(IJNCAA)

ISSN 2220-9085(ONLINE)
ISSN 2412-3587(PRINT)

INTERNATIONAL JOURNAL OF NEW COMPUTER ARCHITECTURES AND THEIR APPLICATIONS

Volume 5, Issue 3
2015

SDIWC

www.sdiwc.net

International Journal of
NEW COMPUTER ARCHITECTURES AND THEIR APPLICATIONS

Overview

The SDIWC International Journal of New Computer Architectures and Their Applications (IJNCAA) is a refereed online journal designed to address the following topics: new computer architectures, digital resources, and mobile devices, including cell phones. In our opinion, cell phones in their current state are really computers, and the gap between these devices and the capabilities of the computers will soon disappear. Original unpublished manuscripts are solicited in the areas such as computer architectures, parallel and distributed systems, microprocessors and microsystems, storage management, communications management, reliability, and VLSI.

One of the most important aims of this journal is to increase the usage and impact of knowledge as well as increasing the visibility and ease of use of scientific materials, IJNCAA does NOT CHARGE authors for any publication fee for online publishing of their materials in the journal and does NOT CHARGE readers or their institutions for accessing the published materials.

Publisher

The Society of Digital Information and Wireless Communications
Miramar Tower, 132 Nathan Road, Tsim Sha Tsui, Kowloon, Hong Kong

Further Information

Website: http://sdiwc.net/ijncaa, Email: ijncaa@sdiwc.net,
Tel.: (202)-657-4603 - Inside USA; 001(202)-657-4603 - Outside USA.

The issue date is July 2015.

IJNCAA

ISSN 2220-9085 (Online)
ISSN 2412-3587 (Print)

Volume 5, Issue No. 3 *2015*

CONTENTS

ORIGINAL ARTICLES

International Journal of New Computer Architectures and their Applications (IJNCAA) 5(3): 102-106
The Society of Digital Information and Wireless Communications, 2015 ISSN 2220-9085 (Online); ISSN 2412-3587 (Print)

Hybrid Method based Retinal Optic Disc Detection

Arif Muntasa[1], Indah Agustien Siradjuddin[2], and Moch Kautsar Sophan[3]
Informatics Department, University of Trunojoyo Madura, Bangkalan – Madura Island, Indonesia
arifmuntasa@if.trunojoyo.ac.id

ABSTRACT

We propose a hybrid method based for the Optic Disc (OD) detection in the retinal image. This research consists of three main steps. First, blood vessel removal with homomorphic and median filtering. Second, edge detection using canny operator. Third, OD detection using the Hough transform. The Hough transform is used since the objective object is the curve with circle shape, i.e. optic disc region. Therefore, we can find the shape by using the Hough transform with the circle equation. In this research, the generated circles from Hough transform are matched with the edge pixels of the retinal image. The closest match (showed by the maximum value of accumulator) means that the optic disc is detected. The experiments show that the best accuracy is achieved when the distance value between the generated circles is 3. The average sensitivity, specificity, and balanced accuracy are 64.6182575%, 98.58545%, and 81.6018%, respectively.

KEYWORDS

Homomorphic Filtering, Canny Edge Detection, Hough Transform, balanced accuracy, sensitivity, specificity.

1 INTRODUCTION

Nowadays, research on biomedics has been growth increasingly since many advantages are obtained from the research. Especially, the early screening of a certain deathly disease. Some research are conducted to detect the disease automatically using computational artificial intelligence method [1,2]. Therefore, the result of these researches will help the physicist or as the second opinion of finding disease.

This research proposes on automatically finding the location of Optic Disc (OD) in the retinal image using the Hough transform. This research is one of the preliminary researches of early detection of Diabetic Retinopathy disease. Early detection of Diabetic Retinopathy (DR) has been an importance issue since this disease makes the irreversible blindness. To identify the DR disease, some features in the retinal images are important, i.e. microaneurysms, exudates, and haemorrhage [3]. Other features are less important, for instance, blood vessel and optic disc. We will remove the less important features to obtain the important features for DR disease identification; therefore, we have to find the location of blood vessel and optic disc (segmentation), in order to eliminate those features.

