project management and make it more efficient. IoT can monitor project progress, identify potential risks and issues, and provide real-time feedback to project managers. In addition, IoT can help automate project management processes, improve communication between project managers and stakeholders, and provide data-driven decision-making. IoT has the potential to transform project management into a more proactive, predictive, and adaptive field. By using IoT, project managers can gain insights into the behavior of project stakeholders, the status of project resources, and the performance of project deliverables. In addition, IoT can monitor and control project progress, identify and resolve issues in real-time, and improve communication between project managers and stakeholders.

References

[1] R. M. Adzmi, and Z. Hassan, “A Theoretical Framework of Critical Success Factors on Information Technology Project Management during Project Planning,” https://103.227. 140.9/handle/123456789/11593, 2018.

[2] L. Alazzawi, and J. Alotaibi, “Insight into IoT Applications and Common Practice Challenges,” Insight, 4(3):4(3):42-49, 2020, https://www.researchgate.net/profile/LubnaAlazzawi/publication/339697482_Insight_into_IoT_Applications_and_Common_Practice_Challenges/links/5e60080b4585152ce808fbe7/Insight-into-IoT-Applications-and-Common-Practice-Challenges.pdf.

[3] E. Amoah, and J. Y. Oh, “Blockchain in IoT and Project Management. Issues in Information Systems,” 21(3):268-278, 2020, https://iacis.org/iis/2020/3_iis_2020_268-278.pdf.

[4] A. Ancarani, C. Di Mauro, H. Legenvre, and M. S. Cardella, “Internet of Things Adoption: A Typology of Projects,” *International Journal of Operations & Production Management*, 40(6):849-872, 2019, https://doi.org/10.1108/IJOPM-01-2019-0095.

[5] C. Arnesen, *Digital Transformation in Projects: A Study on How Digital Transformation Affects Project Management*, Master’s Thesis, University of Agder, 2020, https://uia.brage.unit.no/uiaxmlui/bitstream/handle/11250/2677068/Camilla%20Arnesen.pdf?sequence=1&isAllowed=y.

[6] L. Babun, K. Denney, Z. B. Celik, P. McDaniel, and A. S. Uluagac, “A Survey on IoT Platforms: Communication, Security, and Privacy Perspectives, *Computer Networks*, 192, 108040. http://dx.doi.org/10.1016/j.comnet.2021.108040, (2021).

[7] P. Brous, M. Janssen, and P. Herder, “The Dual Effects of the Internet of Things (IoT): A Systematic Review of the Benefits and Risks of IoT Adoption by Organizations, *International Journal of Information Management*, 51:101952, 2020, https://doi.org/10.1016/j.ijinfomgt.2019.05.008.

[8] N. Durmic, Factors influencing project success: A qualitative research. TEM Journal, 9(3):1011-1020, 2020, https://www.ceeol.com/search/article-detail?id=894785.

[9] M. El Khatib, “BIM is a Tool to Optimize and Manage Project Risk Management, *Int. J. Mech. Eng*, 7(1):6307-6323, 2022, https://www.researchgate.net/publication/358375029_BIM_as_a_tool_to_optimize_and_manage_project_risk_management.

[10] M. El-Khatib, R. Abu Zitar, and A. Al Nakeeb, *The Effect of AI on Project and Risk Management in Health Care Industry Projects in the United Arab Emirates (UAE)*, https://www.researchgate.net/profile/Mounir-El-Khatib/publication/348097610_The_effect_of_AI_on_project_and_risk_management_in_health_care_industry_projects_in_the_United_Arab_Emirates_UAE/links/5feef183299bf14088612609/The-effect-of-AI-on-project-and-risk-management-in-health-care-industry-projects-in-the-United-Arab-Emirates-UAE.pdf, 2021.

[11] M. El Khatib and A. Al Falasi, “Effects of Artificial Intelligence on Decision Making in Project Management,” *American Journal of Industrial and Business Management*, 11(3):251-260, 2021, https://www.scirp.org/journal/paperinformation.aspx?paperid=107733.

