International Journal of Informatics Society

06/24 Vol.16 No.1 ISSN 1883-4566

Informatics Society

Aims and Scope

The purpose of this journal is to provide an open forum to publish high quality research papers in the areas of informatics and related fields to promote the exchange of research ideas, experiences and results.

Informatics is the systematic study of Information and the application of research methods to study Information systems and services. It deals primarily with human aspects of information, such as its quality and value as a resource. Informatics also referred to as Information science, studies the structure, algorithms, behavior, and interactions of natural and artificial systems that store, process, access and communicate information. It also develops its own conceptual and theoretical foundations and utilizes foundations developed in other fields. The advent of computers, its ubiquity and ease to use has led to the study of informatics that has computational, cognitive and social aspects, including study of the social impact of information technologies.

The characteristic of informatics' context is amalgamation of technologies. For creating an informatics product, it is necessary to integrate many technologies, such as mathematics, linguistics, engineering and other emerging new fields.

Guest Editor's Message

Kei Hiroi

Guest Editor of the Forty-sixth Issue of the International Journal of Informatics Society

We are delighted to have the Forty-sixth issue of the International Journal of Informatics Society (IJIS) published. This issue includes selected papers from the Sixteenth International Workshop on Informatics (IWIN2023), held online from September 1st – 4th, 2023. The workshop was the seventeenth event for the Informatics Society. It was intended to bring together researchers and practitioners to share and exchange their experiences, discuss challenges and present original ideas in all aspects of informatics and computer networks. In the workshop, 27 papers were presented in six technical sessions. The workshop was successfully finished, and precious experiences were provided to the participants. It highlighted the latest research results in informatics and its applications, including networking, mobile ubiquitous systems, data analytics, business and industrial systems, education systems, design methodology, intelligent systems, groupware, and social systems, etc.

Each paper submitted to IWIN2023 was reviewed in terms of technical content, scientific rigor, novelty, originality, and presentation quality by at least two reviewers. Through those reviews, 22 papers were selected for publication candidates of the IJIS Journal, and they were further reviewed as Journal papers. We have three categories of IJIS papers, Regular papers, Practical papers, and Invited papers, each of which was reviewed from different points of view. This volume includes papers among those accepted papers, which have been improved through the workshop discussion and the reviewers' comments.

We publish the journal in print as well as in an electronic form over the Internet. We hope that the issue would be of interest to many researchers as well as engineers and practitioners all over the world.

Kei Hiroi received B.S. in Electronic Engineering from the School of Engineering at Tohoku University and joined Nippon Telegraph and Telephone East Corporation in 2004. Since then, she has worked in telephone transmission network fault repair, telephone switcher management, congestion control, and mainly in telecommunications disaster management. Then, she entered Keio University in 2009 and received her Master of Media Design and Ph.D. in Media Design in 2011 and 2014, respectively from Keio University. She has been an assistant professor in the department of Information and Communication Engineering, Graduate School of Engineering, Nagoya University. She is currently an associate professor in Disaster Prevention Research Institute, Kyoto University. Her research interests include disaster simulation, and crisis computing.

Regular Paper

An Evaluation on a Power Interchange Method for Realizing Net-Zero Energy House for Multiple Small Communities

Masashi Saito*, Tomoki Nomura** and Yuichi Tokunaga*

* Faculty of Graduate School of Engineering, Kanazawa Institute of Technology, Japan
** Graduate School of Engineering, Kanazawa Institute of Technology, Japan
{msaito, y.tokunaga}@neptune.kanazawa-it.ac.jp,
c6101950@planet.kanazawa-it.ac.jp

Abstract - We have developed a simulator to achieve net-zero energy through power sharing among multiple households and communities. In multi-household transactions, when the capacity of one household's electricity storage system is depleted and power supply becomes necessary, power can be purchased from other households to meet demand without having to purchase power from the grid, thus enabling power operations that do not rely on grid power. In transactions in which the scale of transactions was extended to multiple communities, a certain increase in the renewable energy consumption rate was confirmed even in communities with low storage battery capacity and photovoltaic power generation capacity through the inter-community exchange of electricity by scattering large-scale facility households with different electricity consumption styles. The seasonal changes in the self-consumption rate showed that securing electricity in the fall and winter is necessary to maintain a high self-consumption rate and renewable energy consumption rate. In the Hokuriku region, where consumption increases in winter, it is considered necessary to introduce new renewable energy generation facilities in addition to photovoltaic power generation to achieve net-zero power consumption per community. In this simulation, only data from the Hokuriku region was used for electricity consumption and solar power generation. By replacing these data, the feasibility of net-zero energy can be verified for all regions, showing that a wide variety of simulations in social systems is possible.

Keywords: Smart Community, Electric Power Management, Net Zero Energy House, Renewable Energy, Simulation

1 INTRODUCTION

The introduction of renewable energies in Japan has rapidly increased with the introduction of the feed-in tariff (FIT) program, which began in November 2009, requiring electric power companies to purchase electricity generated from renewable energies (solar, wind, hydro, geothermal, and biomass) for a certain period at a price set by the government. For residential PV (residential photovoltaic power generation) of less than 10 kW, surplus power after self-consumption is eligible for purchase for 10 years. The system will gradually expire from November 2019. According to the Agency for Natural Resources and Energy, the total number of residential PV systems that will expire under the FIT will be 730,000 (2.82 million kW) in 2020 and 1.65 million (6.7 million kW) in 2023 [1].

The price of electricity sold during the FIT period was 48 yen/kWh in FY 2009. After the end of the FIT, the price will be about 8 yen/kWh, and compared to the purchase price of about 20 yen/kWh, it is considered more economical to consume the electricity generated on site. For this reason, attention is focused on self-consumption houses that generate electricity during the daytime when the sun is rising, store the surplus in storage batteries without selling it to the grid, and discharge the stored electricity at night.

A house that focuses on self-consumption is similar to a net zero energy house (henceforth ZEH). ZEH is a house that aims for zero primary energy consumption with the three characteristics of self-consumption, thermal insulation, and energy saving [2]. The use of air conditioners is reduced by improving the insulation performance of the house, and the entire house is made energy efficient by high-efficiency equipment and HEMS (Home Energy Management System). Since the electricity consumption is low, a combination of PV power generation and storage batteries is expected to be effective.

Storage batteries play an important role in realizing ZEH. The price of storage batteries is on a downward trend, but the price decline will slow down when the cost of raw materials for lithium-ion batteries is taken into account. Instead of installing storage batteries for home use, storage batteries installed in EVs could be used as storage batteries for home use. If EVs are the vehicles used on a daily basis, this is V2H (Vehicle to Home) which links EVs and homes [3]. The storage batteries of EVs that have reached the end of their service life could be used as a substitute for household storage batteries. In terms of used EVs, EVs that went on sale around 2010 have reached the end of their service life and are beginning to appear on the market.

Furthermore, the unstable global situation is causing electricity prices to skyrocket. According to the Agency for Natural Resources and Energy, the Japanese electricity market, which was at about 8 yen/kWh until September 2021, recorded about 26.2 yen/kWh in March 2022 [4]. Against this background, a shift to distributed energy systems using renewable energy sources that do not depend on existing centralized energy systems is being promoted [5]. The operation of distributed energy systems is expected to

realize "net zero energy," i.e., zero primary energy consumption.

Smart grids and microgrids are effective power transmission systems for building distributed energy systems. A smart grid is a power transmission system that enables control of the electricity flow not only on the supply side but also on the demand side, and this system makes it possible for consumers to buy and sell electricity from the supply side or between consumers. A microgrid is an electric power network in which a smart grid is deployed within a certain region and energy is produced and consumed locally within the community without relying on external supplies. Microgrid operations that do not rely on electricity sales from suppliers can be expected to secure electricity and heat in the event of a disaster and to reduce electricity costs.

In addition, the need to transmit electricity over long distances in rural areas, including mountainous regions of Japan, has raised concerns about the rising cost of infrastructure maintenance. By operating a microgrid and building an independent power infrastructure zone using renewable energy generation within the surrounding area, it is possible to operate power without relying on an external power supply even when the area is remote from urban areas, eliminating the need to manage the infrastructure of a centralized energy system. As a result, infrastructure maintenance costs can be significantly reduced [6].

In this study, we have first developed a simulator that assumes the installation of a PV power generation and storage system in a house for the realization of ZEH and evaluate the economic feasibility by changes in the amount and cost of the system installation. Next, assuming a small-scale community in a mountainous area, we demonstrate the feasibility of independent power operation and economical power management only in the community and its surrounding area when a microgrid is actually constructed, and discuss the evaluation results using the simulator.

2 RELATED WORK

Ishikawa and Matsuo [7] conducted a simulation using storage battery capacity, connection and charge/discharge control methods, and water heater boil-up time as parameters to evaluate the primary energy savings and economic efficiency of installing storage batteries in PV equipment in post-FIT homes. The best results were obtained by shifting the boiling time of the heat-pump water heater without installing the storage battery system, which resulted in an annual profit of 0.37 million yen and a reduction in primary energy consumption of 14 MJ (3.9 kWh). The paper also argues that if 3 kWh of storage batteries are installed at a unit price of 8 yen/kWh, one of the following conditions must be met: the storage battery installation cost must be less than 150,000 yen, the durability period must be extended to 28 years, or the electricity price must rise to 186% of the current price.

Yabe et al. [8] evaluated the economic efficiency of storage batteries by determining the changes in the self-consumption rate and the annual cost before and after the introduction of storage batteries for each consumer based on electricity demand data and actual solar power generation data for residences and businesses with solar power generation throughout Japan. The study concluded that a 5kWh storage battery at 60,000 yen/kWh can be installed in a house, and that the investment can be recovered in about 15 years. The payback period is about 11 years when electricity rates with late-night starting times are selected and varies depending on the demand characteristics and other factors. For commercial customers, the payback period is about 8 years when a 60,000 yen/kWh storage battery with the PV power generation output for 30 minutes is installed.

Goto et al. [9] proposed a framework and a solution method for the problem of determining the supply-demand operation plan of a microgrid by calculating the operating reserve based on the forecast error trends of power demand and PV power output. The validity of the proposed framework is verified through numerical simulations. The demand-supply operation planning problem is formulated as a demand-supply operation planning problem for a general microgrid. In addition, they defined a demand-supply operation planning problem that considers uncertainty by using the probability distribution of the net demand forecast. In the simulation, the operational cost for the forecasted value of net demand resulted in worse solution results compared to the conventional method. On the other hand, the solution succeeded in reducing the expected value of operational cost by approximately 20% compared to the conventional method.

Komiyama et al. [10] evaluated the optimal combination of power generation facilities, considering the introduction of renewable energy and inter-regional power transmission. Wind power generation facilities in Japan are concentrated in regions such as Hokkaido and Tohoku, and there is a need to develop inter-regional power transmission lines. Therefore, an optimal power generation model was developed as a linear programming model that considers the transmission of power using inter-regional transmission lines. As a result, it was shown that wind power can be supplied to the Tokyo metropolitan area, where solar power is the main source of renewable energy supply, by expanding the regional power grid. It was also shown that the introduction of more PV power generation facilities in the Kyushu and Shikoku regions, which have long hours of sunlight, should be promoted.

3 DEVELOPMENT OF ZEH ECONOMIC EVALUATION SIMULATOR

3.1 ZEH Model

This section discusses a house in which a PV power generation system and an energy storage system are installed. Electricity is supplied to the house from the amount of electricity generated by the PV power generation system. If consumption exceeds the amount of electricity generated, the storage batteries are discharged to make up for the power shortfall. If the amount of electricity generated and discharged from the storage batteries is still insufficient,

the remaining shortage is met by purchasing electricity from the grid. Surplus power from solar power generation is charged into storage batteries, and surplus power that cannot be charged due to full battery is sold to the grid. These power transfers are controlled through HEMS.

We will also consider the use of electricity consumption forecasting and PV power generation forecasting. Smart meters measure residential electricity consumption, and PV equipment sensors measure power generation and solar radiation. The power consumption forecast, PV power generation forecast, and remaining storage battery capacity can be used to predict future excesses and deficiencies in power, enabling power control to minimize power costs.

The assumed ZEH is shown in Fig. 1. In this study, the power transfer in Fig. 1 is first simulated.

3.2 Overview of ZEH Economic Evaluation Simulator

The data used in the simulation, such as the amount of electricity consumption and solar power generation, are recorded in CSV and DB formats. When parameters such as the amount of installed PV and energy storage systems and installation costs are entered into the simulator and executed, the data is read and the power transfer simulation is started.

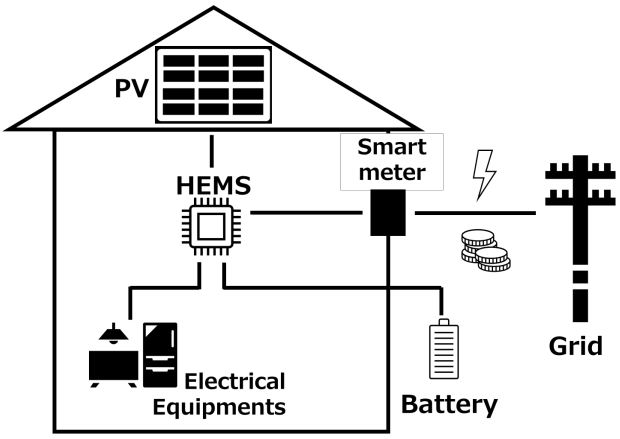


Figure 1: ZEH Model

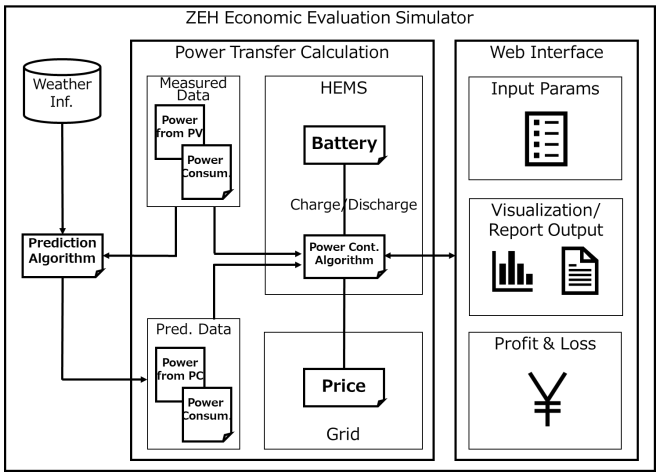


Figure 2: ZEH Economic Evaluation Simulator System Structure

Power transfer calculations are performed in every 30-minute period over a year. After the simulation is complete, a graph is plotted showing the power transfer and profit/loss. The resulting data can be downloaded in CSV format.

The function and configuration of the simulator developed in this study are shown in Fig. 2.

3.3 Solar Power Generation Data

An energy management experiment system for distributed energy is being conducted at the Hakusanroku Campus of the Kanazawa Institute of Technology, and energy-related data is being collected as part of the experiment [11]. Measured solar-generated electricity in 2020 is used for the power transfer simulation. Since the data is measured at 2-second intervals, we calculated the amount of electricity by using the average value of electricity generated every 30 minutes.

3.4 Electricity Consumption Data

The electricity consumption of eight detached houses in the Hokuriku region (Niigata Prefecture) is used to create electricity consumption data from the energy consumption data published by the Architectural Institute of Japan [12]. The electricity consumption of the eight houses was averaged and aggregated by time and classified into three categories. Using the consumption of eight houses as the basic data used in the simulation may lead to a deterioration in accuracy. In the future, if it becomes possible to obtain or measure more data, the accuracy of the simulation can be expected.

Electricity consumption and housing assignments are shown in Table 1.

The classified data are averaged and aggregated by week, day of the week, and time, and scaled with the maximum value set to 1, and used in the simulation. The simulation reproduces electricity consumption according to households by adjusting the size according to a parameter called the average monthly electricity consumption rate. The electricity charges follow the metered B electricity rates of Hokuriku Electric Power Company.

3.5 Power Transfer Simulation

Figure 3 shows an overview of the simulation algorithm. First, a simulator using real data is developed, and later a simulation that takes the electricity supply-demand forecast into account and a simulation in which storage batteries are replaced by EVs are implemented.

Table 1: Three distinguished power consumption models

Model	House ID
Night/Midnight: Concentrated use at night or late at night when the user is away during the day, late-night hot water supply	1,2,6
Evening: Always at home and often used in the evening	4,5
Morning/Evening: Always at home and often used in the morning and evening	3,7,8

4 SETTING UP A ZEH SIMULATION

4.1 Electricity Consumption

Each power consumption model is used in the simulation with the power consumption sized as an average monthly power consumption charge of 10,000 yen. For the adjusted power consumption, Fig. 4 shows the monthly power consumption and Fig. 5 shows the average daily power consumption.

Annual electricity consumption was 5621 kWh. The annual electricity consumption per household in Hokuriku is 6333 kWh, which is about 11% lower than the annual consumption per household in the same area. In all cases, electricity consumption is high in winter and low in spring and fall.

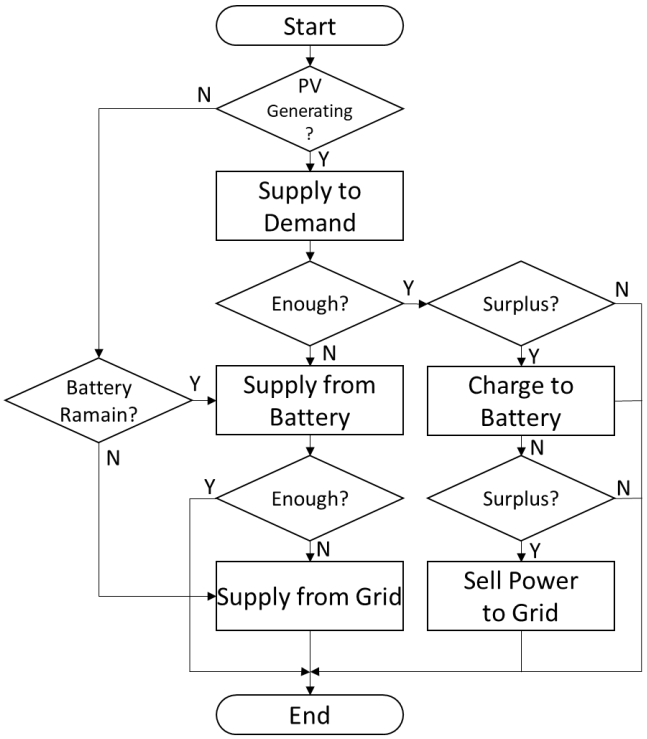


Figure 3: ZEH Economic Evaluation Simulator Algorithm Outline

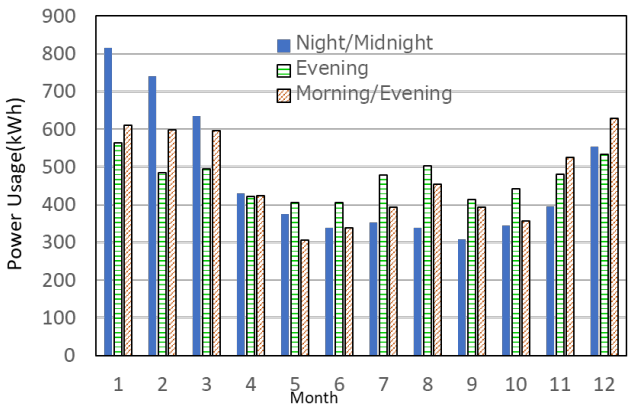


Figure 4: Monthly total power consumption amount

4.2 Setting up PV Power Generation and Energy Storage Systems

We simulate combinations of PV outputs of 5, 10, 15, and 20 kW and storage capacities of 0, 5, 10, and 40 kWh, respectively. The storage capacity of 40 kWh is based on the case where an EV is used as a storage battery. Degradation of the power generation and storage system is not considered in this paper. Table 2 shows the settings of the installed system.

4.3 System Pricing Assumptions

The installation cost of the PV system is assumed to be 336,000 yen/kW, the main unit cost of the stationary storage battery 140,000 yen/kWh, and the main unit cost of the V2H system 493,000 yen [13]. The cost of the EV itself is not included. The construction cost of the energy storage system is assumed to be 336,000 yen per installation. Table 3 shows the unit installation costs of stationary storage batteries, PV power generation, and energy storage systems, and Table 4

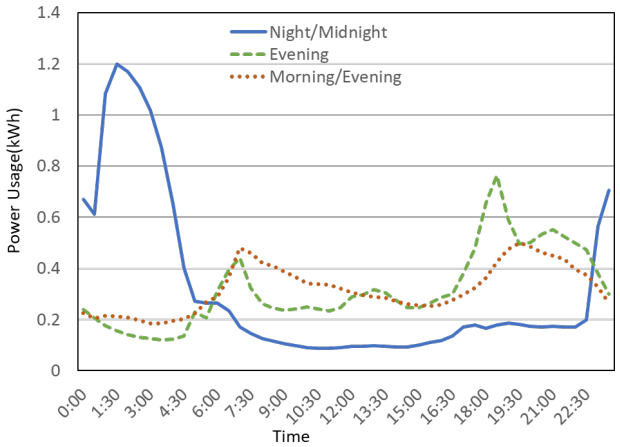


Figure 5: Average daily power consumption

Table 2: System parameters

Parameter	Value
Power conversion efficiency	95%
Maximum charge/discharge efficiency	95%
Maximum charge and discharge power	3kW

Table 3: System costs per unit

System	Cost
solar power (generation)	336,000 yen/kW
Stationary storage battery	140,000 yen/kWh
V2H(EV)	493,000 yen/unit
Energy storage system installation cost	336,000 yen per case

Table 4: System installed costs in this paper

PV (generation)		Battery system	
Spec.	Price(M Yen)	Spec.	Price(M Yen)
5kW	1. 68	0 kWh	0
10kW	3.36	5 kWh	1.036
15 kW	5.04	10kWh	1.736
20kW	6.72	40kWh (EV)	0.829

shows the system installation costs in this paper.

The cost of using an EV for the energy storage system is 493,000 yen for V2H plus 336,000 yen for construction, regardless of the capacity.

Since electricity consumption tends to be low during the day and increases from evening to night, we adapted Hokuriku Electric Power Company's Kutsurogi-night 12 [14] as the electricity rate plan for our simulation. Kutsurogi Night 12 is a rate plan with low electricity prices during the nighttime (20:00~8:00). This report considers only the basic charge (1,650 yen/month) and the electricity volume charge, and does not include the renewable energy surcharge and the unit price of fuel adjustment. The unit price of electricity sold is assumed to be 8 yen.

5 ZEH ECONOMIC EVALUATION SIMULATION RESULTS

5.1 Profits and Losses on Electricity Sales for the Year

The difference between the simulated profit/loss on electricity sales and the conventional electricity rate (120,000 yen) was calculated and used as the profit/loss for one year. Basic charges and system installation costs are not included.

Figure 6 shows the one-year power trading profits and losses for the evening model.

The maximum profit was 178,000 yen in the model with 20 kW of PV power generation, 10 kWh of storage capacity, and consumption in the evening. Although profit increases with increasing PV output, it cannot be said that profit increases with increasing storage capacity.

5.2 15-Year Profit/Loss Including Initial Cost

A minimum loss of 26,000 yen was recorded for the concentrated nighttime model with only 5 kW of PV power generation, and a maximum loss of 6,365,000 yen was recorded for the concentrated morning and evening model with 20 kW of PV power generation and 10 kWh of storage batteries. No profit was obtained, including the initial cost.

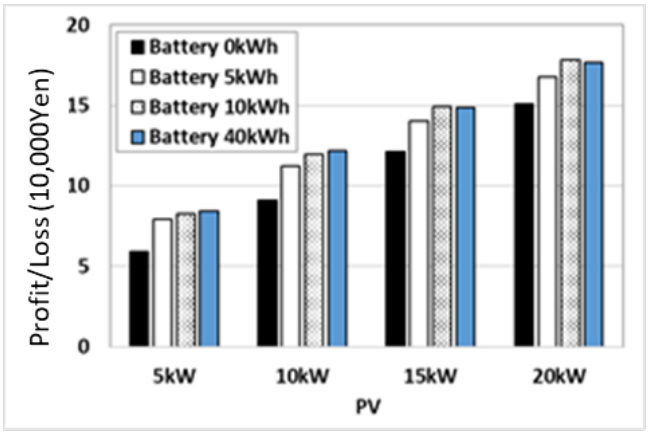


Figure 6: Profits and losses from electric power trading (Evening model)

The 15-year gains and losses for the nighttime concentration model are shown in Fig. 7.

Under the conditions of this paper, it would be more economical not to install the system. If the unit cost of installing a PV power generation system is 330,000 yen, the Night/Midnight model with 5 kW of PV power generation can reduce electricity rates by 0.4 million yen from the conventional electricity rate plus the basic rate.

5.3 Profit/Loss at the Time of Electricity Price Hikes

In the future, the unit price of electricity is expected to increase due to the rising price of fossil fuels.

In this section, we present simulation results assuming electricity unit prices soar up to 10, 30, 50, and 100%. Below are the combinations of systems that resulted in profits when losses were less than without the systems installed.

A 10% spike in the price of electricity would be profitable when only 5 kW of PV power is installed, a 30% spike would be profitable when 5 kW of PV power and an EV is substituted as a storage battery, a 50% spike would be profitable when 5 kW of PV power and 5 kWh of storage batteries are combined, a 100% spike would be profitable when only 10 kW of PV power is installed, and a 100% spike would be profitable when 5 kW of PV power and 10 kWh of storage batteries are combined. Table 5 shows the system combinations that are profitable in the Night/Midnight model due to the increase in the unit price of electricity.