We have done several researches to detect the location of blood vessel pixels in the retinal images[4,5]. This research focus on the OD segmentation using Hough transforms. The remainder of the paper is organized as follows, section 2 will describe the Research Method. Section 3 explains the Hough transform for the detection process, and section 5 describes the result and discussion, and finally the conclusion section in section 6.

2 RESEARCH METHOD

Three main stages are required in the Optic Disc (OD) detection of this research. First, preprocessing using homomorphic and median filtering . Second, edge detection using canny operator, and third OD detection using Hough transform. These stages are depicted in Fig. 1.

Fig. 1. Automatic Optic Disc Detection Research Method

In the first stage, homomorphic and median filtering is implemented for blood vessel removal in the retinal images. Since the blood vessel pixels are removed, then only the optic disc pixels and small noises are remained in the retinal image. Therefore, the accuracy of optic disc will increase. The result of homomorphic filtering and median filtering can be seen in Fig. 2.

Fig. 2. (Left) Original retinal image ; (Right) homomorphic and median filtered image

As seen in the Fig.2, other features besides the optic disc pixels are removed. Hence only optic disc and some noises pixels are considered in the next stage.

The Second stage of this research is edge detection process. In this stage, we use canny detection algorithm since it is an optimal edge detection algorithm [6,7]. Edge detection is required since we need to find the certain shape in the retinal image, i.e. circle that is the shape of an optic disc. To ease the finding of circle shape in the image, we need the extract the edge of the retinal image. Therefore, we implement the edge detection using Canny algorithm for the retinal image. The example of the result in this stage is shown in Fig. 3.

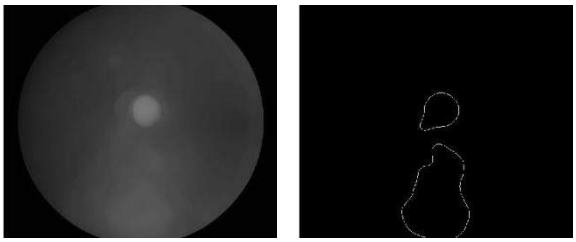


Fig. 3. (Left) Homomorphic and median filtered image ; (Right) edge detection image using Canny algorithm.

As seen in the figure, canny detection algorithm shows the optimal result of edge detection process. This edge detection image is used in the final stage, that is, optic disc detection using Hough transform.

2 HOUGH TRANSFORM

The objective of this research is finding the location of Optic Disc (OD) pixels. The OD pixel in the retinal images is the feature with circle shape. Therefore, in this research we use Hough Transform for this purpose. The Hough transform can detect certain shape in the image using the shape equation [8,9]. The circle shape is represented by the equation as seen in (1).

$$R^2 = (x-a)^2 + (y-b)^2, \quad (1)$$

where (a,b) is the centre coordinate of the circle, R is the radius of the circle, and (x,y) is the pixel coordinates at the edge of the circles shape.

The Hough transform will generate circles using (1) and the information of edge pixels from the previous stage. Hence, the generated circles will only focus in the candidate area of the optic disc pixels. The example of generated circles based on the edge pixels can be seen in Fig.4.