[12] M. Elkhatib, A. A. Hosani, I. A. Hosani, and K. Albuflasa, *Agile Project Management and Project Risks Improvements: Pros and Cons*, Modern Economy, https://www.scirp.org/journal/paperinformation.aspx?paperid=119775, 2022.

[13] A. Gal, I. Filip, and F. Dragan, “A New Vision Over Agile Project Management in the Internet of Things Era,” Procedia-Social and Behavioral Sciences, 238:277-285, 2018, https://doi.org/10.1016/j.sbspro. 2018.04.003.

[14] S. Ghimire, F. Luis-Ferreira, T. Nodehi, and R. Jardim-Goncalves, “IoT-Based Situational Awareness Framework for Real-Time Project Management,” *International Journal of Computer Integrated Manufacturing*, 30(1):74-83, 2017, https://doi.org/10.1080/0951192X.2015.1130242.

[15] K. K. Goyal, A. Garg, A. Rastogi, and S. Singhal, “A Literature Survey on the Internet of Things (IoT), *International Journal of Advanced Networking and Applications*, 9(6):3663-3668, 2018, https://www.researchgate.net/profile/Krishan-Goyal-3/publication/338458451_A_Literature_Survey_on_Internet_of_Things_IoT/links/5e1618b94585159aa4be6a03/A-Literature-Survey-on-Internet-of-Things-IoT.pdf.

[16] V. Hurtoi and D. Avadanei, “IoT Project Management,” Informatica Economica, 24(3/2020):75-80, 2020, https://doi.org/10.24818/issn14531305/24.3.2020.07.

[17] C. Iriarte and S. Bayona, IT projects Success Factors: A Literature Review. International Journal of Information

Systems and Project Management, 8(2):49-78, 2020, https://doi.org/10.12821/ijispm080203.

[18] S. Khalafi, M. Khalilzadeh, S. Razaghi, and E. Vaezi," Investigating the Impact of the Internet of Things on the Performance of Knowledge Areas of Project Management," *International Journal of Internet Manufacturing and Services*, 8(2):89-102, 2021, https://doi.org/10.1504/IJIMS.2021.122702.

[19] M. Kozak-Holland and C. Procter, "The Challenge of Digital Transformation," Managing Transformation Projects, Palgrave Pivot, Cham, pp. 1-11, 2020, https://doi.org/10.1007/978-3-030-33035-4_1.

[20] W. Liu, "Integrating Data Mining with the Internet of Things to Build the Music Project Management System," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1-13, (2020, https://doi.org/10.1007/s12652-020-02327-x.

[21] C. D. Martens, L. F. Silva, D. F. Silva, and M. L. Martens, „Challenges in Implementing Internet of Things Projects and Actions to Overcome Them, *Technovation*, 118:102427, 2022, https://doi.org/10.1016/j.technovation.2021.102427.

[22] J. H. Nord, A Koohang,.and J. Paliszkiewicz, "The Internet of Things: Review and Theoretical Framework," Expert Systems with Applications, 133:97-108, 2019, https://doi.org/10.1016/j.eswa.2019.05.014.

[23] Parteek, "A Review Paper on IoT Advantages and Disadvantages," *International Journal of Research and Analytical Reviews*, e ISSN 2348 –1269, Print ISSN 2349-5138, , 2019.

[24] B. C. Patel, R. S. Tripathi, and N. Goel, "IoT an Overview: Advantage, Disadvantage, and Applications," *International Journal of Computer Applications Technology and Research*, 10(05):119-122, 2021, https://ijcat.com/archieve/volume10/issue5/ijcatr10051003.pdf.

[25] P. Percudani and M. Batrawi, "The Impact of Internet of Things Unification with Project Management Disciplines in Project-Based Organizations," http://www.diva-portal.org/smash/get/diva2:1187433/FULLTEXT01.pdf, 2017.