Assuming that the price per unit of electricity spiked to 100%, installing only 5kW of PV power generation would be the most effective.

When we compare PV 5kW + storage battery 5kWh and PV 5kW + EV 40kWh at the price rises to 50%, the latter is 207,000 yen cheaper to install, but even after subtracting 207,000 yen from the latter's profit, the latter's profit is still 18,000 yen higher. This confirms the effectiveness of the introduction of a large-capacity energy storage system or EV.

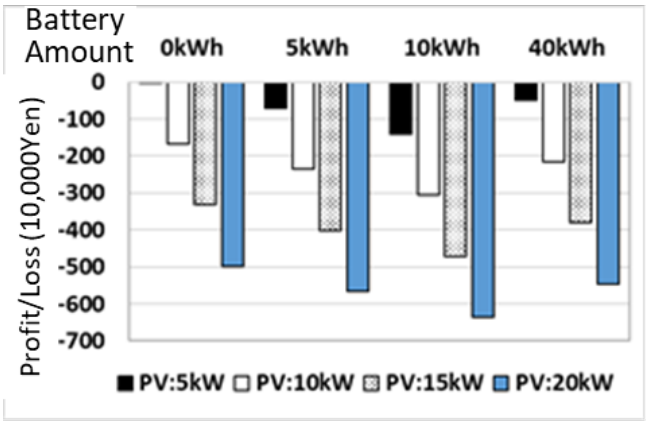


Figure 7: 15-year profits and losses (Night/Midnight model)

6 ELECTRICITY DISTRIBUTION TO MULTIPLE HOUSEHOLDS

6.1 Multi-Households Model

In this chapter, we assume a situation such as a community with several households equipped with PV power generation equipment and energy storage systems in the neighborhood by using electricity exchange among them to identify how electricity sharing resolve the electricity shortage. The multi-households model can also be used to simulate diverse and realistic electricity transactions by changing the electricity consumption patterns and the scale of the PV power generation and storage systems for each household. Figure 8 shows the electricity exchange model for multi-households' electricity exchange transactions.

6.2 Overview of Multi-Households Electricity Sharing Simulator

We simulate the flexible distribution of electricity among multiple households by running multiple simulators simultaneously and having each household disclose its electricity status and trade electricity through communication. In addition, we consider the power sharing network among multiple households as a single community and evaluate large-scale power sharing among communities with different power consumption patterns and power generation patterns.

Table 5: Rate of increase in unit price of electricity and profitable system combination

Price increase rate	System Configuration	Installation Cost (million Yen)	Profit/Loss (Yen)
10%	PV5kW	0.168	147,000
30%	PV5kW	0.168	493,000
	PV5kW+BT40kWh (EV	2.509	31,000
50%	PV5kW	1.68	839,000
	PV5kW+BT40kWh (EV	2.509	384,000
	PV5kW+BT5kWh	2.716	164,000
100%.	PV5kW	1.680	1,703,000
	PV5kW+BT40kWh (EV	2.509	1,269,000
	PV5kW+BT5kWh	2.716	1,040,000
	PV5kW+BT10kWh	3.416	355,000
	PV10kW	3.360	54,000

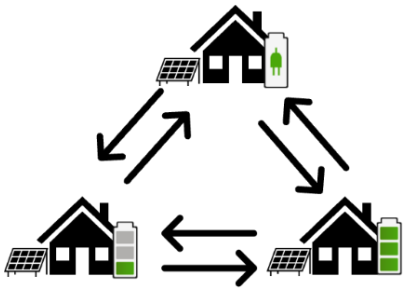


Figure 8: Multi-Households Electricity Sharing Model

In addition to the functions of the ZEH economic evaluation simulator, the following functions in Table 6 are implemented to simulate electricity sharing among multiple households.

Details are explained in Section 6.4.

6.3 Electricity Calculation Method for Electricity Sharing

When conducting electricity sharing among multiple households, the amount of electricity that each household wishes to trade is determined, and simulations are conducted in a form that more closely resembles realistic transactions. As we explain in Section 3.4, the energy consumption used in the simulation is classified into three patterns. These power patterns were mixed in the community, and the simulator is operated. In addition, a minimum power configuration that assumes facilities that do not consume power but only generate, store, and supply power is also included.

6.4 Multi-Households Electricity Sharing Algorithm

Figure 9 shows the algorithm outline of the multi-households' electricity sharing algorithm. The left process shows the requester and the right one is the responder.

Table 6: Additional Function

implementation function	
Household-to-household communication capability	Reproduction of power trading
Algorithm for determining the amount of electricity during power sharing	Process for determining content at the time of transaction

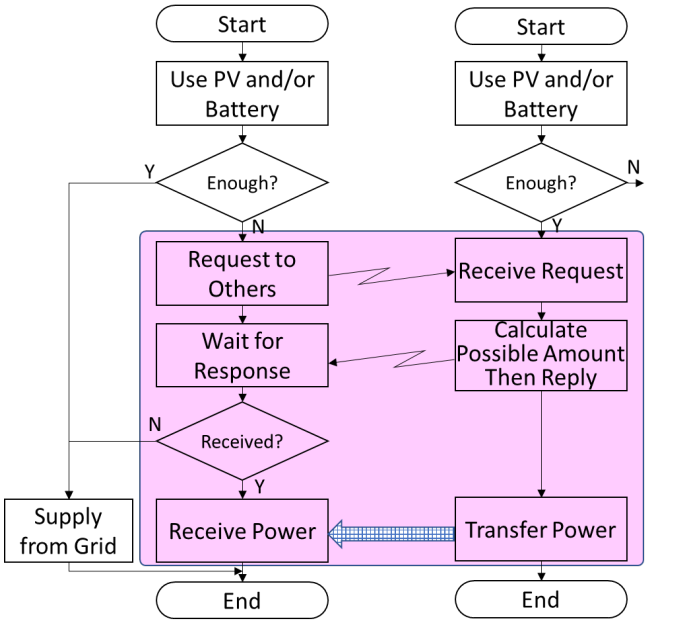


Figure 9: Multi-Households Power Sharing Simulator Algorithm Outline

Households whose remaining battery capacity has been reduced below a certain level due to electricity consumption are approached by other households to sell their electricity. The households with sufficient remaining battery capacity present their tradable electricity to the electricity-deficient households. Prospective buyers adopt the optimal power trading details based on the offers from the responding households and charge their batteries until a certain capacity is met or the trading partner ceases to exist. If the battery cannot be recharged, electricity is purchased from the power company to meet the remaining demand.

7 SIMULATION OF ELECTRICITY DISTRIBUTION AMONG MULTI-HOUSEHOLDS

7.1 Two Cases for Simulation

First, we set up a household model case for one-to-one electricity sharing among households in a community and conducted simulations according to this setting.

Table 7 shows the electricity consumption plans and household equipment configurations used in the simulations. The reason why we choose Morning/Evening and Night/Midnight households is that their average daily power consumption is complementary.

In the first case, we simulate the storage battery capacity and PV generation capacity settings that could realistically be installed by an average household at this moment.

In the second case, we assume a scenario in which the benefits of installing large-scale facilities increase due to the future low cost of energy storage and power generation equipment and soaring electricity prices and set the scale of facilities to be installed by Household-2 to a large scale.

7.2 Multi-households' Simulation Results and Evaluation

First, simulations are conducted for the first facility case. The simulator is operated from July 6 to July 13, when the supply to other households by PV power generation could be verified, to confirm that sharing is achieved. Figure 10 shows the electricity demand-supply ratio for Household-1.

On days when PV power generation is high during the day, electricity is supplied from the storage battery in Household 2, indicating that peak-time electricity demand could be met through sharing. However, on many days, the grid power supply is still relied on.

Table 8 shows the amount of electricity purchased and the consumption rate of PV power generation compared to the case without electricity sharing.

The results show that the consumption rate of PV power consumption decrease, and the amount of electricity purchased from the grid increases because of the electricity sharing. These results indicate that even small-scale facilities can consume a portion of the demand through electricity sharing during the season when PV power generation can be expected.

The simulation is then performed in the Case 2 setup, with the following electricity demand-supply ratios for Household 1 for July 6~July 13 as shown in Fig. 11.

Table 7: Case1 Parameters

Case 1	Household-1	Household-2
Electricity Consumption Plan	Morning/Evening	Night/Midnight
Storage battery capacity (kWh)	5	5
PV power generation equipment (kW)	5	5
Case 2	Household-1	Household-2
Electricity Consumption Plan	Night/Midnight	Morning/Evening
Storage battery capacity (kWh)	5	15
PV power generation equipment (kW)	5	15

Table 8: Case 1, Electricity Purchased and PV power Consumption Rate for Household 1

	One household	Two households Sharing
PV power consumption rate (%)	54.8	52.0
purchased from the grid amount (kWh)	46.3	50.0

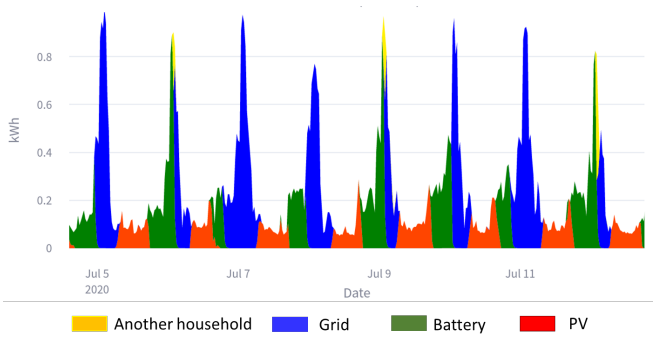


Figure 10: Case1: Household-1 Summer Electricity Supply Ratio

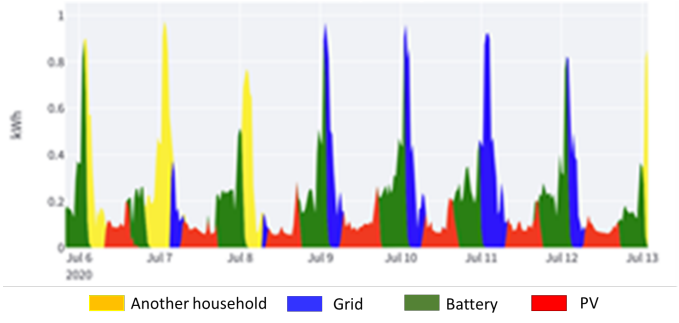


Figure11: Case2, Household1 summer electricity supply ratio

Compared to Fig. 10, the period during which the nighttime peak demand can be met only by storage battery discharge and discharge from household-2 is longer. By differentiating the amount of system installation among households, the sharing amount of electricity supply by households that could afford it increased, resulting in a change in the electricity supply ratio. Table 9 shows the changes in the amount of electricity purchased and the PV power consumption rate compared to the case without electricity sharing.

The PV power consumption rate increase by about 20% compared to when only one household is operating. The amount of electricity purchased from the grid also decrease. It was found that increasing the scale of the installed system in some households have a significant impact on the PV power consumption rate in other households.

8 SIMULATION OF MULTI-COMMUNITY ELECTRIC POWER SHARING

8.1 Overview of Electricity Transactions between Communities

We assume that there are communities with multiple households in the vicinity to trade electricity. Since the electricity storage pattern by PV power generation differs from community to community (Fig. 12), a stable electricity supply in winter, which is difficult to achieve in a simulation of only one community, can be expected.

8.2 Smart Marginalized Communities

In this study, we define "Smart Marginalized Community" as a community in a smart grid system that achieves net-zero energy in mountainous areas and does not rely on external power supply.

In this simulation, we are working to achieve a form of energy self-sufficiency by producing electricity and heat for their own consumption, called net-zero energy. This study aims to "achieve complete net-zero energy by means of renewable energy (mostly PV energy) and energy storage systems that do not rely on external power," and hypothesizes that it is possible to achieve these goals and realize smart marginal community. We then verify the configuration of the smallest marginal community that can satisfy the conditions and the combination of power generation and storage systems.

The optimal configuration and changes in the degree of net-zero energy realization is also examined when considering the realization of net-zero energy in regions with different demand and generation conditions.

Table 9: Case 2: Electricity Purchased and PV Power Consumption Rate for Household-1

	One household	Two households Sharing
PV power consumption rate (%)	54.8	73.0
purchased from the grid amount (kWh)	46.3	27.6

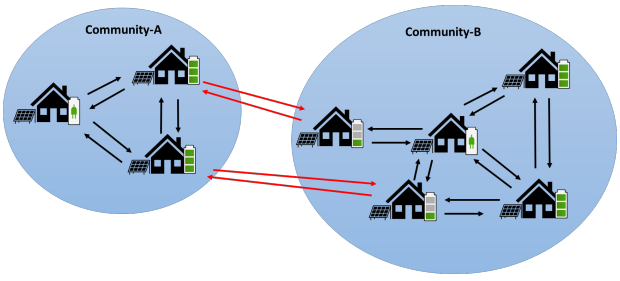


Figure 12: Model of inter-community electricity sharing

8.3 Model Case in Two Communities

A model case is then set up for two neighboring communities to share electricity with each other, and simulations are conducted according to this set-up. Table 10 and 11 show the electricity consumption plans and household facility configurations used in the simulations.

8.4 Multi-Community Simulation Results and Evaluation

The simulator is operated for one year. Figure 13 and Fig. 14 show the percentage of electricity demand supplied by Household-1 and the amount of electricity supplied by Facility 1 during the period from July 1 to July 31.

The results in Fig. 13 show that household 1 receives a large amount of electricity from Facility 1 and also receives electricity from Facility 2. Table 12 shows the amount of electricity purchased and the consumption rate of renewable

Table 10: Case 3: Community-1 Parameters

Community-1			
Case 3	Household-1	Household-2	Equipment-1
Electricity Consumption Plan	Night/Midnight	Morning / Evening	a little
Battery capacity (kWh)	5	5	15
Solar power generation (kW)	5	5	15

Table 11: Case 3: Community-2 Parameters

Community-2			
Case 3	Household-3	Household-4	Equipment-2
Electricity Consumption Plan	Night/Midnight	Morning / Evening	a little
Battery capacity (kWh)	5	5	15
PV generation (kW)	5	5	15

energy by Household 1 compared to the case without electricity supply.

The consumption rate of renewable energy is significantly increased, and the amount of electricity purchased from the grid is also reduced through the sharing power supply system. In addition to the intra-community transactions, the power acquired through transactions with power generation facilities in other communities is considered to have enabled more stable operation that is not dependent on grid power.

Figure 15 shows the relation between the amount of equipment and the amount of electricity transactions. From Fig. 15, we can recognize that increase in battery capacity can reduce both the amount of electricity discarded and the amount of electricity purchased from the grid.

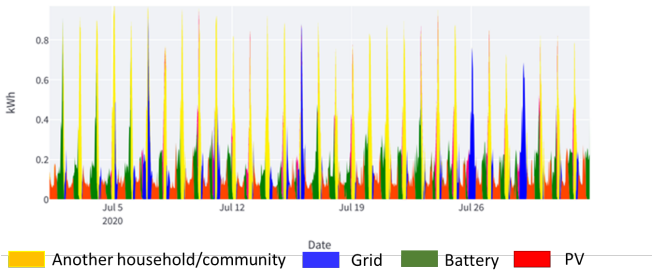


Figure 13: Household1 Summer Electricity Supply Ratio

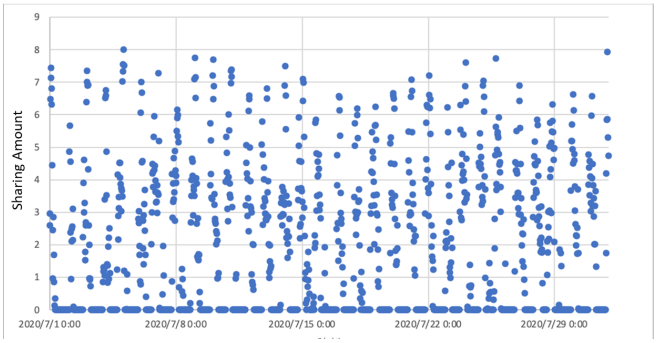


Figure 14: Amount of Electricity Sharing for Equipment-1

Table 12: Case 1, Electricity Purchased and Renewable Energy Consumption Rate for Household 1

	One household	Inter-community Sharing
PV power consumption rate (%)	37.3	54.0
purchased from the grid amount (kWh)	3523	2484

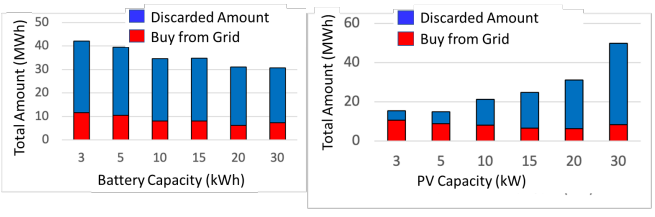


Figure15: Amount of Capacity and Grid Power Transactions

However, the effect gradually decreases when the capacity is increased beyond 20 kWh, and little change in the total amount of electricity was observed between 20 kWh and 30 kWh.

In the verification by changing the PV power generation facilities, the amount of electricity discarded tended to increase as the power generation facilities become larger, resulting in a large increase in the total amount of electricity consumed. However, the amount of electricity purchased from the grid decrease in a negative correlation with the size of the facility, resulting in an electricity consumption of 5 kWh less than the total electricity consumption for a facility with a capacity of 3 kWh. This result suggests that a smart marginal community that does not rely on buying and selling power from the grid can be realized by reducing the size of PV power generation facilities to a size appropriate for the scale of the community and increasing the size of storage batteries.

8.5 Seasonal Changes in PV Power Consumption Rates

In order to investigate the factors that reduce the PV power consumption rate, the PV power consumption rate is calculated for each of the following seasons: winter (January to March), spring (April to June), summer (July to September), and fall (October to December).

We set up three households each at two communities and all households equip 5kW PVs and 20kWh storage batteries. Table 13 shows each household consumption model.

Table 14 shows the seasonal PV power consumption rates.

Table 14 shows that the PV power consumption rate is close to 100% in spring and summer, and that electricity consumption can be achieved without power supply from the grid. However, in the fall season, the consumption rate is in the 70% range, and in the winter season, the consumption rate drops to the 50-60% range. This is due to the decay of PV power generation and the increase in electricity consumption.

The amount of solar power generation decreases during the winter season. In winter, weather conditions tend to

Table 13: Consumption model for seasonal evaluation

Community-1	
Household-1	Night/Midnight
Household-2	Evening
Household-3	Morning/Evening
Community-2	
Household-4	Daytime (every time same)
Household-5	Evening
Household-6	Morning/Evening

Table 14: Seasonal PV power consumption rates

	H-1	H-2	H- 3	H- 4	H- 5	H-6
winter	53.6	63.8	62.5	68.8	62.8	62.8
spring	100.0	100.0	100.0	100.0	100.0	100.0
summer	100.0	100.0	100.0	100.0	99.8	99.8
fall	71.4	77.7	73.8	72.2	77.7	73.8

worsen, and when the PV panels are covered with snow, power cannot be generated. If the inability to generate electricity persists for several days, the grid power supply will have to be relied on if PV power generation and storage batteries are the only means of supplying electricity, resulting in a decrease in PV power consumption. Therefore, in order to further increase the average rate, it is necessary to solve the energy problem, especially heating, during the fall and winter seasons with renewable energy sources that do not rely on PV power.

9 CONCLUSION AND FUTURE WORKS

In this research, we have developed a simulator to evaluate the economic efficiency of ZEH. The simulator is useful for each household to consider the installation of equipment. The simulator can be operated through a web interface, so it can be operated by many people.

If economics is the only consideration when installing PV power generation and storage systems, it is best to refrain from installing PV power generation and storage systems. The introduction of these systems should be based on the assumption that losses will be accepted. However, we have shown that it is possible to profit from the introduction of PV and EVs when electricity prices soar.

In addition, a simulator was improved to investigate the feasibility of net-zero energy by integrating electricity among multiple households or communities.

In a one-on-one simulation assuming intra-community transactions, the amount of electricity purchased decrease, and the PV power rate increase when the storage batteries and capacity of one household are increased, while the demand during peak hours is only partially compensated by the sharing of the other household when the size of the storage batteries and PV generation facilities of both households are small.

In the case of transactions in which the scale of transactions is expanded to multi-communities, by scattering large facility households with different electricity consumption styles, the surplus electricity of large households was passed on to small facility households, and as a result, a high renewable energy rate is confirmed.

One of the challenges for the future is to change the electricity transaction price, which is fixed within and between communities, using an algorithm. In addition, because electricity consumption is high during the winter season, the smart marginal communities could not be realized, so it is necessary to derive constraints on environmental changes, such as improvements to power generation and storage facilities and price decreases.

ACKNOWLEDGEMENT

This work was partially supported by Grants-in-Aid for Scientific Research(C) numbered 21K11856.

REFERENCES

[1] Agency for Natural Resources and Energy, “Actions toward the end of the FIT purchase period for residential photovoltaic power generation equipment,” (2019).

[2] Agency for Natural Resources and Energy, “Information Disclosure on ZEH (Net Zero Energy House) - Energy-Saving Housing,” https://www.enecho.meti.go.jp/category/saving_and_new/saving/general/housing/ index03.html, (accessed July 26, 2022).

[3] New Energy and Industrial Technology Development Organization, “NEDO Renewable Energy Technology White Paper, 2nd Edition,” (2014).

[4] Agency for Natural Resources and Energy, “Recent Trends in Electricity Supply and Demand and Wholesale Electricity Market,” https://www.meti.go.jp/shingikai/enecho/denryoku_gas/denryoku_gas/pdf/046_03_02.pdf, (accessed July 15, 2022).

[5] Agency for Natural Resources and Energy, “Actions Toward the Formation of Next-Generation Distributed Energy Systems,” https://www.meti.go.jp/shingikai/enecho/denryoku_gas/denryoku_gas/pdf/046_04_02.pdf, (accessed July 15, 2022).

[6] Agency for Natural Resources and Energy, “A Guide for Establishing Regional Microgrids,” https://www.meti.go.jp/shingikai/energy_environment/energy_resource/pdf/015_s01_00.pdf, (accessed July 20, 2022).

[7] R. Ishikawa and H. Matsuo, “Evaluation of Energy and Economy of Battery System Installation in residential PV System after FIT Expiration,” Journal of Japan Solar Energy Society, Vol.46, No.5, pp.63-70 (2020).

[8] K. Yabe and Y. Hayashi, “Economic Evaluation of Increased In-House Consumption of Photovoltaic Power Output by Storage Batteries,” IEEJ Transactions B (Journal of the Electricity and Energy Division), Vol. 139, No. 5, pp. 363-371 (2019).

[9] R. Goto, H. Takano, and H. Asano, “Scheduling Method of Microgrids’ Operation Considering Uncertainty in Values of Forecasted Net Load,” Transactions of the Japan Society of Energy and Resources, Vol. 42, No. 1, pp. 11-17 (2021).

[10] R. Komiyama, S. Shibata, and Y. Fujii, “Analysis on Optimal Power Supply Structure Considering Output Fluctuation of Photovoltaic and Wind Power Generation and Interregional Electricity Flexibility,” IEEJ Transactions on Electric Power Industry B Vol. 133, No. 3, pp. 263-270 (2013).

[11] E. Fukuda et al, “Construction of Energy IoT Data Collection System (2)”, Information Processing 82nd National Convention, (2020).

[12] Committee for Research and Study on Energy Consumption in Houses, Architectural Institute of Japan, “Database of Energy Consumption in Houses,” http://tkkankyo.eng.niigata-u.ac.jp/HP/HP/database/index.htm, (accessed July 1, 2021).

[13] Nichicon Corporation, “Product Specifications/EV Power Station/Grid-connected V2H System,” https://www.nichicon.co.jp/products/v2h/about/specs/, (accessed Jan. 10, 2022).
[14] Hokuriku Electric Power Company, “Hokuriku Electric Power Company Customers at Home (unit price of electricity),” https://www.rikuden.co.jp/ryokin/minsei.html, (accessed Dec. 3, 2021).

(Received: Novemver 19, 2023)
(Accepted: February 15, 2024)

Masashi Saito received B.E. degree from Tokyo Institute of Technology in 1983, M.E. degree from Cornell University in 1992 and Ph.D. degree from Osaka University in 2006. In 1983, he joined Mitsubishi Electric Corporation and has developed engineering workstations, Internet TVs, cellular phones, car navigation systems especially for operating system extension, Internet services and distributed processing. In 2006, he moved to Mitsubishi Electric Research Laboratories as a senior principal technical staff. Since 2015, he has been a professor at College of Informatics and Human Communication, Kanazawa Institute of Technology. His current research interests include ITS, distributed systems, smart communities and smart social systems. He is a member of IEEE and IPSJ.

Tomoki Nomura received Master of Engineering from Graduate School of Engineering, Kanazawa Institute of Technology. He joined INTEC Inc. in April 2023.

Yuichi Tokunaga received a ph.D. degree in science and engineering from Ritsumeikan University in 2009. He joined Mitsubishi Electric Corporation in 1990 and engaged in R&D of the high-reliability computer, wireless sensor network, the algorithms of time-synchronization and positioning, the network protocol for industrial applications, and data analytics for Condition Based Maintenance. He has been a professor at Kanazawa Institute of Technology since 2019. He is a member of the ITS steering committee of IPSJ and a member of IEEE and ISCIE.