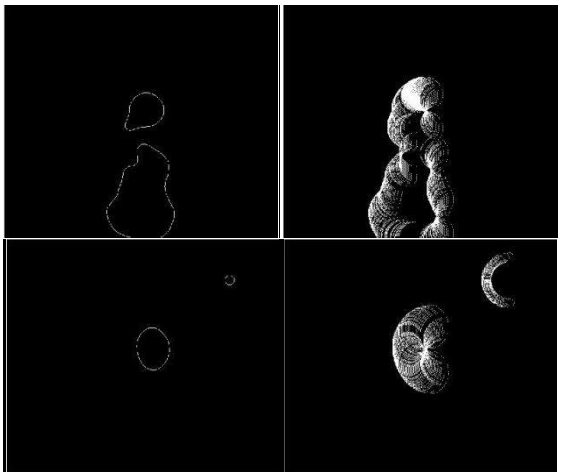


Fig. 4. (Left) Edge Detection Image; (Right) Generated circles using Hough transform

The entirely generated circles are then matched with the edge pixels from the previous stage. The matched pixels between every generated circle and the edge pixels from the edge detection image are calculated. The most maximum number of

matched pixels of the generated circle is considered as the closest match. Hence, the generated circle is the detected optic disc pixels.

4 RESULT and DISCUSSION

Forty retinal images from INSPIRE (Iowa Normative Set for Processing Images of Retina) dataset [10] are used for the experiment in this research. The dataset provides the retinal image and also its Ground truth image. Therefore, we can justify our result of the experiment with this ground truth image. Three kinds of performance measure are used to measure the accuracy of the detection result, i.e. sensitivity, specificity, and balanced accuracy, as seen in (2) , (3), and (4).

$$BalancedAccuracy = \frac{Sensitivity + Specificity}{2}, \quad (2)$$

$$Sensitivity = \frac{TruePositives}{TruePositives + FalseNegatives}, \quad (3)$$

$$Specificity = \frac{TrueNegatives}{TrueNegatives + FalsePositives}. \quad (4)$$

Sensitivity is the probability of true detected optic disc pixel (foreground), specificity is the probability of true detected background pixel, and meanwhile the balanced accuracy is the average accuracy of sensitivity and specificity.
Description of True Positive pixels (TP), True Negative pixels (TN), False Positive pixels (FP), and False Negative pixels (FN) are shown ini Table. 1.

TABLE I. DEFINITION OF TP, TN, FP, AND FN

	Optic Disc Pixels	**Background Pixels**
Detected as Optic Disc Pixels	True Positives	False Positives
Detected as Background Pixels	False Negatives	True Negatives

Three scenarios are conducted in this experiment based on the distance of interval between the generated circles in the Hough transform process, i.e. 5, 3, and 1. The average accuracy of all scenarios is shown in Table 2.

The scenario 1 obtains the lowest sensitivity rate, i.e. 50.883%. Meanwhile, the scenario has the highest specificity rate, i.e. 99.4862. The lowest sensitivity rate is obtained, since in the maximum accumulator of generated pixels from Hough transform only cover the small part of optic disc in the retinal image (see blue pixels on the right image of Fig 5). On the contrary, this result makes the specificity is high, since all pixels beside the detected optic disc pixels considered as background pixels. Left images in the figure show the detected optic disc pixels. Meanwhile the right images show the detected optic disc pixels are superimposed on the original of the retinal images.

TABLE II. AVERAGE ACCURACY RATE FOR THE OPTIC DISC DETECTION

No	**Dist**	**Average Accuracy (%)**		
		Sensitivity	***Specificity***	***Balanced Accuracy***
1	5	50.88359	99.4862	75.1849
2	3	64.6182575	98.58545	81.6018
3	1	56.623305	99.386	78.004675

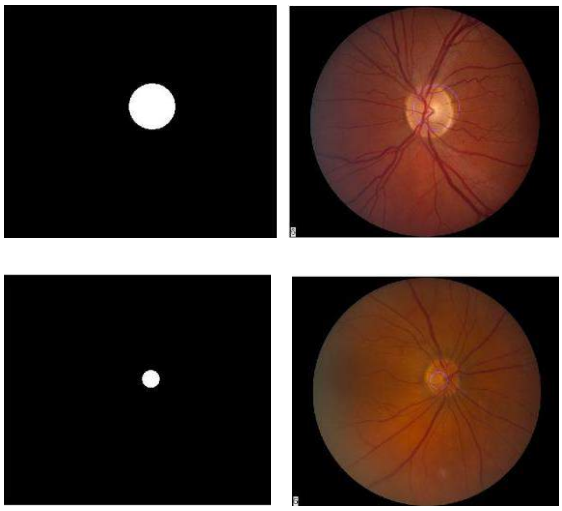


Fig. 5. (Left) Detected Optic Disc pixels and (right) Detected Optic Disc pixels superimposed on original retinal image (blue pixels)