[26] V. S. Prasher, "Internet of Things (IoT) and Changing Face of Project Management," https://digitalcommons.harrisburgu.edu/pmgt_dandt/49/, 2018.

[27] V. S. Prasher, S. Onu, The Internet of Things (IoT) upheaval: Overcoming management challenges. The Journal of Modern Project Management, 8(2):24-37, 2020, https://doi.org/10.19255/JMPM02402.

[28] R. M. L. Rebelo, S. C. F. Pereira, and M. M. Queiroz, "The Interplay between the Internet of Things and Supply Chain Management: Challenges and Opportunities Based on a Systematic Literature Review," *Benchmarking: An International Journal*, https://doi.org/10.1108/BIJ-02-2021-0085, 2021.

[29] C. Rivett, *Cisco Survey has Revealed that Three-Quarters of IoT Projects are Failing*, Technology Magazine., from https://technologymagazine.com/ai-and-machine-learning/cisco-survey-has-revealed-three-quarters-IoT-projects-are-failing, May 17, 2020, Retrieved November 26.

[30] M. A. Sanchez, "How the Internet of Things is Transforming Project Management," Research Anthology on Digital Transformation, Organizational Change, and the Impact of Remote Work, IGI Global, pp. 463-484, 2021, Https://10.4018/978-1-7998-7297-9.ch024.

[31] S. Shokouhyar, S. Zarrin, and S. Shokoohyar, "Analyzing the Impact of IT Governance on the Performance of Project-Based Organizations," *International Journal of Business and Systems Research*, 14(4):411-433, 2020, https://www.researchgate.net/profile/Sajjad-Shokouhyar/publication/343497406_Analysing_the_impact_of_IT_governance_on_the_performance_of_project based_organisations/links/60674477299bf1252e24318f/Analysing-the-impact-of-IT-governance-on-the-performance-of-project-based-organisations.pdf.

[32] C. P. Tang, T. C. K. Huang, ands S. T. Wang, „The Impact of Internet of Things Implementation on Firm Performance," Telematics and Informatics, 35(7), 2038-2053, 2018, https://doi.org/10.1016/j.tele.2018.07.007.

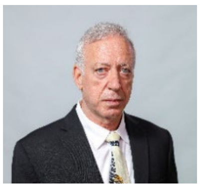
[33] Y. M. Tukur, D. Thakker, and I. U. Awan, Multi-layer approach to internet of things (IoT) security. *2019 7th International Conference on Future Internet of Things and Cloud (FiCloud)* IEEE, pp. 109-116, August 2019, https://doi.org/10.1109/FiCloud.2019.00023.

[34] N. K. Vishwakarma, R. K. Singh, and R. R. K. Sharma, "Internet of Things Architectures: Do Organizational Strategies Matter? *Business Process Management Journal*, 26(1):102-131, 2020, https://doi.org/10.1108/BPMJ-03-2018-0092.

[35] B. Wang, T. Y. Fan, X. T. Nie, "Advanced Delay Assured Numerical Heuristic Modelling for Peer to Peer Project Management in Cloud Assisted Internet of Things platform," Peer-to-Peer Networking and Applications, 13(6):2166-2176, 2020, https://doi.org/10.1007/s12083-020-00883-9.

[36] H. Wang, X. Luo, and X. Yu, „Exploring the Role of IoT in Project Management Based on the Task-Technology Fit Model," *Procedia Computer Science*, 199:1052-1059, 2022, https://doi.org/10.1016/j.procs.2022.01.133.

[37] H. Xie and Z. Yang, "The Risk Management Mode of Construction Project Management in the Multimedia Environment of the Internet of Things," Mobile Information Systems, https://doi.org/10.1155/2021/1311474, 2021.



Mounir El khatib has a long and varied experience in Digital Government and Digital Transformation in the UK, Middle East and UAE. He has worked as a practitioner and in academic focused areas like Project and Program Management, Information Systems, Strategy and Operations, Business process and Service Management, providing teaching, training, consulting, advising, research and professional services.