Regular Paper

Identifying Risk Factors for Jaywalking Using a Risk Mapping Framework

Yuichi Tokunaga*, Atsushi Yamamoto**, and Masashi Saito*

*Faculty of Graduate School of Engineering, Kanazawa Institute of Technology, Japan
**Department of Management Systems, Kanazawa Institute of Technology, Japan
{y.tokunaga, msaito}@neptune.kanazawa-it.ac.jp

Abstract -Crossing outside of a crosswalk or ignoring a traffic signal is called jaywalking. Jaywalking is a major road safety issue because even vehicles with safety features have difficulty predicting jaywalking suddenly crossing the road from the shadow of an obstacle. We hypothesize that the possibility of jaywalking is related to the road and its surrounding environment and are investigating ways to statistically estimate the risk of jaywalking from map information. To realize these ways, it is essential to identify risk factors for jaywalking, but there are a wide range of candidate risk factors, and an efficient method to identify them is needed. Therefore, we have developed a framework to generate a wide-area risk potential map from local observations and attempted to identify risk factors. The risk map framework helped streamline the procedures from factor hypothesis to verification and analysis, and we gained new insights into risk factors.

Keywords: Jaywalking, Risk map, Bayesian network.

1 INTRODUCTION

In recent years, the number of accidents involving automobiles has been on a downward trend due to the improvement of the road traffic environment, the spread of traffic safety awareness, and improvements in safety performance of vehicles. However, the number of accidents involving people versus vehicles has remained flat in recent years [1]. In particular, the number of people killed in accidents while walking continues to be the leading cause of death by condition and has begun to increase slightly. In addition, crossing violations and disregard for traffic signals account for more than 45% of the causes [2]. These dangerous behaviors of pedestrians are called "jaywalking" and countermeasures against them have become an issue in road traffic, including overseas [3]. It is assumed that elderly pedestrians jaywalk due to the overconfidence brought on by years of experience despite their declining physical ability and sensory perception. Although automobile pedestrian detection has improved over the years, the braking distance is insufficient for an emergency stop against a Jaywalker emerging from between vehicles in the oncoming lane. Prioritizing safety requires a significant reduction in speed, resulting in traffic congestion. As a result, it is difficult to avoid jaywalkers who may cross from anywhere at any time.

Common countermeasures against jaywalking include the installation of anti-crossing barriers (guardrails) and regulatory signs. Anti-crossing barriers are installed between the roadway and sidewalk to physically deter people from crossing and are considered to be an effective measure to prevent jaywalking. Regulatory signs can be installed in various locations due to their ease of installation, but they are not very effective, and neither of them is a fundamental solution to the problem.

Therefore, instead of eliminating jaywalking, we should consider foreseeing pedestrian's jaywalking and avoiding them on the vehicle side. As mentioned above, it is difficult for vehicles to be prepared for jaywalking from all points, but by identifying points with a high probability of jaywalking, it is possible to reduce speed only there, thereby achieving both safety and avoidance of traffic congestion. Based on this idea, this paper aims to statistically determine the points of high risk of jaywalking and presents a method to identify the risk factors.

Section 2 introduces related works on jaywalking countermeasures, Section 3 describes the risk map framework we constructed to efficiently evaluate risk factors derived from these works. Section 4 uses this framework to obtain specific jaywalking risk maps, and Section 5 evaluates the results. Finally, Section 6 provides a summary and future outlook.

2 RELATED WORK

Since more than 70% of those killed in walking accidents are elderly with older than 65 years old, the behavioral psychology of road crossing for the elderly has been actively studied [2]. According to an analysis of accidents involving elderly people crossing the road by the Traffic Accident Analysis Center [4], 51% of elderly people who had accidents while crossing the road did not notice the approach of a car because they did not check for safety or check insufficiently. A typical accident scenario is shown in Fig. 1. The pedestrian first checks the path to be crossed. At that time, they see an A1's car approaching in the lane in front of them and wait for

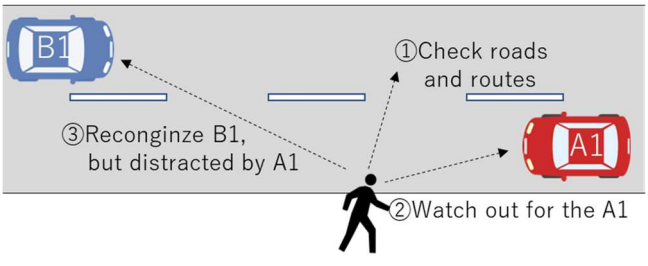


Figure 1: Mechanism of accident while crossing.

the car to pass by. They also confirm the presence of a B1's car on the opposite lane, but their attention is directed to A1. They start crossing immediately after A1 passes, continues crossing without checking left and right directions, and crosses without noticing B1 approaching from the opposite direction.

Technology to detect and avoid invisible pedestrians is being considered as the features of Advanced Driver Assistance System (ADAS) for automobiles [5]. For example, research is being conducted on technologies to recognize the movement of invisible objects from images reflected in curved mirrors and reflective objects [6][7], and to share sensing information of other vehicles, GPS information of pedestrians, and image or sensing information from infrastructure facilities through V2V, V2P, and V2I communications respectively and to detect the danger range [8]. However, all these technologies are dependent on outdoor reflective objects, sensors and cameras, and pedestrian communication devices, and thus lack versatility and do not provide fundamental solutions to the problem.

The use of risk potential maps has been proposed to reduce accidents caused by unseen hazards [9]. A risk potential map shows not only the hazards physically present in the driving path, but also the degree of danger in a space based on past cases and probability, using numerical values and colors like contour lines. For example, an intersection without a signal, which is a blind spot at a wall, has a higher danger level because of the risk of ejection. The risk map generated in this way is used to realize safe driving by presenting the driver using Head-Up-Display (HUD) and Augmented Reality (AR) or by referring in the automatic driving control to plan the low-risk routes or to control speed. Several studies and demonstrations have been conducted, but the challenge is how to narrow down the risk; if the risk map is applied to jaywalk, generalized to roads without pedestrian crossings, the jaywalk risk is high, and cars are always forced to drive at low speeds. On the other hand, if only points where jaywalking has occurred in the past are considered high risk, accidents caused by jaywalking at other points cannot be prevented. It is necessary to narrow down the risk value with more accurate information.

On the other hand, there have been many studies on pedestrian crossing behavior from the viewpoint of road design in the field of civil engineering planning [10][11]. Hamamoto et al. studied pedestrian behavior at a national road intersection leading to a station connection [12] and explained that pedestrians want to take the shortest route possible, and if there is a possible crossing on the route, they will cross even if it is outside the pedestrian crossing area. The study notes that jaywalking is more likely to occur at locations where crossing cannot be physically deterred, such as street intersections, entrances to roadside facilities, and openings in anti-crossing barriers. Takehira also studied the relationship between roadway facilities and jaywalking on several roads in densely populated urban areas [13]. Unlike national roads, anti-crossing barriers on urban streets are often characterized by disconnection at the approaches to stores and parking. In all the study locations observed, more than 90% of the disorderly crossings occurred at locations where there were openings at both the start and end points.

These studies indicate that the likelihood of jaywalking occurring depends on the condition of the road environment surrounding the pedestrian, such as intersections and anti-crossing barriers. Studies have been published that focus on this point and incorporate it into risk assessment for pedestrians. Wang et al. propose a risk assessment method for pedestrians who jump from the shadow of a car onto the roadway for automated driving on public roads [14]. They used Dynamic Bayesian Networks (DBN) to determine the probability that a pedestrian exists in the shadow of a car and that the pedestrian will enter the roadway. They hypothesized that pedestrian behavior of entering the roadway was due to the road and traffic environment and attempted to identify the risk factors by observing crossing behavior. Observations showed a link between the road environment and disorderly crossing, but the proposed risk assessment method was only validated in a computer simulation and not evaluated in a real environment.

Thus, it is shown that the potential risk of jaywalking may be inferred from the road environment, and anti-crossing barriers are an important clue in estimating jaywalk risk. On the other hand, many community streets and suburban roads are not equipped with anti-crossing barriers, and the presence or absence of such barriers alone cannot narrow down the potential risk area of jaywalking. It is essential to discover environmental impact factors other than anti-crossing barriers for risk mapping that aims at both safety and traffic efficiency.

3 RISKMAP FRAMEWORK

3.1 Target Setting

Related works have shown that the opening of an anti-crossing fence affects the probability of jaywalking occurring. Other road traffic environment influences have also been suggested. So, as shown in Fig. 2, if we create a model to estimate the probability of jaywalking based on these road environment characteristics in a certain limited area and expand it to the whole country using map information and this estimation model, we can create a general jaywalking risk

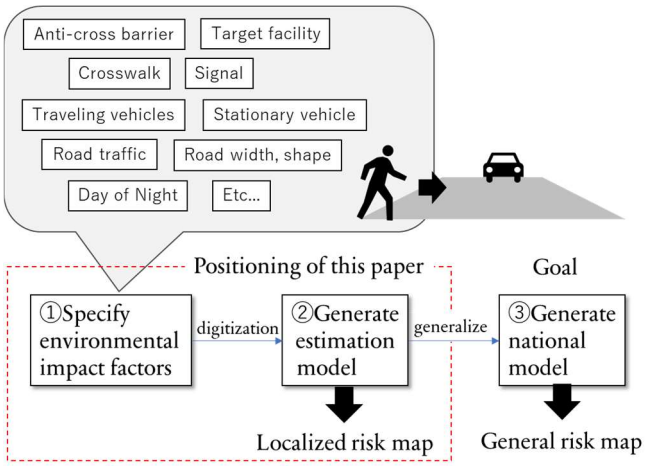


Figure 2: Approach to create a jaywalking risk map.

map. However, to narrow down the risk area, it is necessary to specifically identify features other than anti-crossing barriers.

Therefore, we formulate a hypothesis based on use cases. Facilities where people gather, such as train stations and schools, are often located on main streets, and pedestrians targeting the facility flow into the main street from side streets near the facility, bypassing nearby pedestrian crossings to follow them to the facility. If there are no crosswalks nearby, the behavioral psychology of choosing the shortest route is at work, and pedestrians are more likely to cross at the point where they exit the side street onto the main street. Based on this use case, it is highly likely that the mutual location of the facility where people congregate, the influx of side streets in the vicinity, and the crosswalks affect the potential risk probability of jaywalking. The goal of this study is to substantiate this hypothesis.

To achieve this goal, we will generate a jaywalking estimation model based on the location of facilities and other features and crosswalks and evaluate whether the risk map derived from the model is valid outside of the observation points.

3.2 Framework Construction

Creating a jaywalking estimation model requires the development of factors such as facility type and feature extraction, as well as observational data. On the other hand, whether an appropriate risk map can be obtained depends on the hypotheses based on use cases, and there is no guarantee that the results will be worth the effort of creating the model. Therefore, the initial stage is to generate a model from a minimal dataset to prove the hypothesis, the process of expanding the missing variables and data is repeated. In this iterative process, as shown in Fig. 3, the collection of observation data, generation of data sets, model generation, and risk visualization can be automated, and therefore, these can be built as a framework to improve efficiency.

An overview of the framework is shown in Fig. 4. First, location information of features in the road environment that induce jaywalking, such as facilities and pedestrian crossings, are automatically obtained using open data. Next, location dataset is created, and an estimation model is generated. Utilizing the created estimation model, the probability of occurrence of jaywalk is estimated. Finally, a heat map is superimposed on the map using the occurrence probability as the risk value to create a jaywalking risk map. Each of these functions is described below.

3.2.1 Extraction of Location Information on the Road Environment

Google Maps is used to automatically acquire location information of features on the road environment that trigger jaywalk. Facilities such as stations and schools are listed by the search function and their latitude and longitude information is obtained. On the other hand, pedestrian crossings cannot be extracted by the search function. As a method to address this problem, a study has been reported that analyzes aerial photo images to detect pedestrian crossings

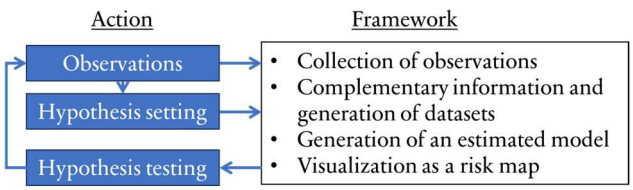


Figure 3: Positioning of the framework.

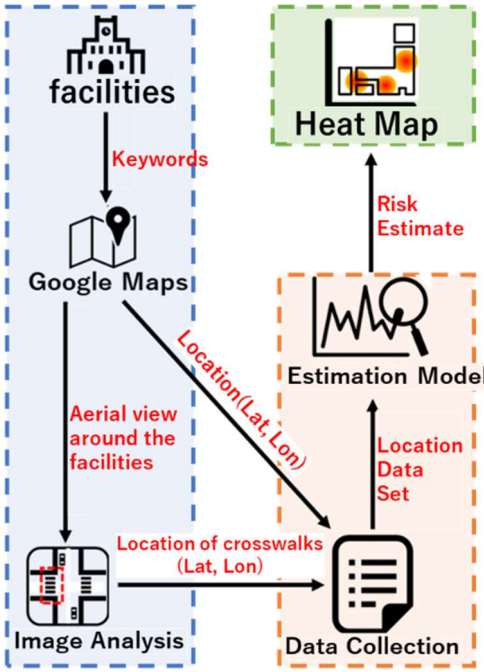


Figure 4: Framework overview.

[15][16], and we will use the same method. The method of extracting crosswalks is explained in 4.1.2.

3.2.2 Generating an Estimation Model

Next, a location dataset is created from the calculated latitude and longitude information. The presence or absence of jaywalking is observed in advance from visual surveys, and the data is created by calculating the distance and direction for each latitude and longitude of the observation location. The created data is compiled as location data, and an estimation model is created using a Bayesian network.

Bayesian network is a probabilistic inference model that shows multiple causal relationships in a weighted graph structure (network) and indicates each causal relationship by probability [17]. Even when the full picture of causal relationships is not known, the estimation accuracy can be improved by accumulating findings obtained through observation as partial causal relationships. We believe that this is the best method for this project to clarify the causal relationship of the jaywalking outbreak through repeated hypotheses and observations.

Bayesian network can be expressed by the following equation:

$$P(x_1, \ldots, x_N) = \prod_{i=1}^{N} P\left(x_i | parent(x_i)\right) \qquad (1)$$

where x_i denotes each event and $parent(x_i)$ denotes the upper event in the relationship. The probability of occurrence of a given combination of events is expressed as the product of the conditional probabilities of each causal relationship.

From the hypotheses set up in this paper, x_i indicates the relationship between the location of the facility at the observation point, the road structure, and the location of the pedestrian crossing. The model is generated by data on the jaywalking situation at each observation point, analyzing the dependencies between each causal event and determining the conditional probability of each relationship.

Bayesian estimation was used to create the estimation model. Details of the location dataset and estimation model are described in 4.2.2 and 4.2.3.

3.2.3 Generating a Jaywalking Risk Map

Finally, a jaywalking risk map is created by overlaying a heat map on a map with the probability of jaywalking as the risk value. The colors displayed on the heatmap should be separated according to the risk values so that the information can be obtained visually.

4 APPLYING THE FRAMEWORK

The constructed framework is used to generate risk maps based on the risk factors hypothesized in the use cases in Section 3.1. Specifically, the latitude and longitude of structures are automatically obtained from Google Map searches to create a dataset of the location relationship between gathering places and pedestrians, and crosswalks are detected and located from aerial images to create a dataset of the location relationship between pedestrians and crosswalks. Based on these location relationships and the observed information on the availability of jaywalking, a Bayesian estimation model is created, and a jaywalking risk map is generated. This section describes the acquisition of crosswalk location information, creation of the Bayesian estimation model, and generation of the risk map.

4.1 Extracting Crosswalk Locations

4.1.1 Image Collection

Using the location data (latitude and longitude) of the facility retrieved by Google Maps, aerial images centered on the facility (or its center of gravity in the case of multiple facilities) are collected from the Google Maps Platform (hereinafter referred to as "GMP"). GMP outputs a 70-meter square aerial image by specifying the location data of the center of the image to be acquired. Assuming a 1 km square risk map, the location data sets of 15 x 15 = 225 images must be specified to GMP. To minimize gaps and overlaps at image boundaries, the location data of the center point of each image was determined using the Vincent method from the latitude and longitude of the center point of the target area, and the distance and direction of each image. These data sets were input into GMP to obtain the images.

4.1.2 Extraction of Crosswalks

Although several methods have been proposed for extracting pedestrian crossings from aerial images, we apply YOLO v5, which is widely used for general object detection applications. We trained on 300 crosswalk images extracted from aerial photographs and obtained a high detection accuracy with an F value of 0.95 by setting confidence>= 0.8.

Using the acquired crosswalk coordinate information and the latitude and longitude information of the images, the latitude and longitude of the crosswalk are calculated YOLO outputs the coordinate information of the bounding box of the detected object. Since the latitude and longitude information of the center of the original image is known, the latitude and longitude information of the center of the bounding box was calculated from this information, and this was used as the location of the crosswalk.

4.2 Creating a Bayesian Network Model

4.2.1 Collection of Training Data

To obtain the training data, a visual survey of jaywalking was conducted. The survey covered an area of 80 meters from north to south and 800 meters from east to west, centered on the main road that passes in front of the main gate of the university, and observed jaywalking for 20 minutes starting at 8:20, when students are concentrated at school. Six surveyors were assigned to the survey area to record the starting point of jaywalking and the crossing direction. In order to investigate the usual situation, the surveyors were students, and they pretended to meet up with others so that they would not be recognized as being in the middle of a survey, and no photographs were taken. Although the percentage of pedestrians killed in accidents while walking is higher among the elderly, we observed jaywalking among all age groups of pedestrians, as jaywalking is done regardless of age. The overview of the study area is shown in Fig. 5 and the time trends of observed jaywalking are shown in Fig. 6. Red boxes in Fig. 5 indicate locations where jaywalks were observed, and black boxes indicate locations where a particularly large number of jaywalks were observed. Many jaywalkers were observed at the intersection of side streets to the main street, and many pedestrians were observed jaywalking not only to and from the university but also to and from the supermarket.

4.2.2 Creation of Dataset

The dataset consists of five study area divided into a 1-meter square mesh, and four items, distance and direction to the nearest facility (in this case, a university), distance and direction to the nearest pedestrian crossing, were calculated based on the location information of facilities and pedestrian crossing groups obtained by the methods described in Sections 3.2.1 and 4.1. Note that although road structure is also a candidate for jaywalking risk factor in the hypotheses of section 3.1, it was excluded from the item because the criteria for categorization had not been clarified. This is one of the issues to be addressed in the future.

Map data (c) OpenStreetMap contributors, CC-BY-SA

Figure 5: Observation points of jaywalking.

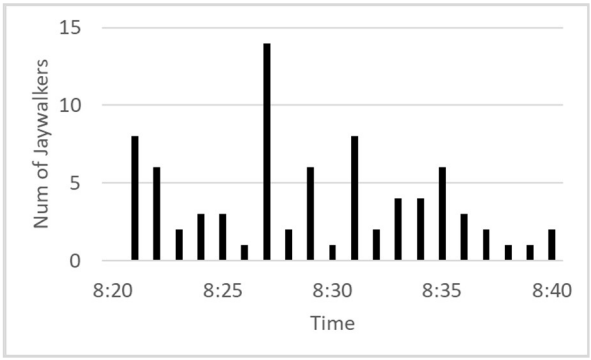


Figure 6: Trends in occurrence of Jaywalking.

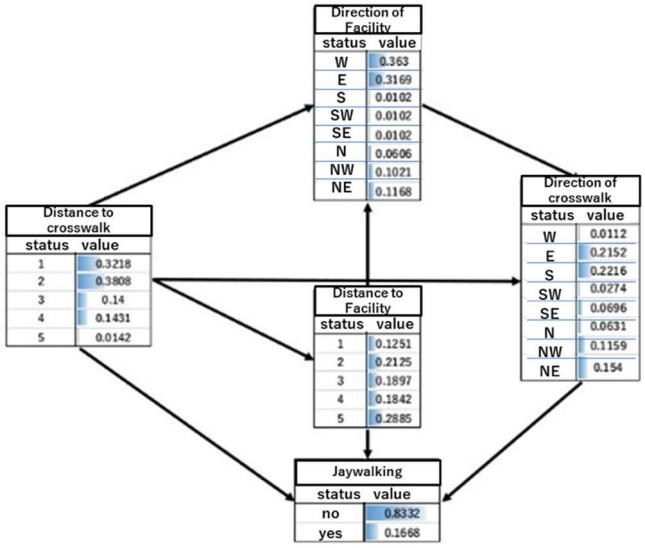
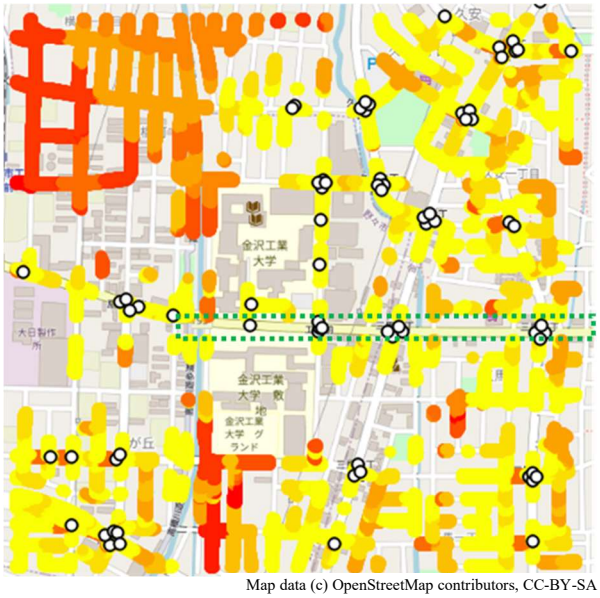


Figure7: Created Bayesian network.

The results of the visual survey are added here: the number of observations is added to the mesh corresponding to the starting point of the jaywalking. If a mesh was not observed during the 20 minutes of the survey period, it is set to 0. If multiple jaywalking was observed, the number of times it was observed is recorded. As described above, the survey results for an area of 80 meters from north to south and 800 meters from east to west are divided into 80 meshes x 800 meshes = 68,000 records, and values for 5 items in each record are generated.



Map data (c) OpenStreetMap contributors, CC-BY-SA

Figure 8: Jaywalking risk map.

4.2.3 Estimation Model Generation

BayoLinkS [18], software for building Bayesian network, was used for model generation. This software can learn structures from data, extract dependencies (Bayesian networks), and generate estimation models. Since the data needed to be discretized, the distances in the dataset were equally divided every 50 meters, and the directions were labeled in eight levels: east, west, south, north, and south. The number of divisions for discretization has a significant impact on model performance, so the appropriate number of divisions is an issue to be considered in the future.

The Bayesian network constructed by structural learning is shown in Fig. 7. The solid line shows the dependencies, and the table below the items shows the probability of occurrence of each state relative to jaywalking probability. In the current study, the distance to the facility and the crosswalk and the direction of the crosswalk are directly causally related to the occurrence of jaywalking, while the direction of the facility is causally related to the direction of the crosswalk. This is considered to be a result of the influence of the learned terrain and is a challenge when generalization, i.e., when the model is used to estimate other areas.

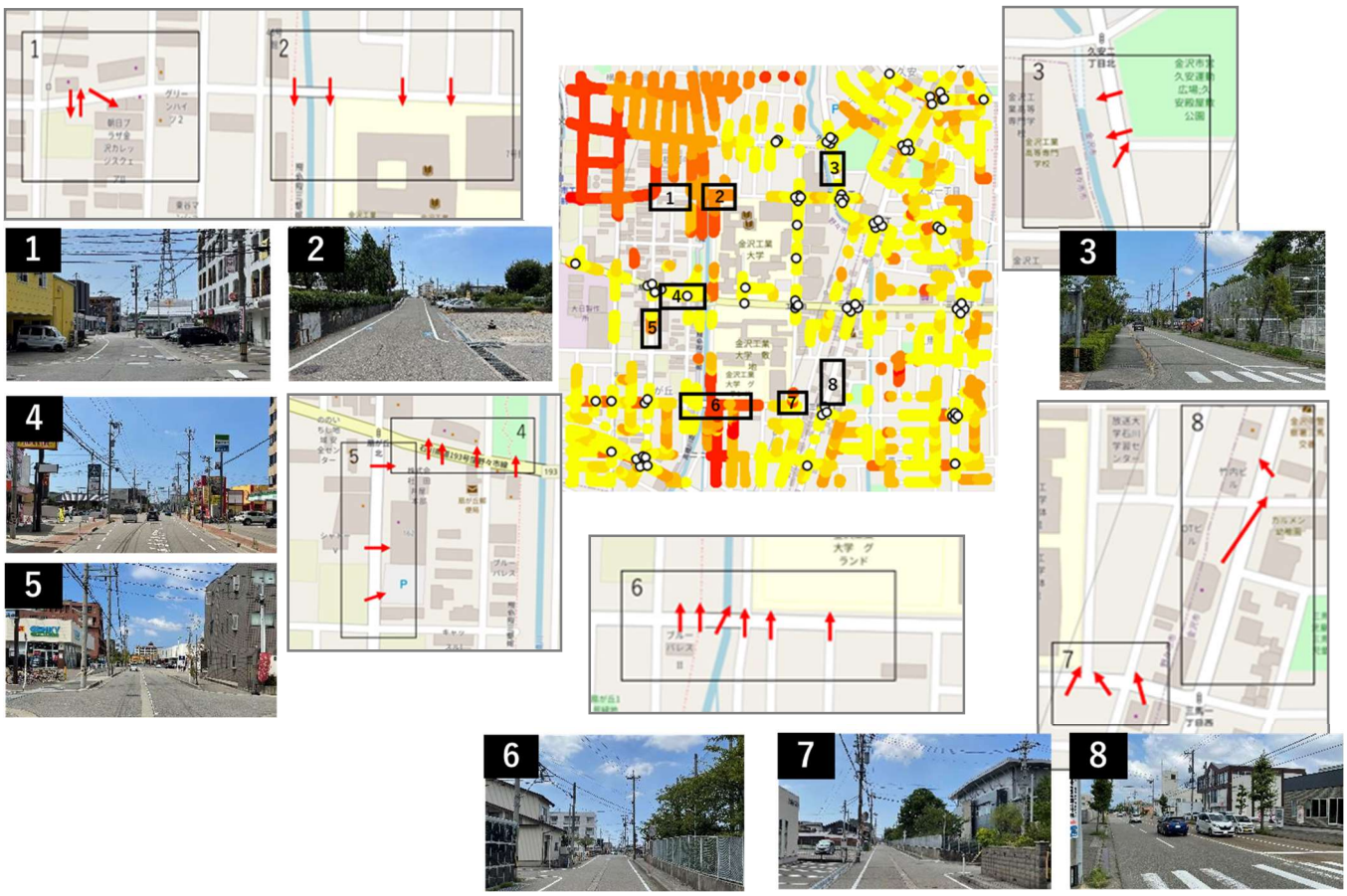


Figure 9: Observation results.