The scenario 2 is the highest balanced accuracy that is 81.6018% accuracy. Moreover, the scenario 2 has the highest sensitivity rate, i.e. 64.6183%, and unfortunately it has the lowest specificity rate, i.e. 98.585%. The example result of this scenario is depicted in Fig 6. As seen in the Figure, the detected optic disc pixels almost cover up all the entire pixels of the optic disc pixels in the original retinal image. Therefore in this scenario, the

highest sensitivity rate is achieved. The accuracy of Hough transform also depends on the edge detection process since generated circles on Hough transform are checked on the edge pixels only.

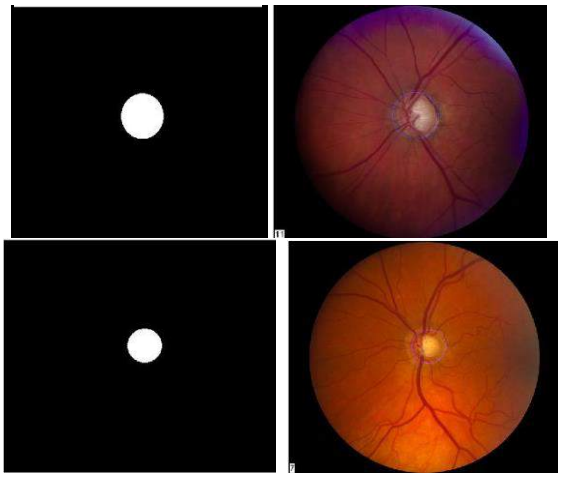


Fig. 6. (Left) Detected Optic Disc pixels and (right) Detected Optic Disc pixels superimposed on original retinal image (blue pixels)

6 CONCLUSION

Detection of the optic disc of the retinal image in this research implements the combination of Homomorphic filtering and Hough transform. The preprocessing stage uses the Homomorphic filtering. The result of this stage is blood vessel pixels are reduced. The main stage of this research, i.e. the detection of optic disc pixels uses the Hough transform. Hough transform generates a number of circles and the circles is matched to the edge image of the retinal image.

The closest match is used as the detected optic disc pixels. The experiment results on the INSPIRE dataset prove that the best value of the distance between generated circles in Hough transform is 3 pixels. The sensitivity, specificity, and balanced accuracy of this detection process are 64.6182575%, 98.58545%, and 81.6018%, respectively. The closer distance of the generated circles makes the accuracy decreased since many noises in the edge image; therefore the Hough transform accumulator makes the false detection.

On the contrary, the further distance makes the accuracy is also decreased since there will be candidate optic disc pixels are missed during the Hough transform process. Hence the best value of the distance in this research is 3 pixels. The Hough transform depends on the edge pixels in the retinal image since the generated circles are matched to the edge pixel, therefore in the future research improvement of the edge detection pixels is recommended.

Acknowledgment

This research is funded by Directorate General of Higher Education of Indonesia for the Hibah Kompetensi Grant.

7 REFERENCES

1. Khan J., Wei J. S., Ringnér M., Saal L. H., Ladanyi M., Westermann F., Berthold F., Schwab M., Antonescu C. R., Peterson C., and Meltzer P. S. : Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. Nature Medicine 7, 673 – 679 (2001).

2. Middleton I. and Damper R. I. : Segmentation of magnetic resonance images using a combination of neural networks and active contour models. Medical Engineering & Physics, 26, 71-86 (2004).