Raed Abu Zitar (BS 88, MS 90, and PhD 93) is a senior research scientist in the context of Thales Endowed Chair of Excellence, Sorbonne Center of Artificial Intelligence (SCAI), Sorbonne University- Abu Dhabi (SUAD), UAE. Prof. Abu Zitar has joined SCAI-SUAD in Feb-2021. Prof. Abu Zitar is directing the research at Thales/SCAI in the fields of Intelligent detection and tracking of drones in cooperation with industrial experts from Thales and Aveillant (France and the UK). Prior to that date, he was the founding coordinator for the Master of Artificial Intelligence Program (one of the first in UAE), Head of the Center of Artificial Intelligence, and Manager for the Teaching and Learning Center at Ajman University, UAE. He has more than 30 years of experience in teaching, research, and administration in the USA, Jordan, and the UAE. He has published more than 100 Journal and conference papers.

Ahmed Alnaqbi (photo not available) is an operations plant supervisor at ADNOC Offshore company. In 2018, Ahmed got his bachelor degree in chemical engineering at Khalifa University. Ahmed is graduate master in Project Management at Hamdan Bin Mohamed University. Ahmed has 4 years of experience in the oil and gas Industry.

Abdulla Alnaqbi (photo not available) is a Trading Operations Specialist at ADNOC Trading. In 2018 Alnaqbi graduated with a Bachelor's in Chemical Engineering from Pennsylvania State University. From 2019 to 2022 he supervised the refinery operations for ADNOC Refining. Abdulla is a graduate of Master degree in Project Management from HBMSU and just started his research journey by publishing his articles.

Hamad Alsuwaidi (photo not available) is an Inspection Engineer at ADNOC Onshore Under Projects & Technical Centers Division. In Dec 2018 Alsuwaidi graduated with a Bachelor's in Mechanical Engineering from the HCT. Joined ADNOC Onshore since 2019. Hamad is a graduate of Master degree in Project Management from HBMSU.

Meera Al-Marri (photo not available) holds a PhD in Project Management. She is an Assistant Professor at the School of Business and Quality Management at HBMSU. Research interests include project management, online learning, sustainability, and CSR.

Ahmed Ankit is an accomplished academic leader with a passion for promoting research excellence. He earned his Bachelor's degree from Mohammed the First University in Morocco, and currently serves as the Dean of Research and Doctoral Studies at Hamdan Bin Mohammed Smart University, Dubai, U.A.E. Prof. Ankit has played a key role in securing licensure and accreditation for several academic programs, showcasing his dedication to academic quality. His research interests lie in the areas of smart education and higher education management, and he has co-authored numerous papers on these topics. Through his work, Ahmed strives to foster a culture of innovation and excellence in academia.

Appendices

Interview Questions

After a brief introduction, the interviewer welcomed the interviewee in the next chapter, where they were required to answer questions based on IoT. The research questionnaire progressed using the following questions.

i. Please tell me about yourself – your area of specialization and occupation.
ii. In brief answers, please describe your experiences in research.
iii. In your previous research, have you ever handled IoT? If yes, how did you engage in this topic?
iv. Based on your knowledge, what are the advantages of an IoT system?
v. What barriers do you face while implementing IoT in project management and using it?
vi. What are your thoughts on using IoT in project management?
vii. How can IoT be used to improve project management?
viii. Are the barriers mentioned above and difficulties prevalent in all industries or organizations?
ix. Do those challenges, difficulties, and barriers vary based on the situation?
x. If those challenges were prevalent in a different organization, what is prevalent among all industries?
xi. In project management, the program has several phases; which one is more affected by implementing IoT technology?
xii. Why does IoT mainly impact the phase you selected in (ix) above? Give examples.
xiii. Will the introduction of IoT technology impact the role of the project manager? If yes, describe your opinion on how the effect will be.
xiv. If a project incorporates new technology, are there common skills required from the manager?
xv. Will the project manager develop a new skill?
xvi. What are the new skills that project managers will acquire?
xvii. Since the project manager has different functions, which is more affected by IoT technology, and which is less impacted?