4.3 Risk Map Generation

The risk maps are generated in QGIS, using the estimation model generated by BayoLinkS to obtain the probability of jaywalking occurrence for each mesh location, which is then converted into a heat map. Since we want to show a risk map for a road, we obtain road information from OpenStreetMap and display the probability of occurrence only for the mesh location that overlaps the road. The created risk map is shown in Fig. 8. The circled dots represent pedestrian crossings.

The green boxes indicate visual survey locations, i.e., areas of correct data. The upper left corner of the map indicates a high-risk area. This area is considered high-risk because it is a residential area and there are no pedestrian crossings. On the other hand, there are no areas near pedestrian crossings that are indicated as high risk, indicating that the distance from the pedestrian crossing has a significant impact.

5 EVALUATION

5.1 Evaluation Method

To evaluate the hypothesis using the created signal neglect risk map, a visual survey of jaywalking will be conducted again. Eight sites will be randomly selected from a wider area than during the survey to obtain training data, and jaywalking will be observed for 20 minutes. We will evaluate the consistency between the frequency of jaywalking indicated by the results and the hazard level of the risk map. The location of each point is shown in Fig. 9.

5.2 Evaluation Results

Regardless of the results of the risk estimation, jaywalking was observed at all points and the hypothesized estimation results were determined to be unreliable. To clarify the problem with the hypothesis, the status of jaywalking at each point was analyzed.

Point 1: High risk determination with no pedestrian crossing in the vicinity. A convenience store fronting a two-lane road with a parking lot and café in front of it. jaywalk was identified between the convenience store, parking lot and cafe. The destination was hypothesized to be the university because the observation time coincided with the students' commute to school; however, considering that many students stop at the convenience store, the stopping establishment should be included in the destination.

Point 2: The risk map shows a high-risk level due to the lack of pedestrian crossings in the vicinity. Visual surveys

also identified jaywalking on all streets, and the risk level is correctly estimated.

Point 3: Crosswalks are located approximately 50 m apart, and the risk map estimates the risk level to be low; however, a signal failure in the direction of the entrance to the sports field was observed at the midpoint of both crosswalks. As with point 1, lack of destination setting is considered to be the cause of the discrepancy.

Point 4: Low risk determination with crosswalks approximately 20 m apart, with a convenience store and park fronting the two-lane road. Jaywalking was observed despite the proximity of the pedestrian crossing. The convenience store to stop at is on the opposite side of the road, and the straight road with good visibility suggests that the driver is crossing the street without using a pedestrian crossing. In addition to the additional destination, the factor of road visibility was found to be necessary.

Point 5: Intermediate risk determination near a signalized intersection. There is a bookstore facing a two-lane road with good visibility. Jaywalk was observed avoiding the signalized tail crossing. The situation is considered similar to the situation in point 4.

Point 6: A medium to high-risk determination on a residential street with no crosswalks. As in point 2, jaywalking across a two-lane road from each residential street is identified and the risk is correctly estimated.

Point 7: Medium risk determination near a signalized intersection. Signal failure to use crosswalk was observed on the route from the convenience store to the main gate of the University. jaywalking direction is consistent with the direction toward the University.

Point 8: A four-lane road with signalized crosswalks at approximately 100 m intervals. Low risk determination. Jaywalk was confirmed to be heading toward a family restaurant in the middle of the road. The situation is considered similar to that of Point 1.

As described above, no correlation is obtained between the risk estimate of each point and the frequency of jaywalking, and it is not possible to substantiate the hypothesis that the mutual location of the facilities where people gather, the influx of side streets around them, and the crosswalks affect the risk value of jaywalks. However, for Points 1, 3, 4, 5, 7, and 8, jaywalking targeting stores and facilities where people often stop on their way to school is confirmed, and by registering these as additional facilities that contribute to jaywalks and having the risk estimated based on the location relationship with the nearest facility, an improvement in the accuracy of the risk map This function is expected to improve the accuracy of the risk map. This functionality is an improvement to the framework and will be considered an issue to be addressed in the future.

Furthermore, for points 4 and 5, visibility is thought to induce jaywalking, indicating the need for a new factor that expresses visibility in terms of road width and straightness. For points 2 and 6, the road structure needs to be added as a factor, as the shortest route to the university is a residential street, which is blocked by a two-lane road. The developed framework does not yet support road features, but we believe it is possible to extract them from aerial photographs in the same way as pedestrian crossings, and this is another area for improvement.

6 CONCLUSION

This study attempted to identify risk factors for jaywalking in order to visualize the probability of jaywalking as a potential risk. Based on the results of related research, such as the behavioral psychology of people crossing a road and the effects of facilities that obstruct crossing, such as anti-crossing fences, it was found that the road and its surrounding environment have an influence on the occurrence of jaywalking. Therefore, we hypothesized that the mutual location of facilities where people gather, side roads and crosswalks that flow into the surrounding area affect the occurrence of jaywalking, and constructed an environment that can be used as a framework to materialize this hypothesis. Keyword facility locations from Google Maps and crosswalk locations from aerial photo image analysis were extracted to create a dataset along with Jaywalking observation data. Using this data, a Bayesian network risk estimation model was created to display the risk of jaywalking on the road in a QGIS heat map. Based on this, it is possible to efficiently evaluate whether the hypothesized risk factors can be generalized by re-observation.

The model was generated based on observations of jaywalking in the vicinity of university facilities, and this model was applied to the outer perimeter of the university to evaluate the risk map and actual conditions through observation, but the results were not as expected. After analyzing the observation results, it was found that the hypothetical problem requires the presence of multiple types of facilities. It was also found that the road structure, which has not been implemented, is likely to be an important risk factor.

Although the objective of this study to identify risk factors for jaywalking has not been achieved, the framework we have developed allows us to efficiently move on to the next hypothesis. In the future, we aim to improve the accuracy of jaywalking risk estimation by identifying risk factors through repeated PDCA cycles for the hypotheses.

ACKNOWLEDGEMENT

This research was supported by a Grants-in-Aid for Scientific Research(C) numbered 21K11856.

REFERENCES

[1] Cabinet Office, "FY2022 Status of Traffic Accident and Current State of Traffic Safety Measures," (in Japanese) WHITE PAPER ON TRAFFIC SAFETY IN JAPAN 2023, pp.50-70 (2023).

[2] National Police Agency, "Traffic Accidents in 2022 and Other Matters," (in Japanese) https://www.npa.go.jp/news/release/2023/20230302jiko.html, (accessed July 19, 2023)

[3] G. Millington, "Jaywalking," In Shades of Deviance, Routledge, pp.17-20 (2014).

[4] H. Shibasaki, "Accidents Involving Elderly People Crossing the Street," (in Japanese) Institute for Traffic

Accident Research and Data Analysis, https://www.itarda.or.jp/presentation/19, (accessed Jan. 31, 2023).
[5] L Liu, S. Lu, R. Zhong, B. Wu, Y. Yao, Q. Zhang, W. Shi, "Computing Systems for Autonomous Driving: State of the Art and Challenges," IEEE Internet of Things Journal, vol.8, no.8, pp.6469-6486 (2021).
[6] Y. Shi, X. Ying, and H. Zha, "Unsupervised Domain Adaptation for Semantic Segmentation of Urban Street Scenes Reflected by Convex Mirrors," IEEE Transactions on Intelligent Transportation Systems, vol.23, no.12, pp.24276-24289 (2022).
[7] R. Sato, and K. Onoguchi, "Scene Recognition for Blind Spot via Road Safety Mirror and In-Vehicle Camera," 26th International Conference on Pattern Recognition (ICPR), pp.3714-3720 (2022).
[8] A. Kabil, K. Rabieh, F. Kaleem, and M. A. Azer, "Vehicle to Pedestrian Systems: Survey, Challenges and Recent Trends," IEEE Access, vol.10, pp.123981-123994 (2022).
[9] P. Raksincharoensak, T. Hasegawa, and M. Nagai, "Motion Planning and Control of Autonomous Driving Intelligence System Based on Risk Potential Optimization Framework," International Journal of Automotive Engineering, vol.7, no.1, pp.53-60 (2016).
[10] Y. Nagawaki, and K. Suzuki, "Empirical Analysis of Factors for Jaywalking and Its Preventive Measures on Urban Road," JSTE Journal of Traffic Engineering, vol.8, no.2, pp.A_149-A_158 (2022).
[11] K. Adachi, K. Suzuki, Y. Okada, and H. Ito, "Empirical Analysis of Jaywalking Behavior on Two-stage Crosswalk," JSTE Journal of Traffic Engineering, vol.7, no.2, pp.A_102-A_109 (2021).
[12] K. Hamamoto, and T. Murayama, "The Safety Evaluation Analysis of the Walker in the Crosswalk and Station Junction," JSCE, vol.39, no.402, pp.1-4 (2009).
[13] S. Takehira, "Study on Design and Operation of Urban Arterials Based on Nature of the Pedestrian Behavior," University of Tokyo (2014).
[14] D. Wang, W. Fu, Q. Song, and J. Zhou, "Potential Risk Assessment for Safe Driving of Autonomous Vehicles under Occluded Vision," Scientific reports, vol.12, no.1, pp.4981 (2022).
[15] Y. Ishino, and H. Saji, "Extraction of Road Markings from Aerial Images," 2008 SICE Annual Conference, pp.2180-2183 (2008).
[16] A. Maeda, "Automatic Generation of a Pedestrian Road Network Using Aerial Photographs and Maps," Journal of Information Processing, vol.61, no.2, pp.262-273 (2020).
[17] D. Heckerman, "A Tutorial on Learning with Bayesian Networks, Innovations in Bayesian Networks: Theory and applications," pp.33-82 (2008).
[18] NTT DATA Mathematical Systems Inc., "BayoLinkS," https://www.msi.co.jp/english/index.html, (accessed June 11, 2023).

(Received: November 15, 2023)
(Accepted: May 12, 2024)

Yuichi Tokunaga received a ph.D. degree in science and engineering from Ritsumeikan University in 2009. He joined Mitsubishi Electric Corporation in 1990 and engaged in R&D of the high-reliability computer, wireless sensor network, the algorithms of time-synchronization and positioning, the network protocol for industrial applications, and data analytics for CBM (Condition Based Maintenance). He has been a professor at Kanazawa Institute of Technology since 2019. He is a member of the ITS steering committee of IPSJ and a member of IEEE and ISCIE.

Atsushi Yamamoto graduated in March 2023 from the Department of Management Systems, College of Informatics and Human Communication, Kanazawa Institute of Technology.

Masashi Saito received B.E degree from Tokyo Institute of Technology in 1983, M. E. degree from Cornell University in 1992 and Ph.D. degree from Osaka University in 2006. In 1983, he joined Mitsubishi Electric Corporation and has developed engineering workstations, Internet TVs, cellular phones, car navigation systems especially for operating system extension, Internet services and distributed processing. In 2006, he moved to Mitsubishi Electric Research Laboratories as a senior principal technical staff. Since 2015, he has been a professor at College of Informatics and Human Communication, Kanazawa Institute of Technology. His current research interests include ITS, distributed systems, smart communities and smart social systems. He is a member of IEEE and IPSJ.

Regular Paper

Development of a Safety Confirmation Collection System in a Time of Disasters Using Q-ANPI Service Linked with Web Interface

Mitsuki Sano*, Keiichi Abe*

* Kanagawa Institute of Technology, Atsugi City, Kanagawa, Japan
abelab.sano.0914@gmail.com, abe@he.kanagawa-it.ac.jp

Abstract -In the future, when a large-scale wide-area disaster as Nankai Trough Earthquake occurs in Japan, there is a possibility of the loss of electric power and communication infrastructure. In such a case, a Q-ANPI service for collecting information on the safety of disaster victims using the Quasi-Zenith Satellite MICHIBIKI, which can cover the entire country, would be very anticipated to be useful. The Q-ANPI system allows the use of smartphones for registering safety information on the refuge management PC. However, this requires a special application for Q-ANPI installed on the smartphones in advance. Therefore, s smartphone without this application cannot be used if the communication infrastructure for smartphones is cut off due to a large-scale disaster, etc. For this reason, the Q-ANPI service cannot fully exhibit its advantages despite the use of the information and communication technology.
In order to solve this problem, this research describes a technology that allows users to input safety information from a general web browser without using a dedicated app. We also propose a method to operate the system with lower power consumption than the conventional Q-ANPI service system.

Keywords: QZSS-MICHIBIKI, Victims information, Q-ANPI service, Refuge Management System.

1 INTRODUCTION

In the future, when a wide-area disaster such as the Great East Japan Earthquake occurs, there is a possibility of the loss of electric power and communication infrastructure. In fact, when the Great East Japan Earthquake occurred, ICT (Information and Communication Technology) could not be fully utilized, making it impossible to promptly disseminate information on disaster victims, etc. [1]-[6]. To solve these problems, we have developed a Refuge Management System (hereinafter referred to as RMS) [7]. This RMS [7] utilizes ICT to collect information on disaster victims in each shelter and to manage the long-term health of each victim, and can transmit the collected victim information via simple radio communication, such as amateur radio, as well as other means. The information collected by this RMS can also be stored in a USB flash memory and be taken out manually. RMS can manage on a shelter-by-shelter basis.

However, assuming a scenario in which a wide-area disaster such as a major NANKAI Trough earthquake occurs in Japan in the future, it is necessary to manage disaster victims in a wide-area rather than on a shelter-by-shelter basis. As a solution to this problem, we have previously proposed a wide-area disaster victim management system [8] in which RMSs are installed in each evacuation center, and all evacuation centers and disaster victims can be managed centrally by linking these RMSs in cloud.

However, there are two issues with this approach. The first point is that the RMS requires an environment that can be connected to the Internet even in the event of a communications infrastructure cutoff. Second, a wide-area disaster victim management system developed independently will require considerable sales efforts to be accepted by local governments, making it difficult to spread the system. As a technology to solve these problems, there is a safety information service, Q-ANPI, using the Quasi-Zenith Satellite MICHIBIKI [9] promoted by the Secretariat of the Strategic Secretariat for Space Policy and Management, Cabinet Office, Government of Japan. MICHIBIKI covers the entire communication area of Japan, and the Q-ANPI service is promoted by the Cabinet Office, so it is easy to be accepted by local governments, etc.

Therefore, we considered this technology to be a solution to the problem. Therefore, this study focused on the Q-ANPI service. Currently, Q-ANPI provides two input interfaces for registering the safety information of disaster victims on a refuge management PC installed in the refuge. One is to download a smartphone application to a smartphone used by a disaster victim, and register the victim's safety information on the refuge management PC. The second method is to register the information directly on the PC using a keyboard. In the case of registration using a dedicated smartphone application, when the communication infrastructure is cut off due to a large-scale disaster, those who have installed the dedicated application on their own smartphones in advance can register their safety information from their smartphones. However, those who had not installed the application beforehand would not be able to download the dedicated application during the communication infrastructure breakdown, and would not be able to register their safety information from their smartphones. For this reason, safety information may also be registered directly from the refuge management PC (Personal Computer) using a keyboard. In the midst of a large-scale disaster, if a large number of people register directly on the refuge management PC, a great deal of congestion will probably occur. Therefore, we believe that the current Q-ANPI system will not be able to

take full advantage of the convenience of inputting safety information from smartphones in the event of a large-scale disaster when the communication infrastructure is cut off. In addition, when electricity and communication infrastructure are cut off due to a large-scale disaster, it is also necessary to operate the refuge management PC with low power consumption when collecting safety information on the refuge management PC.

This paper describes a technology proposed to solve this problem. It also describes the development of the prototype system proposed in this paper and the evaluation of the effectiveness of the system from various perspectives, including an actual simulated evacuation drill using the prototype system and the results of a demonstration questionnaire evaluation.

In Chapter 2, we provide an overview of the Q-ANPI service. Chapter 3 describes the position of this study with respect to related studies, and Chapter 4 gives an overview of the proposed system and the system flow. Chapter 5 evaluates the proposed system, and Chapter 6 presents the conclusion.

2 Q-ANPI SERVICE

Session 2 will introduce existing Q-ANPI services.

2.1 Overview of Q-ANPI Service

Q-ANPI [9] is a safety confirmation service for disaster victims using the Quasi-Zenith Satellite "MICHIBIKI", which is promoted by the National Space Policy Secretariat, Cabinet Office, Government of Japan. Figure 1 shows an overview of Q-ANPI. The Q-ANPI service consists of a Q-ANPI refuge management PC for collecting information on disaster victims and a Q-ANPI terminal for communicating with the Quasi-Zenith Satellite "MICHIBIKI", as shown in Figure 1. The communication between the Q-ANPI terminal and the Quasi-Zenith Satellite "MICHIBIKI" is in the 2 GHz band, the communication between the Q-ANPI terminal and the refuge management PC is in the 920 MHz band, and the communication between the refuge management PC and smartphones is via Wi-Fi. The victims in the refuge can register their safety confirmation information in the refuge management PC from their smartphones or other devices using a dedicated application.

There are two ways to register, one is to send it via wireless communication, and the other is to display a QR code and read it into the refuge management PC using a web camera. After that, the safety information collected by the refuge management PC is sent to the QZSS server via the Q-ANPI communication terminal and the Quasi-Zenith Satellite "MICHIBIKI". The QZSS server is the control station server where various items of information such as the refuge information, victims information, det k ailed refuge information, and rescue support information are stored. In addition to smartphones, disaster victims can also register their safety information directly on the refuge management PC. In Q-ANPI, telephone number is used as the ID number for confirming the safety of a disaster victim, and relatives of the victim can confirm the safety of the victim by entering the victim's telephone number on the QZSS server's safety confirmation Web site. In addition, the system allows related ministries, agencies, and local governments in charge of disaster prevention to access the QZSS server via the Internet to monitor the evacuation status of each evacuation site. There is no charge for using the Q-ANPI system and no communication fees.

2.2 Registration of Safety Information of Disaster Victim

Figure 2 shows the procedure for registering the safety information of individual disaster victims. There are two methods for registering the safety information of individual disaster victims.

The first method is to enter the information directly into the refuge management PC. Fig. 3(a) shows the registration screen of the refuge management PC. The registered information includes telephone number, name, date of birth, gender, admission/exit status, whether or not to disclose, address, presence or absence of injuries, need for nursing care, presence or absence of disabilities, and presence or absence of pregnant women.

The second method of registration is through a smartphone application. Fig. 3(b) shows the smartphone application registration screen. This registration application is available on iOS and Android smartphones. The personal safety information entered in the smartphone application can be transferred to the Q-ANPI information collection management PC by two methods. One is to display a QR code on the smartphone application and hold it over the camera of the Q-ANPI information collection management PC to transfer the safety information to the Q-ANPI information collection management PC. The other method is to transmit the safety information via Wi-Fi. The contents registered on the dedicated application are the same as those registered on the Q-ANPI information management PC.

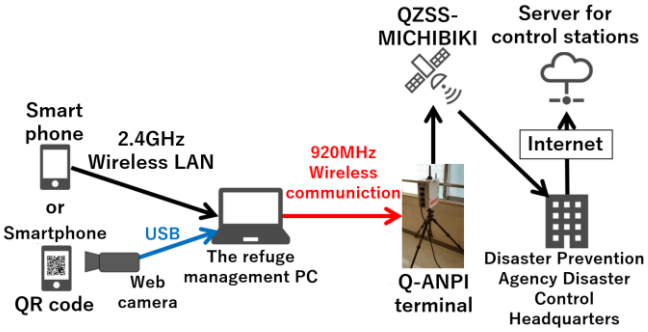


Figure 1: Overview of Q-ANPI service.

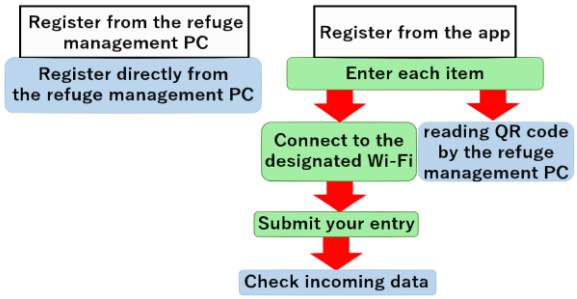


Figure 2: Victim safety information registration method.

(a)Safety confirmation screen of Q-ANPI information collection management PC (b) Smartphone application safety registration screen

Figure 3: Victim safety information registration screen.

Figure 4: Safety information confirmation screen on the disaster prevention organization server.

2.3 Safety Confirmation at Quasi-Zenith Satellite Control Station Server

Figure 4 shows a page of the safety confirmation site of the QZSS server on the Internet. If the safety confirmation of the victim has been registered in advance, the safety status of the victim can be confirmed by entering the victim's phone number and pressing the search button on the QZSS server's safety confirmation site shown in Fig. 4.

2.4 Types of Information Handled by Q-ANPI Service

Table 1 shows the types of information handled by the Q-ANPI service. The types of data handled by Q-ANPI can be broadly classified into (1) through (4) as shown in Table 1. (1) Refuge information is information on the location and closure/opening of the refuge. (2) Personal safety information is information on the safety of individual disaster victims, such as personal numbers (telephone numbers) and whether safety information should be disclosed or not. (3) Detailed refuge information includes the number of injured people, shortage of supplies, etc. (4) Rescue support information is information addressed to

Table 1: Types of information handled by the Q-ANPI service.

Types of Information	Overview	Transmitted Data
①Refuge Information (Collected Information) Refuge ⇒Satellite	Provide refuge conditions and number of victims collected at refuge. Number of content : 84bit	Location of refuge, Refuge closed/opened, Number of victims(up to 131,071)
②Victims Information (Collected Information) Refuge ⇒Satellite	Provide victims information collected at refuge. Number of content : 84bit	Personal ID(phone number), Victims information disclosure, availability of victim's information, supplemental information,
③Detailed Refuge Information (Collected Information) Refuge ⇒Satellite	Provide detailed refuge Information collected at refuge. Number of content : 22characters/348bit	It can handle textual information and free-format binary information.
④Rescue Support Information (Simultaneous transmission /Individual transmission) Satellite ⇒Refuge	Collect Rescue support information addressed to refuge from user agencies and provide the information to refuge. Number of content : 59characters/944bit	Rescue support information (Simultaneous transmission, Individual transmission)

refuges regarding rescue support. (1), (2), and (3) are information transmitted from refuges, while (4) is information addressed to refuges. Of these four types of information handled by the Q-ANPI service, only the information (2) is used in this study, i.e., the information on the safety of individual disaster victims. The information volume of the personal safety information for disaster victims is 84 bits.

3 RELATED WORK

A previous study of disaster victim information dissemination in response to a wide-area disaster is the RMS development [7]. This RMS utilizes ICT to collect information on disaster victims in each refuge and to manage the long-term health of each victim and can transmit the collected disaster victim information. The RMS can transmit information by simple radio communication and amateur radio, in addition to storing disaster victim information on USB flash memory devices and allowing users to manually take it outside. This RMS can be managed on a shelter-by-shelter basis.

However, assuming a wide-area disaster such as the Nankai Trough earthquake that is expected in the future, it is necessary to manage victims in a wide-area rather than in refuge. As a solution to this problem, we have previously proposed a wide-area disaster victim management system[8] in which RMSs are installed in each refuge, and all refuges and disaster victims can be managed centrally by linking these RMSs with a cloud server.

However, there are two problems with this approach. First, it requires an environment that can be connected to the Internet even when the communication infrastructure is cut off. Second, a wide-area disaster victim management system developed independently by a private sector will require

considerable sales efforts to be accepted by local governments, making it difficult to spread the system.

Q-ANPI, a safety information service using the Quasi-Zenith Satellite MICHIBIKI [9] promoted by the Secretariat of the Strategic Secretariat for Space Policy, Cabinet Office, Government of Japan, exists as a technology to solve these problems. The Quasi-Zenith Satellite MICHIBIKI can communicate throughout Japan even in the event of a large-scale disaster, and the Q-ANPI service is promoted by the Cabinet Office, making it easy to be accepted by local governments. In particular, since the control station servers used in Q-ANPI are owned by the national government, they are more easily accepted by local governments than those developed by a private sector. This technology can therefore be used to solve the two issues mentioned above. Therefore, this study focuses on the Q-ANPI service.