3. Marin, D. Aquino, A., Gegundez-Arias, M.E., Bravo, J.M.: A New Supervised Method for Blood Vessel Segmentation in Retinal Images by Using Gray-Level and Moment Invariants-Based Features. IEEE Trans on Medical Imaging, 146-158 (2011).

4. Muntasa A., Siradjuddin I.A., and Sophan M.K.: Matrix Mask Overlapping and Convolution Eight Directions for Blood Vessel Segmentation on Fundus Retinal Image. Telkomnika 12, 631-638 (2014).

5. Muntasa A., Siradjuddin I.A., and Sophan M.K. : Region Based Segmentation Using Morphological Mathematics And Laplacian Of Gaussian To Automatically Detect Retinal Blood Vessel Image. IJACT 6, 93-104 (2014).

6. Zhou P., Wenjun Y., Xia Y., and Wang Q. : An Improved Canny Algorithm for Edge Detection. Journal of Computational Information System 7, 1516-1523 (2011).

7. Sheikh A., Mandavgane R. N., Khatri D. M.: Review On Canny Edge Detection, IJAICT 1, 744-745 (2015).

8. Yaseen K., Tariq A., and Akram M.U.: A Comparison and Evaluation of Computerized Methods for OD Localization and Detection in The Retinal Images. International Journal of Future Computer and Communications 2, 613-616 (2003).

9. Hassanein A.S., Sherien M., Sameer M., and Ragab M.E.: A Survey on Hough Transform, Theory, Techniques and Applications. IJCSI 12, 139-156 (2015).

10. Sivaswamy J., Krishnadas S.R., Chakravarty A., Joshi G.D., Ujjwal, and Syed T.A.: A Comprehensive Retinal Image Dataset for the Assessment of Glaucoma from the Optic Nerve Head Analysis. JSM Biomed Imaging Data Papers 2, 1-7 (2015).

International Journal of New Computer Architectures and their Applications (IJNCAA) 5(3): 107-118
The Society of Digital Information and Wireless Communications, 2015 ISSN 2220-9085 (Online); ISSN 2412-3587 (Print)

A Framework for Deployment of Easy to Use Wireless Sensor Networks for Farm Soil Monitoring and Control: A case Study of Horticulture Farms in Pwani Region Tanzania

Kosmas Kapis and Mathias Ombeni
College of Information and Communication Technologies,
University of Dar es Salaam, P.O. Box 33335, Dar es Salaam,
Tanzania.
kapis@udsm.ac.tz, kkapis@gmail.com, ombenitz@yahoo.com

ABSTRACT
Monitoring and control of soil parameters such as temperature, moisture and micronutrients play an important role in producing quality horticulture crops. The farms at Pwani Region needed close monitoring and control of soil parameters because the type of soil in this place is alluvial which can drain water and loose macronutrients easily.
Recently Wireless Sensor Network (WSN) has been used to monitors soil parameters, however many systems developed from existing frameworks were unable to meet requirements for small and medium scale horticulture farmers.
This paper addresses the challenges mentioned above and provides solution that fits the left gap by improving previous frameworks. The gap was identified after comparing different WSN related technologies and requirements obtained from experts and farmers. The framework designed from this information was used to develop WSN system which can be easily used by farmers of different levels of education.

KEYWORDS

Precision agriculture; Sensor; Wireless Sensor Network; Monitoring; WSN Framework; Horticulture Crop; Soil Parameters.