xviii. Why do you think those functions are most and less impacted? Give examples.
xix. Do you think IoT is a necessary tool for project management? Why or why not?
xx. How would you go about implementing IoT in project management?
xxi. Are there other factors that need to be considered before implementing IoT in project management?
xxii. What changes will be made to the organization's IT infrastructure to support IoT in project management?
xxiii. What training and support will be necessary for project managers to use IoT in project management effectively?
xxiv. Thank you for your time. Is there anything else you would like to share about your organization's IoT plans or considerations?

Journal Submission

The International Journal of Computers and Their Applications is published four times a year with the purpose of providing a forum for state-of-the-art developments and research in the theory and design of computers, as well as current innovative activities in the applications of computers. In contrast to other journals, this journal focuses on emerging computer technologies with emphasis on the applicability to real world problems. Current areas of particular interest include, but are not limited to: architecture, networks, intelligent systems, parallel and distributed computing, software and information engineering, and computer applications (e.g., engineering, medicine, business, education, etc.). All papers are subject to peer review before selection.

A. Procedure for Submission of a Technical Paper for Consideration

1. Email your manuscript to the Editor-in-Chief, Dr. Ajay Bandi. Email: ajay@nwmissouri.edu.
2. Illustrations should be high quality (originals unnecessary).
3. Enclose a separate page (or include in the email message) the preferred author and address for correspondence. Also, please include email, telephone, and fax information should further contact be needed.
4. **Note**: Papers shorter than 10 pages long will be returned.

B. Manuscript Style:

1. **WORD DOCUMENT**: The text should be **double-spaced** (12 point or larger), **single column** and **single-sided** on 8.5 X 11 inch pages. Or it can be single spaced double column.

   LaTex DOCUMENT: The text is to be a double column (10 point font) in pdf format.
2. An informative abstract of 100-250 words should be provided.
3. At least 5 keywords following the abstract describing the paper topics.
4. References (alphabetized by first author) should appear at the end of the paper, as follows: author(s), first initials followed by last name, title in quotation marks, periodical, volume, inclusive page numbers, month and year.
5. The figures are to be integrated in the text after referenced in the text.

C. Submission of Accepted Manuscripts

1. The final complete paper (with abstract, figures, tables, and keywords) satisfying Section B above in **MS Word format** should be submitted to the Editor-in-Chief. If one wished to use LaTex, please see the corresponding LaTex template.
2. The submission may be on a CD/DVD or as an email attachment(s). **The following electronic files should be included:**
   - Paper text (required).
   - Bios (required for each author).
   - Author Photos are to be integrated into the text.
   - Figures, Tables, and Illustrations. These should be integrated into the paper text file.
3. Reminder: The authors photos and short bios should be integrated into the text at the end of the paper. All figures, tables, and illustrations should be integrated into the text after being mentioned in the text.
4. The final paper should be submitted in (a) pdf AND (b) either Word or LaTex. For those authors using LaTex, please follow the guidelines and template.
5. Authors are asked to sign an ISCA copyright form (http://www.isca-hq.org/j-copyright.htm), indicating that they are transferring the copyright to ISCA or declaring the work to be government-sponsored work in the public domain. Also, letters of permission for inclusion of non-original materials are required.

Publication Charges

After a manuscript has been accepted for publication, the contact author will be invoiced a publication charge of **$500.00 USD** to cover part of the cost of publication. For ISCA members, publication charges are **$400.00 USD** publication charges are required.

Revised 2020

ISCA INTERNATIONAL JOURNAL OF COMPUTERS AND THEIR APPLICATIONS Vol. 30, No. 2, June 2023