As described in section 2.2, the current Q-ANPI provides two input interfaces for collecting information on disaster victims. One is to download a dedicated smartphone application and register the safety information of disaster victims through the application, and the other is to input and register the safety information directly from the Q-ANPI refuge management PC using a keyboard. In the case of the registration method using a dedicated smartphone application, assuming a scenario in which the communication infrastructure is cut off for a long period of time due to a large-scale disaster, those who have installed the dedicated application in advance can register their safety information, but those who have not installed it will not be able to download the dedicated application during the communication infrastructure cut-off period. For this reason, the current system additionally allows people to register their safety information directly from the refuge management PC using a keyboard. Since it is assumed that one such PC will be placed in each refuge, a great deal of congestion will probably occur if there are a large number of people registering directly on the refuge management PC. Therefore, the authors believe that the current system (Q-ANPI) cannot fully utilize the advantages of using a smartphone.

On the other hand, the previous study on the Q-ANPI service proposed a system [10] that can collect safety information of disaster victims more widely than the conventional Q-ANPI service by combining Q-ANPI and ad hoc communication of smartphones. The proposed system is very useful when disaster victims do not know how to access Wi-Fi access points in the refuge, when they cannot connect to the refuge management PC, or when collecting safety information on victims outside of refuges. However, this related technology [10] also has a problem in that it cannot be used unless a dedicated application is downloaded in advance. In addition, the refuge management PC used in Q-ANPI uses an ordinary notebook PC, which consumes more than 12W of electricity. Therefore, the authors believe that in the event of a large-scale disaster, a system that can operate the refuge management PC with low power consumption is needed when not only the communication infrastructure but also the electric power infrastructure is cut off.

To solve these problems, the authors propose a method for inputting the safety information of disaster victims using a Web browser already installed on smartphones. In addition, we propose a method that can be operated for a longer period of time than a laptop PC by using a power-saving embedded system instead of a laptop PC to collect information on the safety of disaster victims.

4 PROPOSED TECHNOLOGY

In this study, we considered that Q-ANPI using the Quasi-Zenith Satellite MICHIBIKI is suitable as a solution to the issues of loss of communication infrastructure in the event of a large-scale disaster and ease of acceptance by local governments and other entities. However, as described in Chapter 3, when the communication infrastructure is cut off due to a large-scale disaster, it is expected that Q-ANPI cannot fully exhibit its merits because the dedicated application for registering safety information from smartphones cannot be downloaded. In addition, since the refuge management PC installed in the refuge use an ordinary laptop PC, they consume a lot of power, and considering the case where the power infrastructure is cut off due to a large-scale disaster, it is essential to operate the refuge management PC with even lower power consumption.

To solve this problem, this study proposes a method to input the safety information of disaster victims using a Web browser already installed on smartphones. In addition, this study considers that the power infrastructure will be cut off at the same time during a large-scale disaster. Therefore, we propose a highly practical safety information collection system that is robust even in the event of power infrastructure failure by linking the Q-ANPI service with a web server system built on an embedded microcontroller board that can operate with low power consumption, without building a web server and DB (Data Base) on the refuge management PC.

4.1 Proposed System

Figure 5 shows the system proposed in this study, which adds a web interface to the conventional Q-ANPI system so that disaster victims can input safety information from a web browser of their smartphones. This allows users to enter safety information without using a dedicated app. In addition, in order to achieve long-term operation even when the power infrastructure is cut off in the event of a large-scale disaster, the Web interface developed by this research uses a Raspberry PI 3B+ (hereinafter referred to as RPI), which is a small and low-power-consumption microcontroller. We created a system in which a web server and DB (Database) were built on this RPI. Fig. 6 shows the appearance of the prototype developed in this study. The RPI is used as an intermediary between smartphones owned by disaster victims and the refuge management PC. When a smartphone owned by a disaster victim connects to the web interface from its web browser, a page for inputting the victim's safety information is displayed, as shown in Fig. 7. When the safety information is entered on this page, it is automatically registered in the DB in the web interface. This collects the safety information of multiple victims in a

refuge. The system then sends the information to the refuge management PC in a batch. When the person in charge of the refuge management presses the data transmission button on the application for the refuge management PC, the data are transmitted to the server of the control station via the Quasi-Zenith Satellite “MICHIBIKI”. The system enables relatives of disaster victims to confirm the safety of the victims by accessing the server of the control station via the Internet.

4.2 How to Display the Safety Information Input Page from a Smartphone

As with Q-ANPI, the system in this study is built on a LAN (Local Area Network) with commercially available Wi-Fi wireless routers. In addition to receiving safety information from smartphones and the Web interface, the refuge management PC serves as a gateway to the Q-ANPI terminal (920 MHz). For the connection to the web interface from a smartphone, the SSID of the Wi-Fi wireless router (in this case, "QANPI_179") is made open to allow connection to the network without a password. A printout of the URL of the safety information input page converted into a QR code is prepared in advance. By holding the QR code over the camera of a smartphone, anyone can easily display the safety information input page on the Web interface. This enables smooth registration of the safety information of disaster victims.

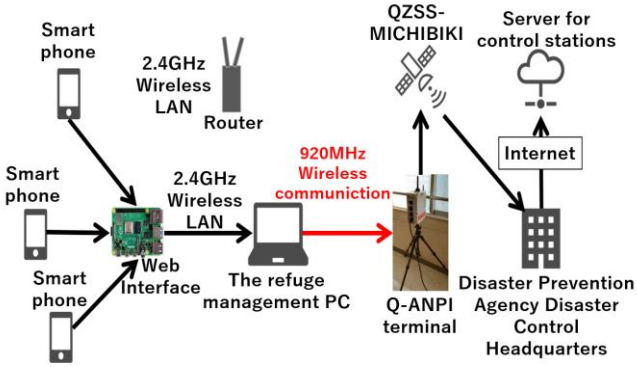


Figure 5: Overview of our proposed system.

Figure 6: Prototype of Web Interface.

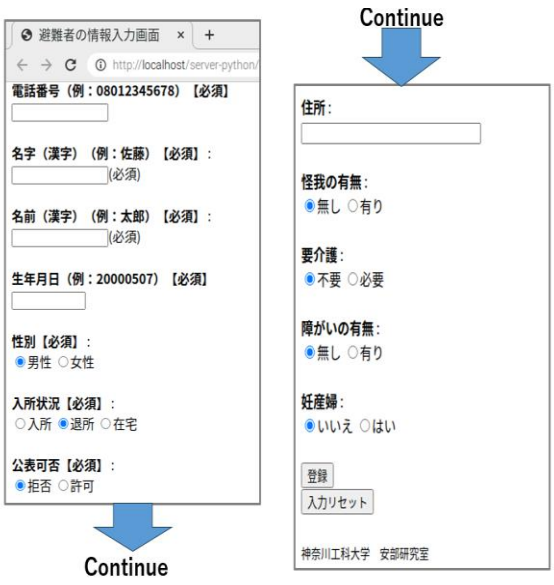


Figure 7: Safety information input page screen for disaster victims.

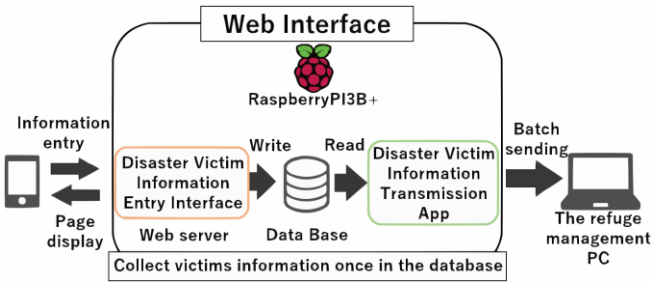


Figure 8: Overview of Web Interface.

4.3 Prototype System of Web Interface

Figure 8 shows an overview of the prototype of the web interface proposed in this paper. The web server was built on RPI using nginx, and MariaDB was used for the DB. The reason for using nginx for the web server is to achieve high processing speeds that allow multiple inputs of victim information at the same time. The application for the victim safety information input interface shown in Fig. 7 was developed in PHP and HTML languages. The "disaster victim information transmission application," which transmits the safety information of disaster victims registered in the DB server to the refuge management PC, was developed in python 3.7, and is an application that transmits all safety information registered in the DB. Since the application for the refuge management PC automatically discards duplicate information, we decided to use a simple application that sends all DB data at once. Fig. 6 shows the appearance of the prototype of the Web interface developed this time. This is a prototype with a web server and DB (Data Base) built on RPI. The Web interface is operated remotely from a terminal to eliminate the need for a display.

4.4 Power Operation Method of Our Proposed System

Assuming that power and communication infrastructures are cut off in a large-scale disaster, power supply is important in a real refuge. To operate the proposed system, it is considered necessary to use one solar panel and two or more battery units. One battery unit is used for the refuge management PC, web interface, wireless router, Q-ANPI terminal, etc. The other battery unit is charged with electricity generated by the solar panel, and the batteries should be exchanged alternately.

Table 2 shows the results of power consumption measurements for each device used in the proposed prototype system. Power consumption was measured using a digital power meter. Among the devices used in this prototype, the Q-ANPI terminal consumes the largest amount of power, 15W. The second highest is the refuge management PC with a maximum of 12W, the third is the RPI, and the fourth is the wireless router. The RPI alone used with a Web interface consumes approximately 3W, a dramatically low level of power consumption. The Q-ANPI terminal and the refuge management PC consume a lot of power. Therefore, when not transmitting data using the QZSS MICHIBIKI, the Q-ANPI terminal can be turned off, the refuge management PC can be put to sleep, and long-term. Operation can be achieved with a single battery charge by having only the proposed Web interface powered on.

Figure 9 shows the operation of the Q-ANPI service using the proposed system. When collecting safety information, only the Web interface is put into operation, the Q-ANPI terminal is turned off, and the refuge management PC is put into sleep mode. In this way, the overall power consumption can be reduced to a total of approximately 3.5W during the collection of safety information. For example, using a commercially available 240Wh portable battery power supply, the system can be operated for approximately 60 hours on a single charge.

5 EVALUATION

We evaluated the effectiveness of the prototype of the system proposed in this study from three perspectives: a demonstration evaluation based on a simulated evacuation drill, a questionnaire evaluation based on a demonstration, and a measurement evaluation of the prototype's power consumption. In the evaluation through a simulated evacuation drill, we conducted a questionnaire to determine whether the proposed system is practically usable, and compared it with the conventional Q-ANPI system to determine how easy it is to input safety information. In addition, a questionnaire survey was conducted by exhibiting a prototype demo to investigate the ease of use of the proposed system and find points to be improved. We also evaluated the power consumption of the prototype system using the Web interface proposed in this study.

Table 2: Power consumption of each device.

Device	Model	Power consumption
Q-ANPI terminal		15W at active
The refuge management PC (Windows10)	Versa Pro VX-5(NEC)	Max 12 W at active
		Max 0.42 at sleep
Wireless router	WMR-433W2-BK (BUFFALO)	2.5 W at active
Web interface(RPI)	Raspberry PI3B+	3W at active

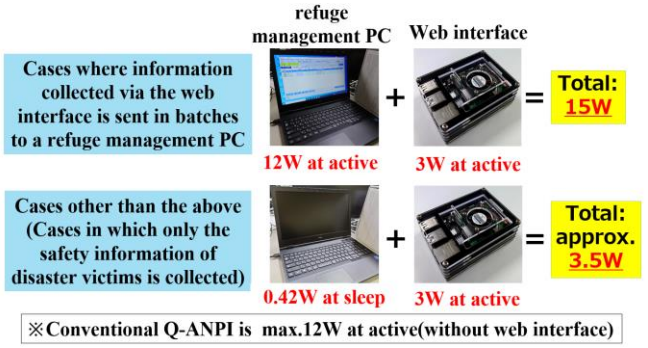


Figure 9: Power consumption in Q-ANPI operation method using proposal method.

5.1 Evaluation through Simulated Evacuation Drill

A prototype of the proposed system was utilized and evaluated through a simulated evacuation drill. The simulated evacuation drill was conducted on August 5, 2022, on the premises of the Kanagawa Institute of Technology. Figure 10 shows the evacuation route of the simulated drill. Table 3 shows the schedule of the actual evacuation drill and the actual action time required. Before the simulated evacuation drill, one Q-ANPI terminal, one refuge management PC, one web interface system, one solar panel, and one battery with 240Wh capacity were placed in the former gymnasium of the university, which served as the evacuation site. It was assumed that a major earthquake occurred at 14:00. All participants in the evacuation drill initially hid under desks, and after the earthquake subsided, they descended the stairs from the classroom (E602) on the 6th floor of Building C2 to the first floor, and then followed the evacuation route shown in Fig. 10 to the entrance of the university's soccer field. After taking roll call at the evacuation assembly, all the evacuees were moved to the old gymnasium, which was the evacuation site, where three evacuation supervisors set up a proposed system and opened the evacuation desk (Fig. 11). Table 3 records the time from immediately after the earthquake until the equipment was removed. The time required from the occurrence of a major earthquake to the gathering at the evacuation site was 9 minutes, which indicates that the evacuation was completed fairly quickly. The time required to open the evacuation site was about 8 minutes. Table 4 shows a comparison of the input time of safety information of disaster victims between the conventional Q-ANPI and the proposed system. The proposed system required only about 3 minutes, while the conventional Q-ANPI system required about 11 minutes.

The proposed system took only about a quarter of the time of the conventional system. Factors that caused the conventional Q-ANPI system to take such a long time include the time required to download the dedicated application and the time required to establish a Wi-Fi connection. The proposed system uses a QR code for easy access to the safety information input page, which is a major factor. But, even with the conventional Q-ANPI system, if the dedicated application is already installed on the smartphone, the input time taken would be similar to that of the proposed system.

Therefore, in the case that the cell phone communication infrastructure is cut off and the dedicated application is not installed on the smartphone in advance, the proposed system is superior in terms for on-site use in actual disasters.

5.2 Questionnaire Evaluation after Simulated Evacuation

In the demonstration evaluation through the simulated evacuation drill, a questionnaire survey was conducted after the simulated evacuation drill, since personal safety information was collected by two systems, i.e., the conventional Q-ANPI and the proposed system. The actual number of participants in the simulated evacuation drill was 9, and 6 males (67% response rate) responded to the questionnaire. Fig. 12 shows the results of the questionnaire. To Question 1, "In the event of a large-scale disaster, would you like to actually use the proposed system used this time, which uses a browser for data input, in an evacuation center?", three respondents answered "Yes, I would like to use it," and three respondents answered "Yes, I would like to use it somewhat." This resulted in a high average score of 3.5 on the 4-point scale in which the highest score of 4 was given to "would like to use" and the lowest score of 1 was given to "would not like to use". Next, to Question 2, "Which system was easier to use, the conventional system or the proposed system?", 83% of the respondents answered that the proposed system was easier to use. Next, in response to Question 3, "In Question 2, what was easy to use about the conventional system?", the answer was "Because there is no need to search in a browser." In Question 4, "In Question 2, what was easy to use in the proposed system?", five respondents answered "The fact that it can be used without downloading a dedicated application." To Question 5, "Is there any other information on disaster victims that you think is necessary besides the existing information (phone number, name, date of birth, gender, admission status, availability, address, injuries, assistance needed, disabilities, pregnant women)?", one respondent answered, "Information such as pre-existing medical conditions, family doctor, and medications used on a regular basis would be helpful." Next, to Question 6, "What do you think needs to be improved in the future regarding the proposed system?", as shown in Fig. 12, the most common response to this question was "Type of evacuee information" (five responses), followed by "Ease of text input" (two responses).

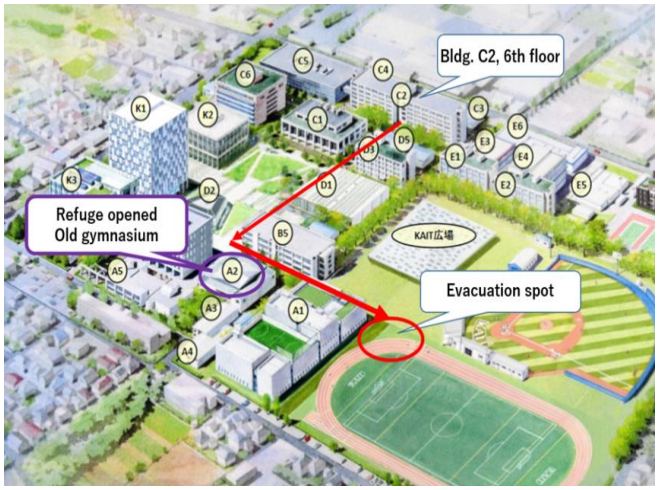


Figure 10: Evacuation route for simulated evacuation drill.

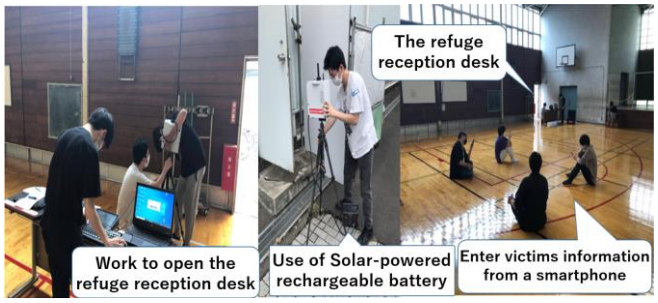


Figure 11: Overview simulated evacuation drill.

Table 3: Schedule for the Demonstration Experiment.

Scheduled time	Action	Actual times required	Remarks
14:00	Earthquake-Hide under the desk	14:00	
14:15	Evacuation to a refuge	14:06	Evacuation by stairs from the 6th floor of Building C2
14:20	Gather at evacuation site	14:09	
14:25	Move to the refuge (old gymnasium)	14:09	
14:40	Establishment of the refuge		
	Completion of evacuation center opening	14:17	
14:55	Enter victim information (Install dedicated app)	Start 14:17 Finish 14:28 (11minutes)	Mistakes in downloading the dedicated app and connecting to Wi-Fi required time to complete the Input.
	Enter victim information by the proposed system	Start 14:34 Finish 14:37 (3minutes)	Fastest input time
15:00	Fill out the questionnaire	Start 14:35 Finish 14:37	
15:15	The withdrawal	14:41	

Table 4:Comparison of proposed and existing systems in terms of victim information input time

	time of inputting disaster victim information	Cause
Proposed system	About 3 minutes	URL of the disaster victim information input page is converted to a QR code
Q-ANPI system	About 11 minutes	Mistakes in downloading the dedicated app and connecting to Wi-Fi

5.3 Questionnaire Evaluation through Demonstration Exhibition of Prototypes

Next, we explained the proposed system to visitors to our open campus and asked them to directly see the actual prototype system and answer a questionnaire. A total of 12 people responded to the questionnaire: seven males (six in their teens and one in his 40s) and five females (two in their teens, one in her 20s, one in her 30s, and one in her 50s). The questionnaire results from this demonstration are shown in Fig. 13. The "score" shown in Fig. 13 is a numerical value that represents the average of the responses of the survey participants. First, as shown in Fig. 13, to Question 1, "Did you understand the overview of the proposed system that you used this time?", the score was 3.33, which is high compared to the highest score of 4: 4 respondents answered "well (4)" and 8 respondents answered "somewhat well (3)".

Next, to Question 2, "In the event of a large-scale disaster, would you like to use the proposed system used this time in an actual evacuation center?", the score was 3.75, which is quite high, with 9 respondents selecting "Yes (4)" and 3 selecting "Yes, somewhat (3)". To Question 3, "How was the operability of the Q-ANPI system and the proposed system?", Fig. 13 shows that the score of the proposed system was 4.25 on the 5-point scale, about 0.50 higher than the conventional system whose score was 3.75.

Next, to Question 4, "What are the points to be improved in the proposed system? (Multiple answers allowed)", many respondents answered "Ease of input" (6), indicating that the proposed system needs to be improved in this respect. Other personal safety information and points to be improved that the respondents thought were necessary for the proposed system were "family information" and "heavy and congested Internet connection".

5.4 Measurement and Evaluation of Prototype Power Consumption

Considering the use of a commercially available 240Wh battery, it is estimated that the RPI alone can operate for about 3.36 days. A more power-saving operation could be achieved by using the RPI as a stand-alone device without the LCD display screen, with the terminal software for the RPI installed on the refuge management PC so that the RPI's "disaster victim information transmission application" could be run remotely.

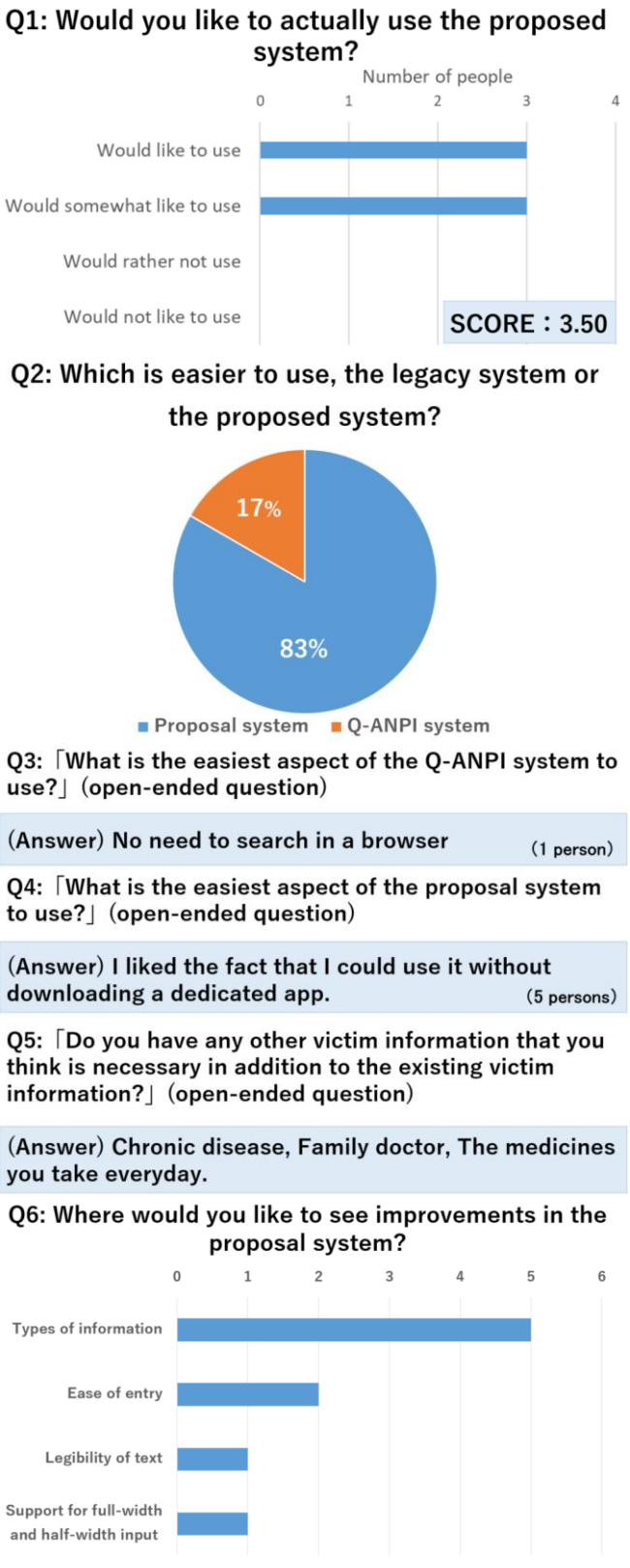


Figure 12: Questionnaire results for simulated evacuation drill.

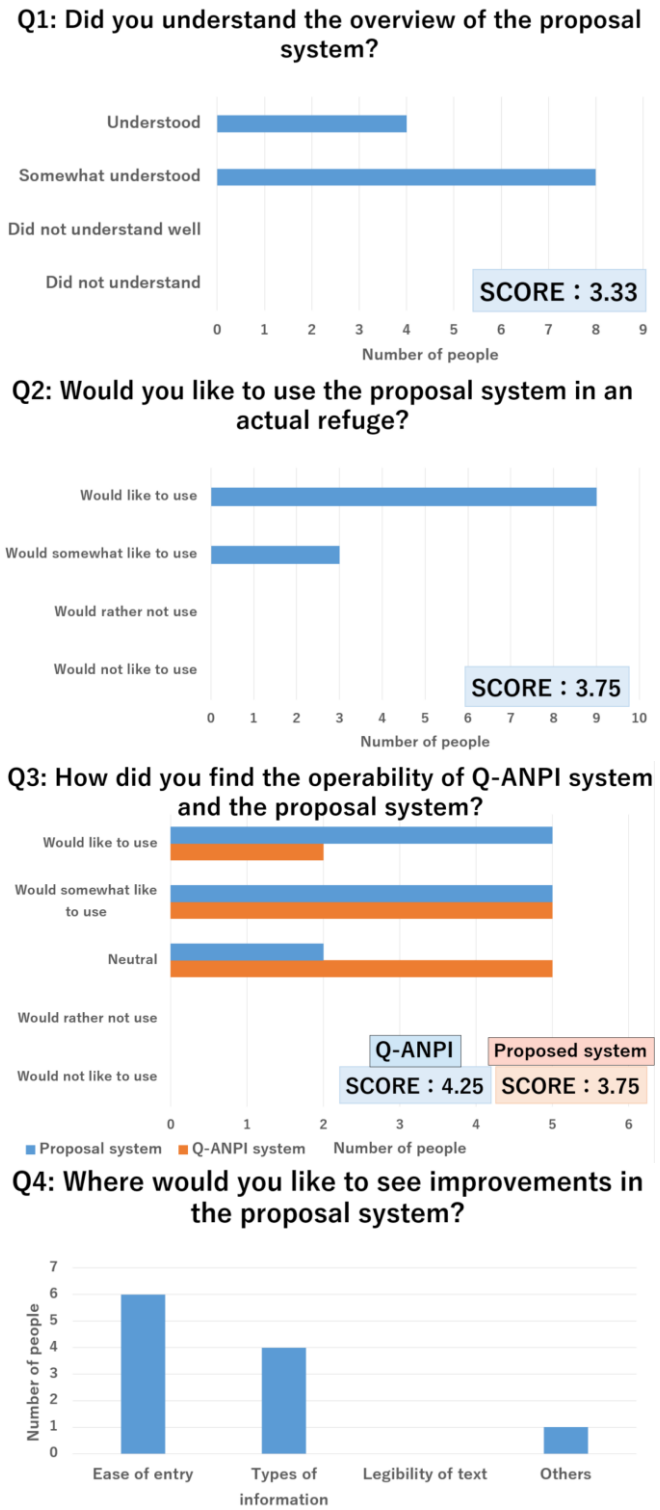


Figure 13: Questionnaire results for the prototype.