1. INTRODUCTION

The Tanzanian government has made a special emphasis on agricultural export diversification stressing the switch from traditional to non-traditional exports such as horticulture products like flowers, vegetables and manufactured goods [5]. The underutilization of available resources and poor soil management has led to land degradation. Agriculture in Tanzania is dominated by small farm holder farmers (peasants) cultivating an average farms of small sizes mostly by using hoes. The major factor affecting the agriculture sector is the lack of precise utilization of resources caused by lack of proper technology to provide current information which can help in proper and timely decision making. [3].
Horticulture is a sector which has been identified as one of the important sectors in the “Kilimo Kwanza” Resolution. The resolution which in English means “Priority Agriculture” , is expected to become one of the main sources of foreign exchange earnings and a significant driver of economic growth [5].
Horticulture crops need close monitoring and control compared to other crops in order to maintain crop qualities and increase production. In response to this, there is a need of having a technology that would help farmers to monitor and control soil parameters in real time.
Currently, there are technologies that have been used to solve this problem. These include grid soil sampling, yield monitoring and crop scouting. However, despite these technologies being non real-time, they involved use of expensive technologies and they were labour intensive [9].
Wireless Sensor Network (WSN) technology has been used for many years in different countries to collect soil parameters information. In contrast to this, in Tanzania, agriculture parameters are still obtained by collecting soil and water samples for laboratory analyses [6]. This method requires a highly human involvement and it is not conducted in timely manner. Furthermore, the method is normally not specifically tailored for specific area or specific crop and reports do not reach farmers in the right time. WSN technology can operate in a wide range of environments and provide advantages in cost, size, power and flexibility compared to other technologies [10].

According to [1], every deployment has its own needs imposed by the type of the monitored crop and other special application design requirements. To develop and deploy WSN which can meet both application and user requirements, the framework has to be well designed. This can be done by using well selected type of sensors, nodes, topology, data fusion algorithms, dissemination protocols, and type of suitable interface [11].

The designed WSN framework makes it easier to work with complex technologies; it ties together a bunch of discrete components into something more useful [12]. In this study, the proposed WSN framework required to meet the specific requirements of horticulture crop and their related design requirements which fit low and medium scale farmers with little or no Information and Communication Technology (ICT) literacy.

Much research has been done on deployment of WSN for different applications like precision agriculture and so forth. Some frameworks /guidelines have been proposed. However, many of these frameworks are not complete and practical as they just mention only few technologies like routing protocols. For those few developed frameworks they do not lead to systems which can be easily used by users who are ICT illiterate and with little knowledge on horticulture sector.

Many current WSN systems use coordinators equipped with both ZigBee module for intra communication and GPRS module for external communication through cellular network. When both modules were powered up, the ZigBee module was losing connection resulting in a total network failure which required manual reset [4]. This type of design is not proper for horticulture crop because as it was mentioned earlier in this paper, horticulture crops are different from other crops as they can easily be affected by slight changes in agriculture parameters (if not quickly identified and corrected). Hierarchical topologies used by many WSNs are not scalable and are difficult to maintain; meaning when someone needs to change position of a node in the farm it can be difficult for a farmer to reconfigure the system [13].

Mafuta[4] presented a Wireless Sensor Network for Precision Agriculture developed in Malawi in Manja Township, City of Blantyre (shown in figure 1). This system combined sensors and actuators in a wireless sensor/actuator network so as to guide successful deployment of WSNs for Precision Agriculture. The system collects soil parameter values like Temperature and moisture and sends to the coordinator node. This node is equipped with both ZigBee and GPRS modules. GPRS module was used to connect to the cellular network and ZigBee was used for intra communication. In this architecture when both modules were powered up, the ZigBee module was losing connection resulting in a total network failure which required manual reset [5]. This design is not good for developing WSN for horticulture crop because it may not able to perform real-time communication.