It is also considered a good idea to turn on the power of the Q-ANPI terminal and the refuge center management PC only when communicating via the Quasi-Zenith Satellite “MICHIBIKI”, and turn them off and in sleep mode for the rest of the time.

Assuming such an operation scenario, only the RPI unit (3W) and the wireless router (2.5W) would be used in constant operation, so the 240Wh battery would last approximately 1.8 days. If the two battery units are operated alternately, long-term operation is possible even when the power infrastructure is cut off.

6 CONCLUSION

This paper addressed the issue of Q-ANPI, a service for collecting safety confirmation information on disaster victims using the Quasi-Zenith Satellite "MICHIBIKI", in the case that the cell phone communication infrastructure is cut off and the dedicated application is not installed on the smartphone in advance, and the safety information from the smartphone owned by the disaster victim cannot be registered on the refuge management PC. The paper discussed how to solve this problem. The paper also discussed energy-efficient operation of PCs used to manage evacuation centers in the event of a power infrastructure cutoff. To solve this problem, we proposed a web interface that enables the input of safety information of disaster victims by utilizing a web browser already installed on smartphones. This Web interface collects the safety information of multiple disaster victims, sends it to the refuge management PC at any given time, and the PC transmits it to the control station server via the Quasi-Zenith Satellite MICHIBIKI. This allows safety information to be entered using a web browser already installed on smartphones without downloading any dedicated applications, making it a general-purpose interface. To reduce the operational power consumption of Q-ANPI, the Q-ANPI terminal is turned off when collecting safety information, the refuge management PC is put in sleep mode, and only the Web interface is operated to collect safety information on disaster victims. It was found that if the Q-ANPI terminal and the refuge management PC are activated only when transmitting data to the server at the control station, the power consumption of the entire system can be minimized to about 4W.

In a questionnaire survey conducted after the simulation training, many respondents answered that they would prefer to use the proposed system rather than the conventional Q-ANPI system, and many of them answered that the proposed system was “easier to use.” However, some respondents pointed out the following points for improvement of the proposed system: "Ease of text input" and "Types of victim information.” In addition, we would like to increase the number of participants in simulated evacuation drills. The prototype of the proposed web interface was developed using RPI of low power consumption.

In the future, if the proposed system and Q-ANPI's previous research [10] are all successfully linked, it will be possible to develop a highly realistic system that can smoothly collect safety information on disaster victims at actual disaster sites where people are in turmoil.

ACKNOWLEDGEMENTS

We would like to express our sincere appreciation to the National Space Policy Secretariat, Cabinet Office, Government of Japan, and the Space Systems Division of NEC Corporation for their guidance and cooperation in providing the Q-ANPI service equipment. and We would

like to express our sincere gratitude to Mr. Takuma Sato and Prof. Masao Isshiki for their cooperation in this study.

REFERENCES

[1] Reconstruction Agency, Cabinet Office (in charge of disaster prevention), and Fire and Disaster Management Agency, "Number of deaths related to earthquake disaster in the Great East Japan Earthquake (in Japanese)", <http://www.reconstruction.go.jp/-topics/main-cat2/sub-cat26/20171226_kanrenshi.pdf>, (Referred:2023-06).

[2] Cabinet Office Government of Japan, "Survey result report on promotion of comprehensive measures on evacuation (in Japanese)",<http://www.bousai.go.jp-/kaigirep/houkokusho/hinan_taisaku/houkoku.html> (Referred:2023-06).

[3] Cabinet Office (in charge of disaster prevention), "Main issues of disaster emergency measures in the Great East Japan Earthquake (in Japanese)", <https://www.bousai.go.jp/jishin/syuto/taisaku_wg/5/pdf/3.pdf〉(Referred:2023-06).

[4] N. Matsumoto, "A Case Study of the Publication of Information on the Japan Earthquake(in Japanese)", Journal of Information Processing Society of Japan, Vol. 54, No. 3, pp. 1021-1027, (2013).

[5] T. Imai, "Current status of issues and responses from the Great East Japan Earthquake (from the aspect of local government ICT) (in Japanese)", Information Policy Department, General Affairs and Planning Bureau, City of Sendai (2011).

[6] Cabinet Office Disaster Prevention Information Page, "Report on the results of a fact-finding survey regarding the promotion of comprehensive measures regarding condemnation (in Japanese)", <http://www.bousai.go.jp/kaigirep/houkokusho/hinan taisaku/houkoku.html > (Referred:2023-06).

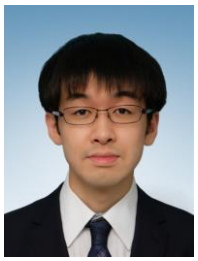
[7] K. Akasaka, T. Namba, M. Isshiki, K. Abe, "Proposal on Victims Information Management System", International Journal of Informatics Society (IJIS), Vol. 12, No. 1, pp. 17-27, (2020).

[8] N. Okamoto, D. Minakawa, M. Anraku, K. Abe, "Proposal of a Wide Area Disaster Victim Management System linked for Large Scale Disaster (in Japanese)", Proceedings of the 83rd National Convention of IPSJ, 2V-06, No. 3, pp. 237-238, (2021).

[9] Cabinet Office, Government of Japan, "Q-ANPI Satellite Safety Confirmation Service (in Japanese)", <https://qzss.go.jp/overview/services/sv09_qanpi.html >, (Referred:2023-06).

[10] N. Ohuchi, M. Nishiura, S. Aikawa, A. Yumura, T. Hamada, "Disaster Safety Confirmation Collection System Using QZSS Michibiki and Smartphones (in Japanese)", IEICE Technical Report, SeMI2019-67, pp. 27-32 (2019).

(Received: November 27, 2023)
(Accepted: May 16, 2024)



Mitsuki Sano received the B. E. degree from Kanagawa Institute of Technology in 2023.
He is currently engaged in re-search on ICT disaster support management systems at Department of Electrical and Electronic Engineering, Graduate School of Engineering, Kanagawa Institute of Technology.

Keiichi Abe received a Ph. D. in informatics from Shizuoka University in 2012. He has been a professor of the Faculty of Creative Engineering, Kanagawa Institute of Technology. His research interests include smart system using IOT technology, Interactive system using Robot and CG, and Victims Information Management System At large scale disaster, HEMS. He is a member of IPSJ, IEICE, ACM and IEEE.

Regular Paper

Fault Localization with Virtual Coverage and Supervised Learning based on Execution Count Information

Takuma Ikeda†, Hitoshi Kiryu†, Satoshi Suda‡, Shinpei Ogata*, and Kozo Okano*

†Graduate School of Engineering, Shinshu University, Japan
‡Mitsubishi Electric Corporation, Japan
Suda.Satoshi@ay.MitsubishiElectric.co.jp
*Faculty of Engineering, Shinshu University, Japan
{ogata, okano}@cs.shinshu-u.ac.jp

Abstract - Automatic fault localization is a technique that helps to reduce the costly task of program debugging. Among the existing approaches, Spectrum-based fault localization shows promising results in terms of scalability. One Deep Neural Network (DNN)-based SFL approach that uses virtual coverage has been proposed. This approach uses a DNN model that classifies whether the test result of an input code coverage is Pass or Fail. Virtual coverage is code coverage that expresses that only certain code blocks are executed. The output value when the virtual coverage is input to a DNN model is treated as the suspiciousness score. We propose a new virtual coverage and a DNN model based on the number of executions. Our idea is that by using execution count-based virtual coverage, higher accuracy can be achieved than existing approaches. We evaluate our proposed approach using six projects available on Defects4j and Software Infrastructure Repository (SIR). As a result of the evaluation, we confirmed that our virtual coverages improve the accuracy by up to 4.2 points compared to the existing approach. We confirmed that our proposed model with our virtual coverages improve the accuracy by up to 5.6 points.

Keywords: Spectrum-based Fault Localization, Virtual Coverage, Deep Neural Network, Supervised Learning

1 INTRODUCTION

In software development, fault localization is a costly task. Testing and debugging are reported to account for up to 75% of the development cost [1]. Automatic fault localization is an effective technique to reduce the cost of program debugging. Among the existing methods, Spectrum-based fault localization (SFL) has shown promising results in terms of scalability, lightweight, and language-agnostic [2–4].

Ochiai [5] and Tarantula [6] are representative SFL techniques. These techniques calculate a failure suspicion score for each statement based on code coverages. Statements with high suspiciousness scores are considered highly suspicious of failure, and developers investigate statements with high suspiciousness scores as priority. The formulas for calculating the suspicion score for each statement in Ochiai and Tarantula are shown below.

$$\frac{e_f}{\sqrt{(e_f + n_f) \cdot (e_f + e_p)}} \quad (Ochiai) \tag{1}$$

$$\frac{\frac{e_f}{e_f+n_f}}{\frac{e_f}{e_f+n_f} + \frac{e_p}{e_p+n_p}} \quad (Tarantula) \tag{2}$$

These formulas are calculated using the following values [7].

- e_f is the number of times the statement is executed in the Fail test case.
- e_p is the number of times the statement is executed in the Pass test case.
- n_f is the number of times the statement is not executed in the Fail test case.
- n_p is the number of times the statement is not executed in the Pass test case.

Ochiai and Tarantula calculate a failure suspiciousness score based on the frequency with which each statement is executed in the Fail and Pass test cases. Thus, the idea in the statistical SFL approach is that statements that are executed frequently in the Fail test case and infrequently in the Pass test case are suspicious.

In recent years, several deep learning-based approaches have been proposed for locating faults, and the learning capability of DNNs is effective in locating faults, showing better identification results than conventional SFL techniques (Ochiai, Tarantula) [8]. As one of the deep learning-based SFL approaches, the approach using virtual coverage has been proposed [8–11]. An overview of this approach [11] is shown in Fig. 1. The virtual coverage used in this approach is shown in Fig. 2. The approach in Fig. 1 takes test coverage as input and learns a Deep Neural Network (DNN) model that classifies whether the input test execution coverage is Pass or Fail. Next, the virtual coverage shown in Fig. 2 is input to a learned DNN model, and a DNN model outputs a score indicating the suspiciousness of failure for each code block.

The virtual coverage in Fig. 2 is generated from the test execution coverage and treats the statements commonly executed in all test cases as code blocks. Therefore, the size of

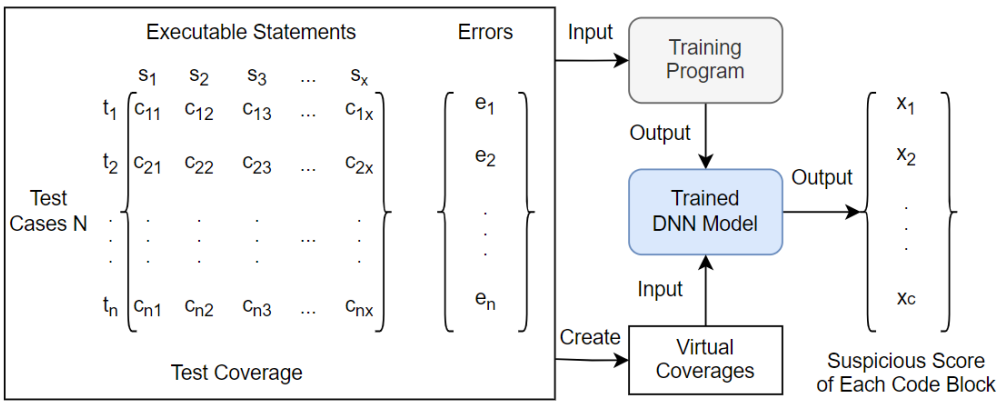


Figure 1: Existing Approach of DNN-based SFL

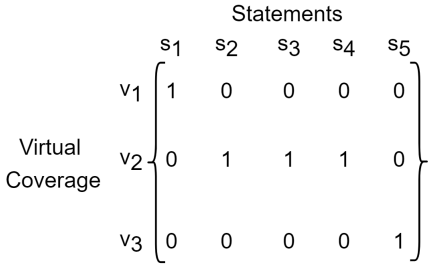


Figure 2: Virtual Coverage at Code-Block Granularity

some blocks can be too large for some programs, which has a negative impact on the accuracy of SFL. We propose a new virtual coverage that can be used with the existing approach shown in Fig. 1. Our proposed virtual coverage is created based on the number of executions (execution count reports) of each statement. Since it is based on the number of executions, it is possible to divide the source code into more code blocks than the existing virtual coverage shown in Fig. 2. Figure 3 shows the creation of existing virtual coverage and the creation of our proposed virtual coverage. In Fig. 3, t_1, t_2, and t_3 represent the three test cases, and $s_1, s_2, ...,$ and s_5 represent statements in the source code. The v_1, v_2, v_3, and v_4 represent the virtual coverages created in each approach. Execution count reports describe execution counts of each statement at test runtime. In Fig. 3, the existing approach divides source code into three code blocks (virtual coverages), and our approach divides source code into four code blocks.

We evaluated our proposed virtual coverage on the six projects (Math, Lang, Chart, Print_tokens, Print_tokens2, and Tot_info) available on Defects4j [12] and Software Infrastructure Repository (SIR) [13]. As a result, we confirmed that the proposed virtual coverage improves the accuracy by up to 4.2 points compared to the existing virtual coverage. Applying the Wilcoxon Signed-Rank Test to the experimental results, we confirmed that the proposed approach is significantly more accurate than the existing approach. In order to further improve the accuracy of our proposed virtual coverage, we also evaluated the proposed virtual coverage on DNN models trained on different training data from the existing approach. As a result, we confirmed that the accuracy is improved by up to 5.6 points compared to the existing approach.

The rest of the paper is organized as follows. Section 2 describes the background of this paper. Section 3 describes the proposed approach. Section 4 describes the experimental setup and Section 5 discusses the experimental results. We conclude in Section 6.

2 BACKGROUND

2.1 Statistical SFL Techniques

Tarantula [6] and Ochiai [5] are representative statistical SFL techniques. These techniques use test coverage collected during test execution to compute suspiciousness scores, which indicate the suspiciousness of failure for each statement. Several approaches [14–16] have been proposed to compute suspiciousness scores similar to Tarantula and Ochiai. Since various metrics have been proposed, the approach [17] to fuse these SFL techniques and calculate suspicion scores from multiple statistical SFL techniques has been proposed.

This paper discusses deep learning-based SFL approaches, which differ from statistical SFL approaches in the idea of fault localization. In this paper, statistical SFL techniques are not further discussed.

2.2 Deep Learning-based SFL Approaches

Existing function-level fault localization techniques [18, 19] use function coverage or statement coverage to compute the suspiciousness values. Murtaza et al. [18]'s approach uses decision trees to identify patterns of function calls related to failures. Sohn et al [19]. proposed the approach to rank faulty methods higher using genetic programming (GP) and linear rank-supported vector machines (SVM). These approaches are function grain SFL approaches, which are different in granularity from the statement granularity SFL approaches discussed in this paper.

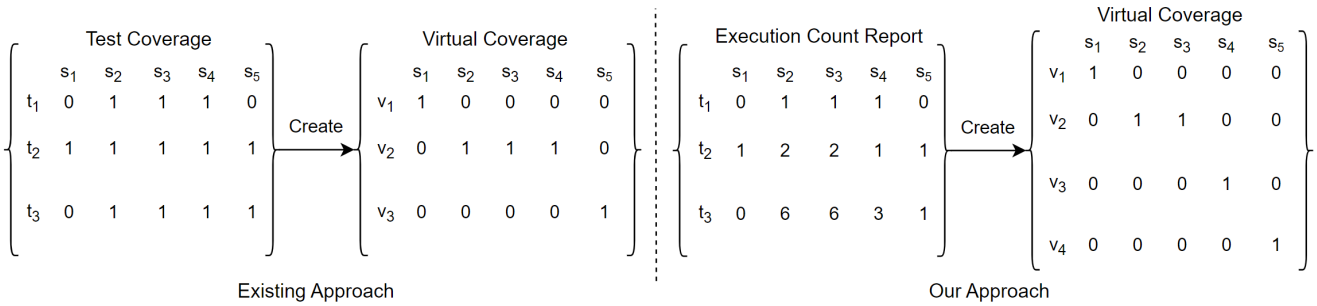


Figure 3: Virtual Coverage Creation in Each Approach

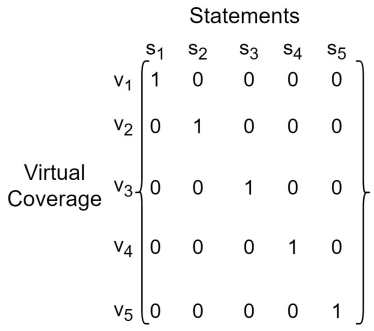


Figure 4: Virtual Coverage at Statement Granularity

As one of the latest techniques for dynamic analysis using machine learning, Li et al. proposed DeepRL4FL, which identifies buggy code by treating fault localization as an image pattern recognition problem [20]. Li et al.'s approach requires marking the statements that are faulty as training data. This is so that a model can distinguish between statements that are faulty and non-faulty at training. In our approach, we treat the label of whether each test case is Pass or Fail as training labels. Therefore, the training data used in Li et al.'s approach is more informative than our approach's training data.

The most relevant researches to this paper are approaches [8, 9, 11] that use virtual coverage to locate faults. In these approaches, a DNN model is first trained that takes test execution coverages as input and classifies whether the test result is Pass or Fail. Next, virtual coverage is constructed that indicates that only certain a statement or a code block is executed. Several approaches have been proposed for constructing virtual coverage, including the approach [8] at the statement granularity and the approach [11] at the code block level. The statement granularity virtual coverage is shown in Fig. 4 and the code block granularity virtual coverage is shown in Fig. 2. By inputting the virtual coverage in Fig. 4 and 2 into a trained DNN model, a failure suspiciousness score is given for each virtual coverage. Since each virtual coverage is a coverage that represents only certain a statement or a code block executed, output values of a DNN model are treated as suspiciousness scores for each statement or code block. It is known experimentally that using virtual coverage at the code block granularity is more accurate for SFL than virtual coverage at the statement granularity [11].

Existing virtual coverage treats statements that are executed at least once in common in all test cases as code blocks. In this approach, the source code is divided into code blocks, which improves the accuracy to the limit of coverage-based SFL. Since the existing approach constructed virtual coverage on a coverage basis, there are possible cases where statements are aggregated in some blocks and the sizes of the blocks are too large. Since our approach constructs virtual coverage based on the number of executions, it is expected that large code blocks can be divided and the accuracy improved.

3 OUR APPROACH

An overview of the fault localization in our approach is shown in Fig. 5 and Fig. 6. Fault localization in our approach consists of the following steps.

1. Execute tests of the System Under Test (SUT).
2. Collect a Coverage Report for each test case at test run time and generate an execution count report.
3. Label the test results (Pass or Fail) in the generated execution count report.
4. A DNN model is trained using generated execution count reports and test result labels.
5. Create virtual coverage from the execution count report.
6. Virtual coverage is input to a trained DNN model and a suspiciousness score is given to each code block.
7. Rank the suspiciousness of each code block in descending order of the suspiciousness score.

Figure 5 shows Steps 3 and 4; Fig. 6 shows Steps 6 and 7. Figure 7 shows the DNN architectures. The coverage report in Step 2 describes information such as how many times each statement of the SUT is executed during each test case. An overview of the execution count report in Step 2 is shown in Fig. 8. The execution count report shown in Fig. 8 indicates how many times each statement of the SUT is executed in each test case. Code coverage is represented as 1 if a statement is executed at least once and 0 if it is not. On the other hand, in the execution count report, each statement is represented by the number of execution at test runtime.

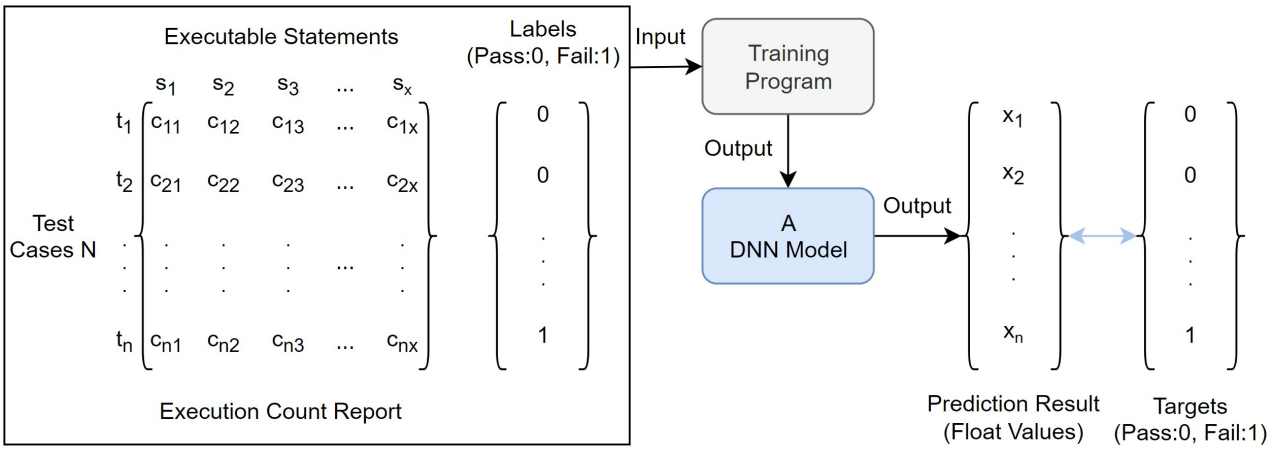


Figure 5: Training A DNN Model in Our Approach

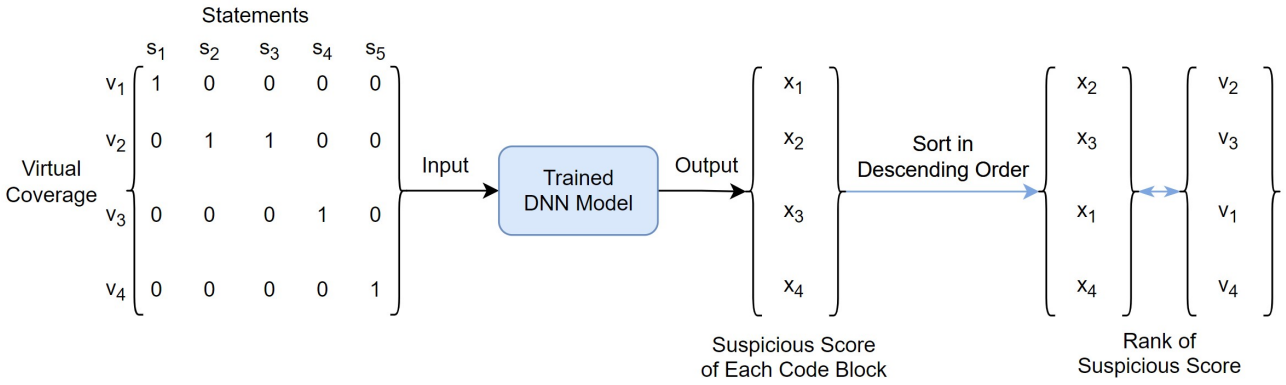


Figure 6: Calculating Suspicious Scores in Our Approach

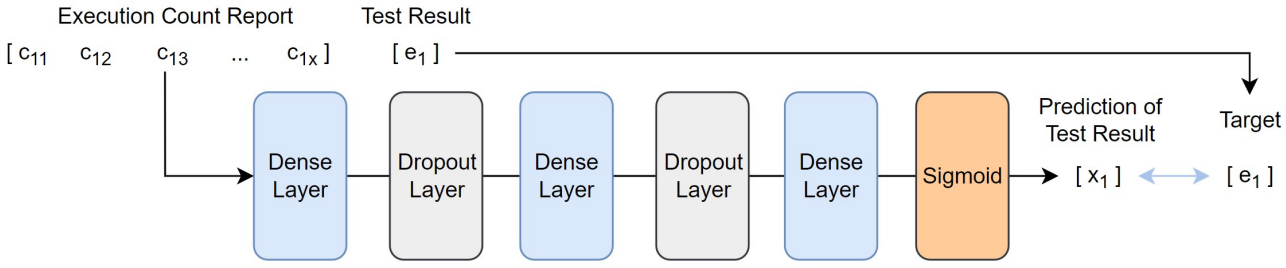


Figure 7: DNN Architecture in Our Approach

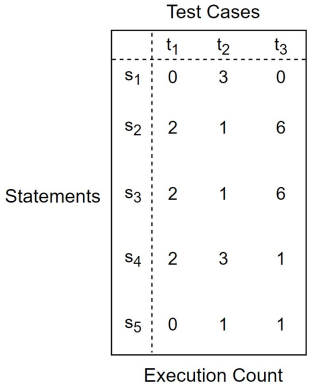


Figure 8: Execution Count Report

In step 4, the developer can use a DNN model that takes the execution count reports as input and classifies the test results. The output value of this DNN model is a float value between 0 and 1, which is regarded as the probability that the classification result is a Fail. A DNN model learns the difference in pattern between the execution count reports of Pass test and Fail test.

The next step is to create virtual coverages that will be input to a trained DNN model. Virtual coverage is code coverage that virtually represents that only a certain code block is executed. The developer creates a virtual coverage and inputs it to a DNN model. Since the output of a DNN model is considered to be the probability that the classification result is Fail, the output value of a DNN model when a virtual coverage containing the faulty statement is input is expected to be higher than other input values. Therefore, the output values of a DNN model in descending order are treated as the rank of the final suspiciousness score, and a suspiciousness rank is assigned to each code block.

The basic idea of SFL (Steps 6 and 7) is the same as the existing approach [11] and is not a new contribution of this paper. We propose a DNN-based virtual coverage fault localization approach based on execution count information. Our proposed approach is used in Steps 4 and 5. The details of our proposed approach are described below.

3.1 Training a DNN Model Using Execution Count Reports

Our approach is to train a DNN model based on execution count information. For this purpose, we use an execution count report. An overview of the execution count report is shown in Fig. 8, which describes the number of times each statement is executed at test runtime.

In our approach, we use execution count reports as training data for a DNN model. Using the execution count reports as training data is expected to improve the accuracy of SFL with our proposed new virtual coverage described in Section 3.2.

3.2 Create Virtual Coverage with Execution Count Reports

The following is a step-by-step description of how to create virtual coverage using execution count reports.

1. In each test case, statements that are adjacent and have the same number of executions shall be temporary blocks.

2. Statements contained in a common temporary block are defined as code block.

3. Create virtual coverage based on defined code blocks.

An overview of each Step is shown in Fig. 9. In Fig. 9, the dotted squares are temporary code blocks and the blue squares are the final code blocks to be defined. In Step 1, temporary code blocks are defined in each test case. In test case 1 (t_1), three temporary code blocks are defined, and in t_2, four temporary code blocks are defined. Each temporary code block is assigned a block id. The color of the dotted line in Step 1 indicates the id of the temporary code block. For example, statement 1 (s_1) belongs to the same temporary code block with the same id in all test cases.

In Step 2, statements that belong to the temporary code block with the same id in all test cases are defined as the final code block. Since s_1 belongs to the temporary code block with the same id (color) in all test cases, we define it as the final code block (blue square). Next, since s_2, and s_3 belong to the same id (color) in all test cases, they are defined as the final code block. At this time, the temporary code block id of s_4 in t_1 is changed to the same id (color) as the next statement, s_5. Next, in Step 2-2, s_4 is defined as the final code block, and the temporary code block id of s_5 in t_1, t_3 is changed to the next temporary code block id (yellow). Finally, in Step 2-3, s_5 is defined as the final code block, and four code blocks are defined in the execution count report shown in Fig. 9.

Our proposed virtual coverage treats statements that are common execution count patterns in all test cases as code blocks. Since code blocks can be created according to the actual execution patterns, source code can be divided into more code blocks than the existing approach. The difference between our proposed code block and the existing approach [11] is shown in Fig. 10.

In Fig. 10, our approach creates four code blocks, while the coverage-based existing approach creates three code blocks. By creating code blocks according to the pattern of execution counts, code blocks can be created with a finer granularity than the existing approach.

The virtual coverage is created from the code blocks and the fault location is identified (Steps 6 and 7). We believe that we can achieve higher accuracy than the existing approach by creating virtual coverage with a finer granularity than the existing approach, and by training a DNN model with data that is more suitable for the proposed virtual coverage.

4 EVALUATION EXPERIMENT

4.1 Research Questions

In this paper, the following research questions are investigated in the evaluation experiment.

RQ1. Comparison of fault localization accuracy with existing approach [11] when virtual coverages created based on

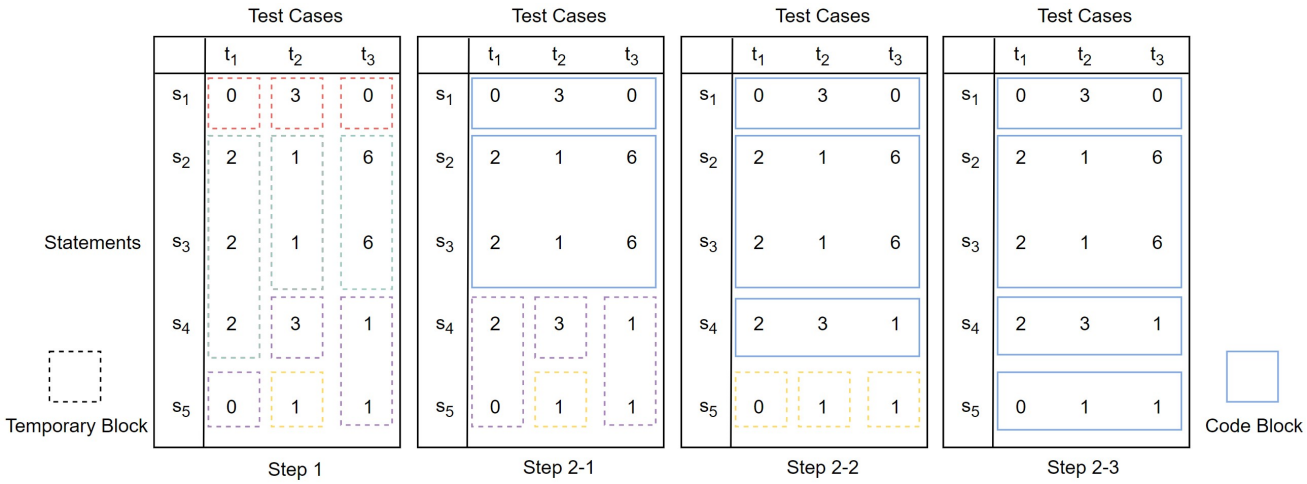


Figure 9: Steps in The Creation of Our Proposed Code Block

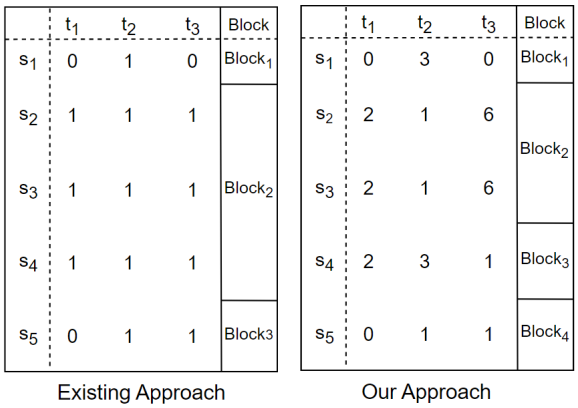


Figure 10: Differences in Code Blocks

the number of executions are input to a DNN model trained with coverages.

In this paper, we propose a new approach for creating a new virtual coverage. We evaluate the effectiveness of our proposed virtual coverage by comparing its fault localization accuracy with existing coverage-based virtual coverage.

The TopN % is used as a measure of the accuracy of fault localization, where the TopN % represents that a bug is classified into the top N % of the total, and a smaller value indicates a better fault localization performance. In the case of multiple bugs, the largest TopN % is used.

RQ2. Comparison of fault localization accuracy with existing DNN model when virtual coverage created based on the number of executions is input.

In our approach (Fig. 5), we use execution count reports as training data to enhance the effectiveness of our proposed virtual coverage. We input the proposed virtual coverage into the existing DNN model and our DNN model, and compare the accuracy of fault localization.

4.2 Subject Programs

We conducted an evaluation experiment using bugs and their fixes provided by Defects4j [12] and Software Infrastructure Repository (SIR) [13]. Defects4J is a database of actual bugs in Java projects. Lang, Math, and Chart from Defects4j and Print_tokens, Print_tokens2, and Tot_info from SIR are selected as the projects for the experiments. We use bugs that meet the following conditions for our experiments.

- Bug fixes only with code addition are excluded. If the bug is fixed by code addition only, the original buggy source code does not have any defects to be pointed out.
- The fault statement is executed at least once in each of the fail and pass tests: the SFL approaches require the bug to be executed in those tests.

Execution coverage and execution count reports are collected using OpenClover [21] and gcov [22]. We use OpenClover to collect coverages of Defects4j's projects, and we use gcov to collect coverages of SIR's projects. Because execution is not recorded for class member variable definitions, etc., due to OpenClover's specifications, we excluded parts of the program that are not recorded as execution coverage from the fault set. The number of lines (LOC) and number of tests for the experimental program are shown in Table 1.

Table 1: Details of Target Programs

Project	Number of Versions	LOC	Number of Tests
Lang	30	58389	54987
Math	29	23623	83364
Chart	15	10094	27036
Print_tokens	7	336	4130
Print_tokens2	9	343	2064
Tot_info	23	268	1052

Figure 11: Results in RQ1 of Defects4J

4.3 Setup for a DNN Model

In performing supervised training, the weight parameters of Dense Layers are initialized with random values. The size of the hidden layer of each Dense Layer is changed according to the size of the coverage size to be trained. The first Dense Layer is the minimum size of 100, and the second Dense Layer is the minimum size of 20 hidden layers. The third Dense Layer is the Output Layer, which is a hidden layer of size 1. The ReLu and Sigmoid functions are used as activation functions. In addition, a Dropout Layer is used to suppress over-learning. Adam optimizer [23] is used, and the learning rate is set to $1.0e-3$. TensorFlow (ver. 2.12.0) is used as the learning framework.

5 RESULTS AND DISCUSSIONS

5.1 RQ1: Result

Figure 11 shows the experimental results for RQ1 of Defects4J. The horizontal axis in Fig. 11 represents the TopN % and indicates the amount of source code examined by the developer. The vertical axis shows the percentage of identified faults, where 100 % on the vertical axis means that all faults are identified. For example, a vertical axis plot with a Top 50% is more than 90%, indicating that more than 90% of the faults are identified by investigating half of the source code. There are two plots in Fig. 11: the gray plot shows the accuracy of the existing approach, and the orange plot shows the accuracy of the proposed virtual coverage.

As a result, except for the 25% and 35% TopN % plots, the proposed virtual coverage has a larger value in the vertical axis plot. This means that for each TopN %, the number of faults that can be identified is higher than with the existing approach.

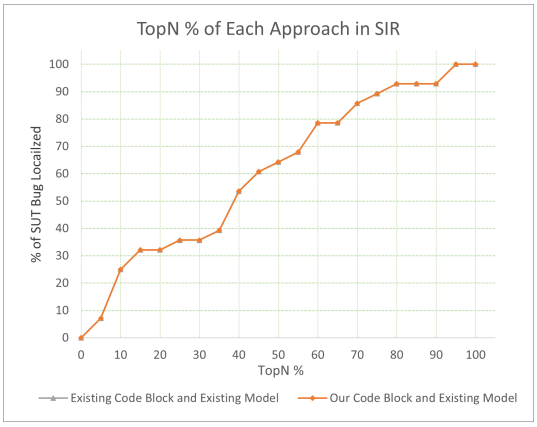


Figure 12: Results in RQ1 of SIR

Table 2: Test Results in RQ1

Subject	1-tailed (left)	1-tailed (right)
Defects4J	3.534E-02	9.647E-01
SIR	5.671E-02	9.433E-01

Figure 12 shows the experimental results for RQ1 of SIR. In Fig. 12, there is no difference in accuracy between the existing and proposed approaches in SIR. Figure 12 shows the accuracy at a granularity of Top 5%, so there is no difference at all, but at a finer granularity, the proposed approach is slightly more accurate ($< 5\ points$) than the existing approach.

Table 2 shows the results of adapting the Wilcoxon Signed-Rank Test to the experimental results. The Wilcoxon Signed-Rank Test is a nonparametric test that tests for differences between two corresponding groups. The null hypothesis indicates that there is no significant difference between the two groups, while the alternative hypothesis indicates that there is a significant difference between the two groups. The 1-tailed test and the alternative hypotheses are listed below.

- **1-tailed (left):** The proposed approach has a smaller TopN % value than the existing approach.
- **1-tailed (right):** The proposed approach has a larger TopN % value than the existing approach.

Since the value of TopN % is the amount of code investigated by the developer to identify the fault location, a smaller value indicates a high accuracy. Therefore, if the left-tailed Wilcoxon Signed-Rank test is accepted, the proposed approach is significantly better than the existing approach in fault localization performance.

The test results of Defects4J show that $p = 0.035 < 0.05 = \alpha$, so the proposed virtual coverage is significantly more accurate than the existing approach. However, the test results of SIR show that $p = 0.057 > 0.05 = \alpha$, so the proposed virtual coverage is not significantly more accurate than the existing approach.

Table 3: Percentage Increase of The Number of Code Blocks in the Defects4J Project

Project	Bugs with Improved Accuracy	Bugs with Decreased Accuracy
Lang	52.36 %	98.54 %
Math	65.34 %	73.01 %
Chart	21.00 %	32.69 %

Table 4: Average Number of Source Code Divisions in Each Project

Project	Existing Approach	Our Approach
Lang	143.86	294.93
Math	80.68	120.93
Chart	65.93	78.33
Print_tokens	120.25	120.25
Print_tokens2	140.63	140.63
Tot_info	72.72	74.00

5.2 RQ1: Discussions

In our approach, the source code is divided into many more code blocks than the existing approach. We consider that defining code blocks with a finer granularity than the existing approach according to the actual execution pattern (number of executions) contributed to the improved accuracy shown in Fig. 11.

On the other hand, for some bugs, the proposed approach showed lower accuracy than the existing approach. Table 3 shows the percentage of increased number of code blocks for each project when the proposed approach achieved higher accuracy than the existing approach, and the proposed approach showed lower accuracy than the existing approach. In Table 3, for all three projects, bugs (Version) with improved accuracy over the existing approach show a smaller percentage increase in the number of code blocks than those with decreased accuracy. Existing research [11] has reported that virtual coverage using code blocks is more accurate than a statement-based virtual coverage approach. Therefore, our approach can define code blocks at a finer granularity than existing virtual coverage, but we consider that the accuracy decreases when the granularity is too fine. In particular, if the granularity of the code blocks defined by our approach is as fine as the statement level, the accuracy is expected to decrease. Future work is needed to determine the level of granularity that will improve accuracy the most.

The results in Fig. 12 show that the proposed approach has no difference in accuracy compared to the existing approach in the SIR project. Table 4 shows the average number of source code divisions (average number of code blocks) for the existing and proposed approaches in each project. In Table 4, the three Defects4j projects (Lang, Math, and Chart) show a difference in the number of source code divisions between the existing and proposed approaches. On the other hand, in the three projects of SIR (Print_tokens, Print_tokens2, Tot_info), there is almost no difference in the number of source code divisions. Therefore, it is considered that the proposed virtual coverage by itself cannot achieve higher accuracy than the existing approach for programs where the number of source code divisions is almost the same as the existing approach.

Figure 13: Results in RQ2 of Defects4J

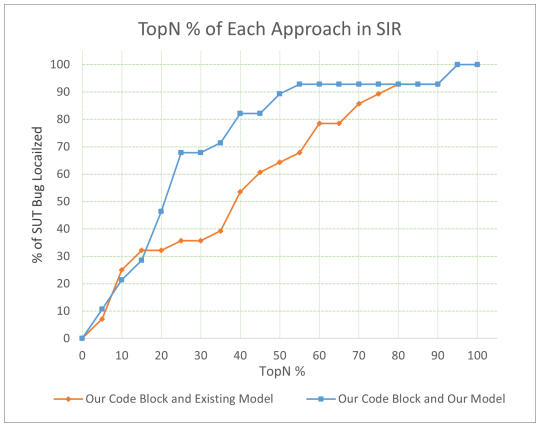


Figure 14: Results in RQ2 of SIR

5.3 RQ2: Result

Figure 13 shows the experimental results for RQ2 of Defects4J. The orange plots in Fig. 13 show the fault localization accuracy when our virtual coverage is input to an existing DNN model. The blue plot shows the accuracy when our proposed virtual coverage is input to a DNN model trained with the execution count reports. The values shown in the two plots are almost identical, and there is no improvement in fault localization performance using our proposed DNN model.

Figure 14 shows the experimental results for RQ2 of SIR. The proposed approach identifies more faults than the existing approach in SIR.

Table 5 shows the results of adapting the Wilcoxon Signed-Rank Test to the experimental results in RQ2. The test results of Defects4J show that $p = 0.421 > 0.05 = \alpha$, so the proposed DNN model is not significantly more accurate than the existing approach. The test results of SIR show that

Table 5: Test Results in RQ2

Subject	1-tailed (left)	1-tailed (right)
Defects4J	4.205E-01	5.795E-01
SIR	1.818E-03	9.982E-01

$p = 0.00182 < 0.05 = \alpha$, the proposed DNN model is significantly more accurate than the existing approach in SIR projects.

5.4 RQ2: Discussions

In Fig. 13, our proposed DNN model did not improve the accuracy compared to the existing DNN model. In Defects4j, we consider that the accuracy based on our proposed virtual coverage (orange plot) is the maximum performance of the SFL method [8] and that the accuracy cannot be improved any further.

Figure 14, the experimental results in SIR, shows that using our DNN model significantly improves the accuracy compared to the existing DNN model. In addition, Table 5 shows that accuracy is significantly improved by using our DNN model. Our proposed virtual coverage is created based on the number of executions of each statement. We consider that training the DNN model with the information used to create our virtual coverage contributed to the improvement in accuracy shown in Fig. 14.

Defects4j showed no improvement in accuracy with our DNN model, while SIR showed a significant improvement (Table 5). It is a future challenge to investigate for which programs our DNN model is effective.

6 CONCLUSION

In this paper, we propose a new virtual coverage to be used for existing DNN-based SFL approaches. Our proposed virtual coverage is created based on the number of executions and is able to divide the source code with finer granularity than the existing virtual coverage. In order to improve the accuracy of fault localization using the proposed virtual coverage, we also proposed a DNN model using execution count reports as training data. Since the proposed virtual coverage is created based on the execution count reports, we consider that training a DNN model with the execution count reports enhances the effectiveness of our proposed virtual coverage.

Our approach is evaluated with six different projects available on Defects4j and SIR. Experimental results show that our proposed virtual coverage is more accurate than existing virtual coverage. We also obtained the prospect of further improving the accuracy by using our proposed DNN model.

Further evaluation of our approach in broader and larger SUTs is needed and is a topic for future work. In the future, we intend to improve a DNN model in our approach to identify faults that cannot be addressed (e.g., faults for which the only bug fix is to add code, performance bugs, etc.).

ACKNOWLEDGMENT

Part of this work is supported by fund from Mitsubishi Electric Corp. The research is also being partially conducted as Grant-in-Aid for Scientific Research C (21K11826).

REFERENCES

[1] G. Tassey, "The economic impacts of inadequate infrastructure for software testing," (2002).

[2] H. A. de Souza, M. L. Chaim, and F. Kon, "Spectrum-based software fault localization: A survey of techniques, advances, and challenges," (2016).

[3] A. Perez and R. Abreu, "A qualitative reasoning approach to spectrum-based fault localization," in *Proceedings of the 40th International Conference on Software Engineering: Companion Proceeedings*, ser. ICSE '18. New York, NY, USA: Association for Computing Machinery, (2018), pp. 372–373.

[4] Q. I. Sarhan and A. Beszédes, "A survey of challenges in spectrum-based software fault localization," *IEEE Access*, vol. 10, pp. 10 618–10 639, (2022).

[5] R. Abreu, P. Zoeteweij, and A. J. Van Gemund, "An evaluation of similarity coefficients for software fault localization," in *2006 12th Pacific Rim International Symposium on Dependable Computing (PRDC'06)*, (2006), pp. 39–46.

[6] J. Jones, M. Harrold, and J. Stasko, "Visualization of test information to assist fault localization," in *Proceedings of the 24th International Conference on Software Engineering. ICSE 2002*, (2002), pp. 467–477.

[7] G. Laghari, K. Dahri, and S. Demeyer, "Comparing spectrum based fault localisation against test-to-code traceability links," in *2018 International Conference on Frontiers of Information Technology (FIT)*, (2018), pp. 152–157.

[8] Z. Zhang, Y. Lei, X. Mao, M. Yan, L. Xu, and X. Zhang, "A study of effectiveness of deep learning in locating real faults," *Information and Software Technology*, vol. 131, p. 106486, (2021).

[9] W. E. Wong, V. Debroy, R. Golden, X. Xu, and B. Thuraisingham, "Effective software fault localization using an rbf neural network," *IEEE Transactions on Reliability*, vol. 61, no. 1, pp. 149–169, (2012).

[10] Z. Zhang, Y. Lei, X. Mao, M. Yan, L. Xu, and J. Wen, "Improving deep-learning-based fault localization with resampling," *J. Softw. Evol. Process*, vol. 33, no. 3, mar (2021).

[11] H. Kiryu, S. Ogata, and K. Okano, "Improve measuring suspiciousness of bugs in spectrum-based fault localization with deep learning," in *Proceedings of International Workshop on Informatics*, ser. IWIN '22. Kii-Katsuura, Japan: Informatics Laboratory, (2022), pp. 3–8.

[12] R. Just, D. Jalali, and M. D. Ernst, "Defects4j: A database of existing faults to enable controlled testing studies for java programs," in *Proceedings of the 2014 International Symposium on Software Testing and Analysis*, ser. ISSTA 2014. New York, NY, USA: Association for Computing Machinery, (2014), pp. 437–440.

[13] H. Do, S. Elbaum, and G. Rothermel, "Supporting controlled experimentation with testing techniques: An infrastructure and its potential impact," *Empirical Software Engineering*, vol. 10, pp. 405–435, 10 (2005).

[14] R. Abreu, P. Zoeteweij, and A. J. van Gemund, "On the accuracy of spectrum-based fault localization," in *Testing: Academic and Industrial Conference Practice and Research Techniques - MUTATION (TAICPART-MUTATION 2007)*, (2007), pp. 89–98.

[15] W. Masri, "Fault localization based on information flow coverage," *Software Testing, Verification and Reliability*, vol. 20, no. 2, pp. 121–147, (2010).
[16] W. E. Wong, V. Debroy, R. Gao, and Y. Li, "The dstar method for effective software fault localization," *IEEE Transactions on Reliability*, vol. 63, no. 1, pp. 290–308, (2014).
[17] Lucia, D. Lo, and X. Xia, "Fusion fault localizers," in *Proceedings of the 29th ACM/IEEE International Conference on Automated Software Engineering*, ser. ASE '14. New York, NY, USA: Association for Computing Machinery, (2014), pp. 127–138.
[18] S. Murtaza, N. Madhavji, M. Gittens, and A. Hamou-Lhadj, "Identifying recurring faulty functions in field traces of a large industrial software system," *Reliability, IEEE Transactions on*, vol. 64, pp. 269–283, 03 (2015).
[19] J. Sohn and S. Yoo, "Fluccs: Using code and change metrics to improve fault localization," in *Proceedings of the 26th ACM SIGSOFT International Symposium on Software Testing and Analysis*, ser. ISSTA 2017. New York, NY, USA: Association for Computing Machinery, (2017), pp. 273–283.
[20] Y. Li, S. Wang, and T. N. Nguyen, "Fault localization with code coverage representation learning," in *Proceedings of the 43rd International Conference on Software Engineering*, ser. ICSE '21. IEEE Press, (2021), pp. 661–673.
[21] "OpenClover," https://openclover.org/.
[22] "gcov - A Test Coverage Program," https://gcc.gnu.org/onlinedocs/gcc/Gcov.html.
[23] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, Y. Bengio and Y. LeCun, Eds., (2015).

(Received: November 15, 2023)
(Accepted: January 19, 2024)

Takuma Ikeda is a graduate student of Shinshu University. His areas of interest include fault localization.

Hitoshi Kiryu is a graduate student of Shinshu University. His areas of interest include formal verification.

Satoshi Suda received his M.S. degree in mathematics from Osaka University, Osaka, Japan, in 2016. He joined Mitsubishi Electric Corp.Currently he is a researcher of Solution Engineering Dept. at Advanced Technology R&D Center.

Shinpei Ogata is an Associate Professor at Shinshu University, Japan. He received his BE, ME, and PhD from Shibaura Institute of Technology in 2007, 2009, and 2012 respectively. From 2012 to 2020, he was an Assistant Professor, and since 2020, he has been an Associate Professor, in Shinshu University. He is a member of IEEE, ACM, IEICE, IPSJ, and JSSST. His current research interests include model-driven engineering for information system development.