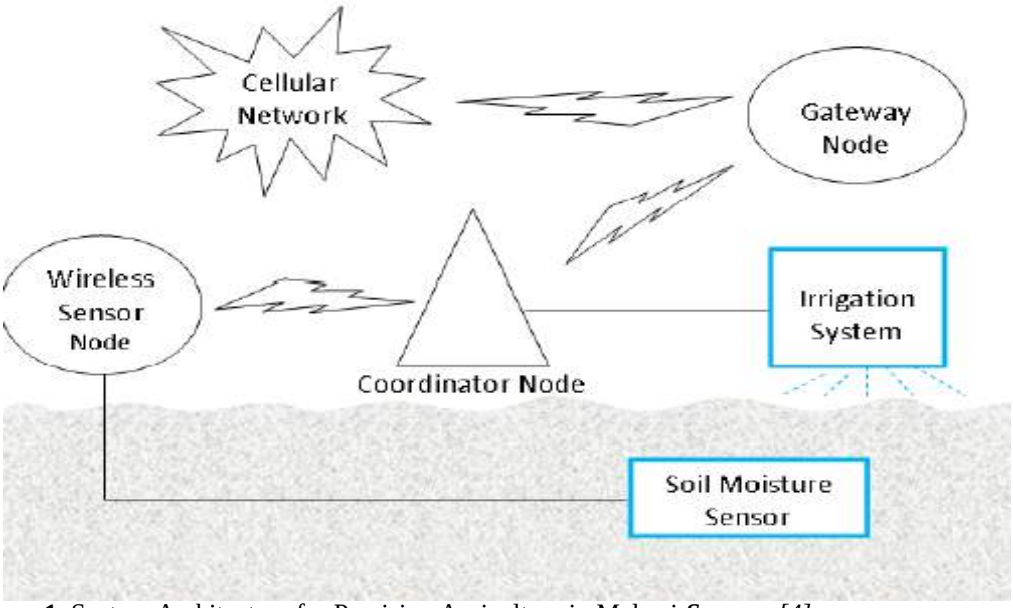


Figure 1. System Architecture for Precision Agriculture in Malawi ***Source:*** *[4]*

Shah [7] presented WSN system developed and deployed at IIT Bombay India for field monitoring of vineyard. The system used a combination of wired and wireless sensors to collect sensory data such as soil pH, soil moisture and soil temperature. Data collected by the sensors were wirelessly transferred in multi-hop manner to a sink node connected with embedded gateway. The sink node is equipped with ZigBee Module for intra communication and IP module for communication to the internet. Data is transferred through the internet to the remote server and to the users with internet enabled devices. Meanwhile users with mobile phones receive information later through cellular network after being processed in the remote server [7]. This type of architecture was trying to solve the problem identified by Mafuta [4] by avoiding conflicts between Zigbee and GPRS communication. However the approach is not effective for developing countries like Tanzania with unreliable Internet, because when the internet connection is down, farmers will fail to receive information to their mobile phones in real time

The system introduced in [7] used cluster architecture for data aggregation. In cluster based architecture all regular sensor nodes send data packet to a cluster head (local aggregator) which aggregates data packet from all the regular sensor nodes in its cluster and sends the aggregated data to a gateway. Cluster based fusion system is not robust to changes. If a node is disconnected from the communications network, it cannot be easily re-introduced and information synchronized. and Once the cluster head fails all nodes under it become useless [8].

As aforementioned, many of existing frameworks are not complete and practical; they just mention only few technologies like routing protocols, or intra-communication technologies within the farm. For those few developed frameworks, they do not lead to systems which can be easily used by users who are ICT illiterate and with little knowledge on horticulture sector. This is due to the fact these systems use complicated and expensive technologies.

The aim of this study was to develop the technical guideline/framework which can

be easily modified, understood and which can incorporate all necessary technologies to guide development of WSN for monitoring agriculture parameters for horticultural crops.

2. APPROACH

In order to solve the mentioned challenges, the author surveyed different WSN related technologies that have been used and presented by other researchers in order to make comparison of such technologies. Specific user and crops requirements were obtained from agriculture experts and farmers at Pwani Region. The information obtained was used to design a framework (shown in figure 2) which was later used to develop a WSN model to represent the designed framework. The components in the framework include devices and technologies such as sensors, sensor nodes, network topology, aggregation algorithms, communication protocols, coordinator, base station and user interface.