Kozo Okano received his BE, ME, and PhD degrees in Information and Computer Sciences from Osaka University in 1990, 1992, and 1995, respectively. He was an Assistant Professor and an Associate Professor of Osaka University. In 2002 and 2003, he was a visiting researcher at the Department of Computer Science of the University of Kent in Canterbury, and a visiting lecturer at the School of Computer Science of the University of Birmingham, respectively. Since 2020, he has been a Professor at the Department of Electrical and Computer Engineering, Shinshu University. Since 2023, he has been the Director of Center for Data Science and Artificial Intelligence. His current research interests include formal methods for software and information system design, and applying deep learning to Software Engineering. He is a member of IEEE, IEICE, and IPSJ.

Invited Paper

A Compiler System Supporting Memory Shared by Heterogeneous Machines

Hitoshi Aida* and Shiro Kawai**

*Emeritus Professor, The University of Tokyo, Japan
**Scheme Arts, LLC, USA
hitoshi-aida@g.ecc.u-tokyo.ac.jp, shiro@acm.org

Abstract - There are several cases in which non-native data representations should be handled. In this paper, a compiler system to support memory shared by heterogeneous machines is discussed. A keyword standard for representation-type-specifier and another keyword shared for storage-class-specifier are added to language C. A prototype compiler is implemented on SPARCstation (SunOS 4.1.3) and IBM compatible PC (386bsd-0.1). Overhead of endian conversion is small for the machines equipped with byte-swap instruction.

Keywords: Heterogeneous shared memory, C language extension, Representation type, Endian conversion.

1 INTRODUCTION

As shown in Table 1, data representation of compiler systems of computer languages such as C [4] varies in many aspects reflecting CPU architecture, OS environment and so on. Usually, the system only supports the representation by which programs are most efficiently executed. We call it 'native'.

There are several cases, however, representations other than native should be handled. For example, in TCP/IP, the most heavily used communication protocol in the Internet, addresses and data lengths should be represented by most significant byte first order (big endian). In TCP/IP programming, macros such as htonl() (host to network long) or ntohs() (network to host short) are usually used to ensure correct data representation in TCP/IP headers and other places.

The authors used to build a sensor data dispatch system in which the sensor data written by one of the workstations to GM(Global Memory) on its external bus is captured by DBC(Distributed Broadcast Controller), is broadcast through communication bus, and is eventually copied to GM of other workstations. The workstations need not be same machine model. In this environment, in addition to the difference of data representation owing to different CPU architectures, difference of the address of GM seen from those CPUs due to the difference of OS environment becomes problem, especially if pointer variables are placed in GM.

In this paper, assuming that memory is shared by machines with different CPU architecture or OS environment, such as shown above, a compiler system which realizes both efficient access to the shared memory and high source code portability by extending C language is discussed.

Table 1: Example difference of data representation

CPU architecture dependent:
- Byte ordering of multibyte data (endian) [5]
- Alignment of multibyte data
- Representation of negative/floating point numbers

OS dependent:
- Character code set
- Memory mapping scheme

Language dependent:
- Representation of character strings

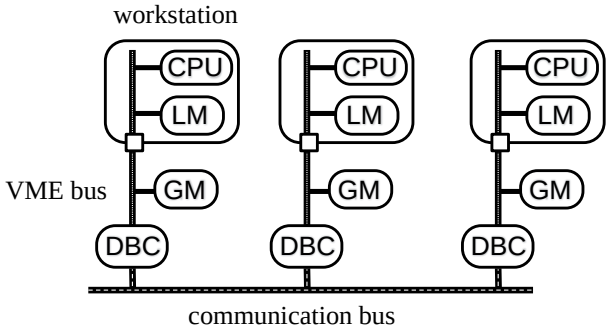


Figure 1: Distributed broadcast memory shared by heterogeneous machines

2 EXTENSION OF C LANGUAGE

2.1 Operational Types and Representation Types

Data type declaration in standard C consists of three parts, 'type-qualifier', 'storage-class-specifier' and 'type-specifier', as roughly shown in Fig 2. Within these, 'type-specifier' is the most essential part which specifies the set of operations applicable to the type and the meaning of those operations. For example, bit shift operations and bitwise logical operations can be applied to char and int, but not to float or double. Another example is that the difference operation of two pointers is only allowed when those pointers point to the same data type except for using explicit cast, and is measured by the size of objects they point.

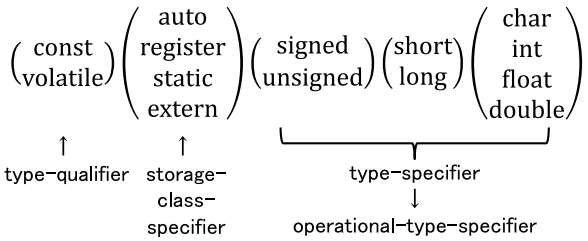


Figure 2: Data types in standard C

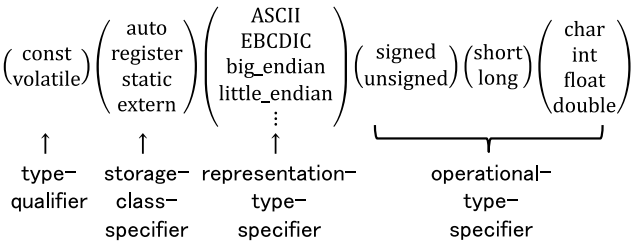


Figure 3: Representation type specifier (not final form)

Here we call the data type specified by 'type-specifier' of standard C as 'operational type'. We also change the name 'type-specifier' to 'operational-type-specifier' for clarity.

As in the case of shared memory described before, different representations of the same operational type could be used. To handle such different representations efficiently and in a portable manner, direct language support is desirable. Here we propose inserting 'representation-type-specifier' in the type declaration in C, such as shown in Fig 3.

When the program accesses variables of which representation type is different from native, automatic conversion is performed. This is much the same as char type variables are converted to int when read and upper bits are truncated when written in standard C.

2.2 Representation Type of Pointers

Pointers are heavily used in C. As for pointers, representation type of both the pointer themselves and the objects pointed by pointers should be considered. Of these, if the representation type of the objects pointed by the pointers are different, the pointers should be supposed to be different operational type. As previously mentioned, in standard C, difference operation of two pointers is only allowed when two pointers point to same data type. Similarly, if pointers are assigned by force using cast to other pointers with different pointed representation type, usage of assigned pointers afterwards may produce unexpected results.

As for the representation type of pointers themselves, in addition to the byte ordering as with integer or floating point data, it should cope with the fact that shared memory might reside different address space reflecting different OS environments. Typical solutions are: (a) represent pointers as the offset from top address of the shared memory, and (b) represent pointers as the relative offset from the pointers themselves. Figure 4 shows a latter example.

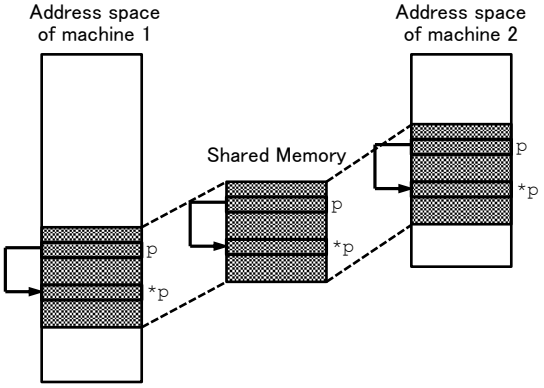


Figure 4: A pointer in the shared memory represented by relative offset

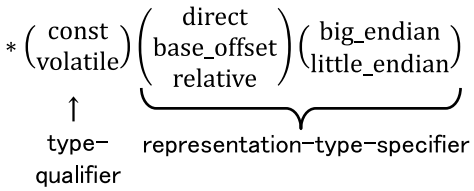


Figure 5: Representation type of pointers (not final form)

In standard C, type-qualifiers such as const can be placed both before and after the pointer declaration mark * to distinguish whether the object pointed to the pointer is not allowed to be assigned or the pointer itself is not allowed to be assigned. In the extended C, in addition to type-qualifiers, we allow representation-type-specifiers to be placed after pointer declaration mark * to specify the representation type of the pointer itself.

2.3 Representation Type Standard and Storage Class Shared

Thinking of typical application of heterogeneous share memory system, it is unlikely to use many different representation types in a single application. It is probably enough to use two representation types in each machine: one for shared variables and one for the native representation type of the machine. Thus, in the proposed extended C language, instead of describing explicit representation-type-specifiers such as big_endian or base_offset, we allow only one representation-type-specifier keyword standard in the source code, and the detailed specification of standard representation type is separately fed to the compiler system. This will increase the portability of source code.

All the data in the shared memory is supposed to be standard representation type. On the other hand, in cases such as swapping two data in the shared memory, we want to hold standard representation type data temporally on registers or in the local memory without converting to native representation.

To specify variables be located in the shared memory, additional keyword shared for storage-class-specifier is added to the language. For standard representation of pointers such as the offset from top address of the shared memory to be meaningful, they should point to the shared memory.

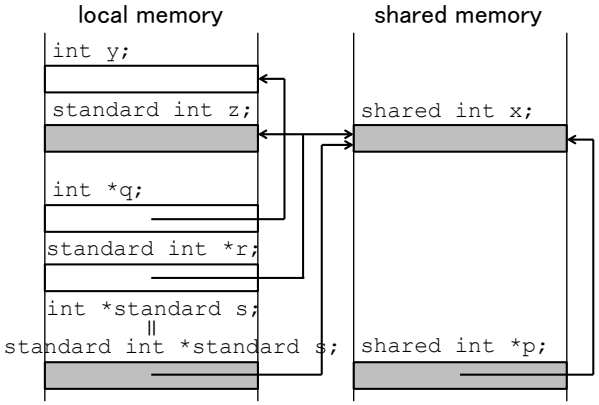


Figure 6: Possible combination of representation type and storage class

Therefore, we assume following constraints between representation type standard and storage class shared:

Constraint 1:
Representation type of variables with storage class shared should be standard.

Constraint 2:
Pointers with representation type standard should point to objects with storage class shared. (For example, they cannot point to functions.)

From these constraints, variables with storage class shared or variables pointed by pointers with representation type standard are implied to have representation type standard:

```
shared int x;
    → shared standard int x;

int *standard s;
    → standard int *standard s;

shared int *p;
    → shared standard int *standard p;
```

There are three combinations for whether the variables are located in the local memory or in the shared memory, and whether the representation type is native or standard for non-pointer variables, and four combinations for direct pointers.

As for structures, all the members of the structures with representation type standard automatically become standard representation type. On the other hand, non-standard representation type structures may have mixture of standard representation type members and non-standard representation type members.

```
standard struct {
    char c;
    int i, j;
    double d;
} t;
            ↓
standard struct {
    standard char c;
    standard int i, j;
    standard double d;
} t;
```

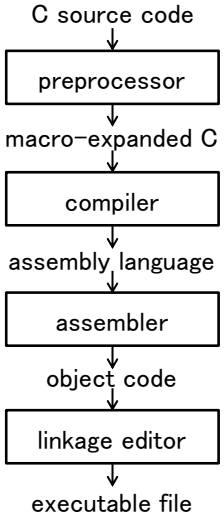


Figure 7: Typical compilation steps for standard C

3 IMPLEMENTATION

Compiler systems of standard C usually consist of several steps as shown in Fig 7. To implement extended C language, best performance is achieved by modifying the compiler in the narrow sense, the second step. However, in cases such as the source code of the compiler for the target CPU architecture is not available, fairly large portion of the features of the extended language can be handled by the preprocessor, the first step. In the following sections, we use preprocessor approach as an example to explain the essence of implementation.

3.1 Representation Type Conversion

Conversion of representation types is the most essential part in the implementation of extended C language.

First of all, compiler system should allocate just enough space for standard representation typed variables. If the conversion is processed by the preprocessor and standard typed pointers are represented as either offset from the top address of shared memory area or relative offset from the pointers themselves, it is natural to convert them to declarations of simple integer variables rather than pointers:

```
standard int x, *p;
        ⇩
long x, p;
```

To access standard representation typed variables, representation type conversion from/to native representation type is needed. In the following, we borrow TCP/IP notations ntohl() for conversion from standard representation to native representation and htonl() for conversion from native representation to standard representation.

```
standard int x;
int y;
```

```
   y = x;
   x = y;
```

⇩

```
long x;
int y;

   y = ntohl(x);
   x = htonl(y);
```

To access the objects pointed by standard-type pointers, conversions are needed for both pointers themselves and the objects pointed by pointers. The following illustrates the possible preprocessor conversion when standard-type pointers are represented by the offsets from top address of the shared memory area. Here _GM denotes the top address of the shared memory area.

```
int *standard p;
int y;

   y = *p;
```

⇩

```
extern void *_GM;
long p;
int y;

   y = ntohl(*(long *)(_GM + ntohl(p)));
```

If standard pointers are represented by relative addresses, preprocessor conversion may look like the following:

```
int *standard p;
int y;

   y = *p;
```

⇩

```
long p;
int y;

   y = ntohl(*(long *)
         ((void *)&p + ntohl(p)));
```

In this case, even simple assignments of pointers needs arithmetic operations:

```
int *standard p, *standard s;

   p = s;
```

⇩

```
long p, s;

   p = htonl(ntohl(s) +
         ((void*)&s - (void*)&p));
```

Byte order (endian) conversion between standard and native representation types can be implemented as C functions. Figure 8(a) shows an example using bit shift and bitwise logical operations, while Fig 8(b) shows an example using a union.

```
long convert_endian(long data)
{
   return (((data << 24) & 0xff000000) |
         ((data << 8)  & 0x00ff0000) |
         ((data >> 8)  & 0x0000ff00) |
         ((data >> 24) & 0x000000ff));
}
```

(a)

```
long convert_endian(long data)
{
   union { long l; char c[4] } in, out;

   in.l = data;
   out.c[0] = in.c[3];
   out.c[1] = in.c[2];
   out.c[2] = in.c[1];
   out.c[3] = in.c[0];
   return out.l;
}
```

(b)

Figure 8: C source code examples for endian conversion

```
or %g0, 255, %l0
sll %i0, 24, %l1
sll %i0, 8, %l2
srl %i0, 8, %l3
srl %i0, 24, %l4
and %l4, %l0, %l4
sll %l0, 8, %l0
and %l3, %l0, %l3
or %l3, %l4, %l4
sll %l0, 8, %l0
and %l2, %l0, %l2
or %l2, %l4, %l4
sll %l0, 8, %l0
and %l1, %l0, %l1
or %l1, %l4, %o0
```

SPARC instruction

```
bswap %eax
```

i486 instruction

Figure 9: Inline assembly expansion for endian conversion

To reduce the overhead of the conversion, expansion of macros similar to Figure 8(a) instead of function call is probably effective.

If the compiler itself can be modified or the existing compiler accepts inline assembler expansions, conversion overhead can be further reduced by embedding machine instructions such as shown in Fig 9. In contrast to SPARC [6] CPU architecture which needs fifteen instructions roughly corresponding to Fig 8(a) for conversion, i486 [7] needs only one instruction which can be executed in about single machine cycle.

3.2 Alignment of Structure Members

For the structures of which representation type is standard, all the offsets of the members from top of the structures should meet the most severe alignment requirement of the machines. To meet this requirement by preprocessor, dummy members are inserted to the declaration of standard structures.

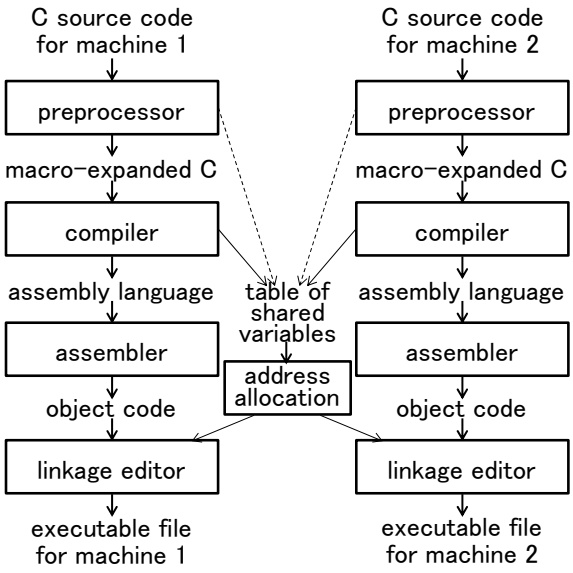


Figure 10: Compilation steps for extended C language

```
standard struct {
    char c;
    int i, j;
    double d;
} t;
```

⇩

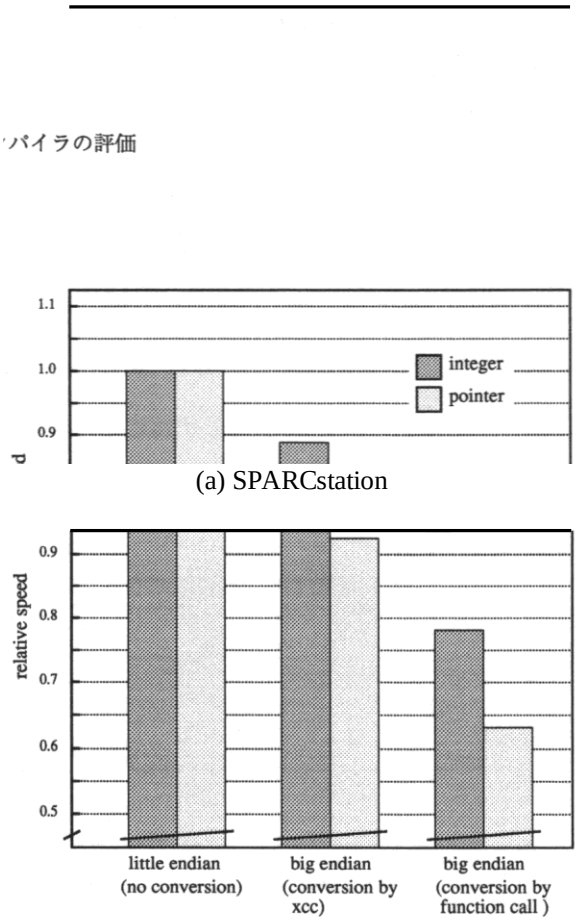
```
struct {
    char c, _dummy1[3];
    long i, j, _dummy2;
    double d;
} t;
```

3.3 Allocation of Shared Variables

If compiled separately, layout of variables in the shared memory may become different among machines. As shown in Fig 10, the compiler (or the preprocessor) is modified to collect all declarations of shared storage class variables in all source files and another compiler step is added to fix the layout of shared memory satisfying the most severe alignment requirement of the machines, which is fed to the linkage editors. Furthermore, a routine to map the shared memory to the address determined at the compilation time should be added to the startup time of the execution in each machine.

4 PERFORMANCE EVALUATION

As one of the experiments to show the practicality and effectiveness of proposed compiler system, we measured the overhead of representation type conversion. In the experiment, two SPARCstations (big endian, SunOS 4.1.3) and one IBM PC compatible machine called i486pc (little endian, 386bsd-0.1) are connected by replicated type shared memory called SCRAMNet [8]. Parallel sorting algorithm known as Batcher's Merge-Exchange Sort [9] was implemented on the hardware. Figure 11 shows performance of single processor execution by relative value to the case of:



(a) SPARCstation

(b) i486pc

Figure 11: Overhead of endian conversion

- standard representation type is set to the same endian as the native representation type and compiled by the proposed compiler system (denoted as ‘no conversion’).

The other two cases are:

- standard representation type is set to the opposite endian to the native representation type and compiled by the proposed compiler system (denoted as ‘conversion by xcc’)

- standard representation type is set to the opposite endian to the native representation type, conversion functions are written in C, and compiled by GNU CC (denoted as ‘conversion by function call’)

In all the three cases, measurement is done for both sorting integer data themselves in the shred memory (denoted as ‘integer’), and sorting shared pointers to the data rather than data themselves (denoted as ‘pointers’).

For both SPARCstation and i486pc, overhead of representation type conversion is reduced by embedding instructions for representation type conversion by the proposed compiler system.

Table 2: Number of instructions in the innermost loop

	SPARCstation		i486pc	
	integer	pointer	integer	pointer
Total number of instructions	52	86	33	42
Number of instructions for endian conversion	30	62	2	8

Especially for i486pc, which has special instruction for endian conversion, overhead is reduced to less than 10%. On the other hand, for SPARCstation, even when the proposed compiler system embed instructions for representation type conversions, overhead reaches 15～25%. Table 2 shows the number of instructions in the innermost loop in the assembly code file. For SPARCstation, the ratio of number of instructions for representation type conversion to the total number of instructions is very large, which results in the large overhead.

Note that, on the hardware used here, absolute performance (MIPS) of SPARCstation exceeds that of i486pc even considering overhead of representation type conversion. To select which endian to use for standard representation type, many factors such as performance of the processors, load distribution, access frequency of shared variables should be considered.

5 CONCLUSION

In this paper, a compiler system to support memory shared by heterogeneous machines is discussed. A keyword standard for representation-type-specifier and another keyword shared for storage-class-specifier are added to language C. A prototype compiler is implemented on SPARCstation (SunOS 4.1.3) and IBM compatible PC (386bsd-0.1). Overhead of endian conversion is small for the machines equipped with byte-swap instruction.

REFERENCES

[1] T. Saito, H. Aida and S. Kawai, “A Compiler System Supporting Heterogeneous Shared Memories”, Journal of Faculty of Engineering, The University of Tokyo, A-31, pp.46-47 (1993) (in Japanese).
[2] S. Kawai, “A Language Processing System for Heterogeneous Distributed Shared Memory”, Master’s Thesis, Graduate School of Engineering, The University of Tokyo (1993) (in Japanese).
[3] S. Kawai, H. Aida and T. Saito, “A Compiler System for Heterogeneous Distributed Shared Memory”, The Special Interest Group Technical Reports of IPSJ, 1992-DPS-058, pp.189-196 (1992) (in Japanese).
[4] B. W. Kernighan and D. M. Ritchie, “The C Programming Language (Second Edition)”, Prentice-Hall (1988).
[5] D. Cohen, “On holy wars and a plea for peace”, IEEE Computer, Vol.14, No.10, pp.48-54 (1981).
[6] SPARC International Inc., The SPARC Architecture Manual Version 8 (1992).
[7] Intel Corporation, i486™ Processor Programmer’s Reference Manual (1990).
[8] SYSTRAN Corporation, SCRAMNet™ Network Reference Manual (1991).
[9] D. E. Knuth, “The Art of Computer Programming”, Volume 3, Addison-Wesley (1973).

(Received: January 27, 2024)
(Revised: February 29, 2024)

Hitoshi AIDA received his Doctor's degrees in electrical engineering from the University of Tokyo in 1985. He joined the University of Tokyo as a research associate in 1985, and was promoted to a lecturer, an associate professor and a professor of the same university in 1986, 1990 and 1999 respectively. He stayed SRI international as an international fellow from 1988 to 1990. His research area includes high quality telecommunication and parallel and distributed computing. He is a fellow of IPSJ and IEICE, a senior member of IEEE, and a member of ACM, EAJ, JSSST, JSAI and IEIEJ.

Shiro Kawai is a software consultant and an actor. He received a Ph.D. in Engineering from the University of Tokyo and worked for interactive graphics, digital content production pipelines, and domain-specific programming languages. His credits include "Final Fantasy" videogame and CG movie franchise at Square USA. He then founded Scheme Arts, LLC, and developed a Scheme scripting engine "Gauche". He also appears as an actor in films, TV shows, and theatres, including "Go For Broke", "Running for Grace", and "Hawaii Five-0". He is a member of ACM and SAG-AFTRA.

Submission Guidance

About IJIS

International Journal of Informatics Society (ISSN 1883-4566) is published in one volume of three issues a year. One should be a member of Informatics Society for the submission of the article at least. A submission article is reviewed at least two reviewer. The online version of the journal is available at the following site: http://www.infsoc.org.

Aims and Scope of Informatics Society

The evolution of informatics heralds a new information society. It provides more convenience to our life. Informatics and technologies have been integrated by various fields. For example, mathematics, linguistics, logics, engineering, and new fields will join it. Especially, we are continuing to maintain an awareness of informatics and communication convergence. Informatics Society is the organization that tries to develop informatics and technologies with this convergence. International Journal of Informatics Society (IJIS) is the journal of Informatics Society.

Areas of interest include, but are not limited to:

- Internet of Things (IoT)
- Smart Cities, Communities, and Spaces
- Big Data, Artificial Intelligence, and Data Science
- Network Systems and Protocols
- Computer Supported Cooperative Work and Groupware
- Security and Privacy in Information Systems
- Intelligent Transportation System
- Distributed Computing
- Multi-media communication
- Information systems
- Mobile computing
- Ubiquitous computing

Instruction to Authors

For detailed instructions please refer to the Authors Corner on our Web site, http://www.infsoc.org/.

Submission of manuscripts: There is no limitation of page count as full papers, each of which will be subject to a full review process. An electronic, PDF-based submission of papers is mandatory. Download and use the LaTeX2e or Microsoft Word sample IJIS formats.

http://www.infsoc.org/IJIS-Format.pdf

LaTeX2e

LaTeX2e files (ZIP) http://www.infsoc.org/template_IJIS.zip

Microsoft Word™

Sample document http://www.infsoc.org/sample_IJIS.doc

Please send the PDF file of your paper to secretariat@infsoc.org with the following information:

Title, Author: Name (Affiliation), Name (Affiliation), Corresponding Author. Address, Tel, Fax, E-mail:

Copyright

Publisher

Address: Informatics Laboratory, 3-41 Tsujimachi, Kitaku, Nagoya 462-0032, Japan

E-mail: secretariat@infsoc.org

CONTENTS