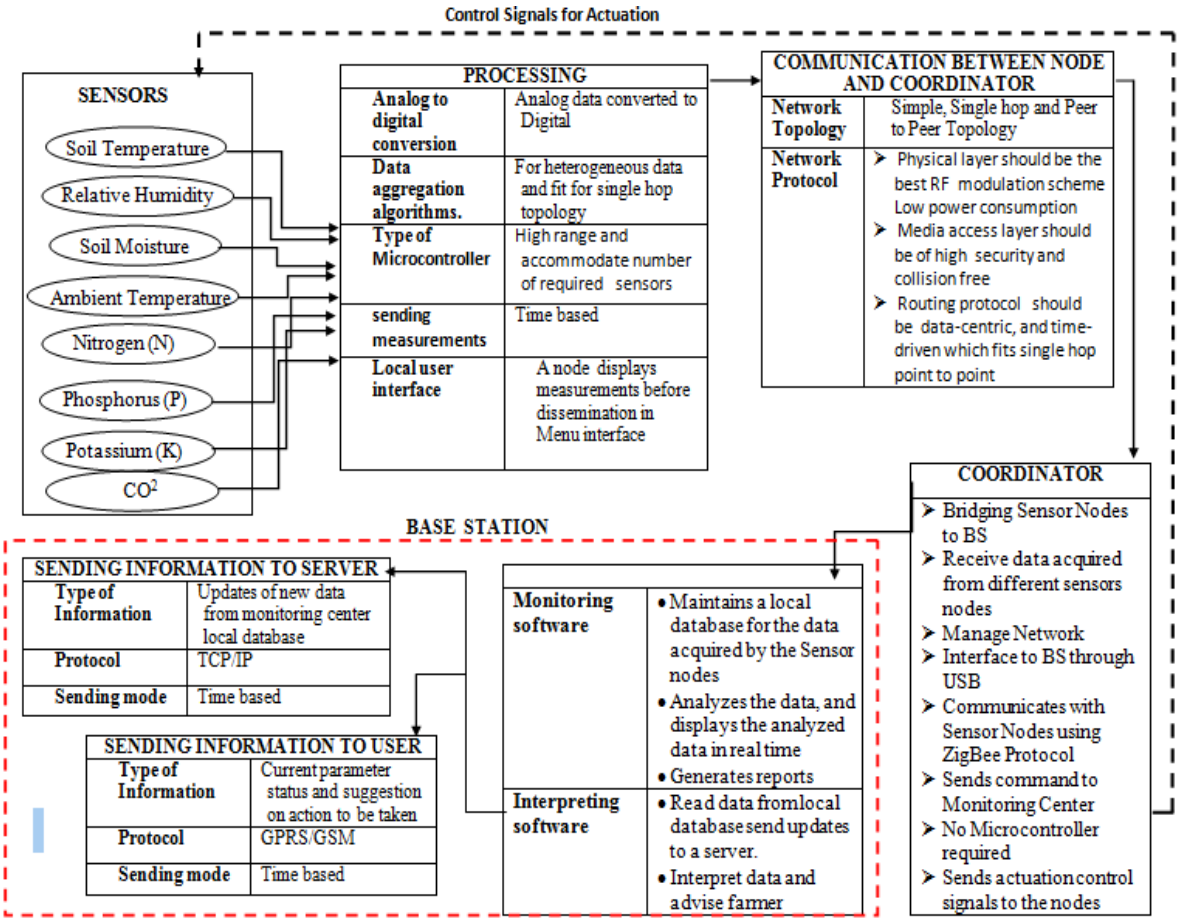


Figure 2. WSN Framework for Horticulture Crop

In this framework, communication within the network (intra communication) is done through ZigBee module in the Coordinator while analysis, interpretation, and transfer of information to farmers and remote server is done in the base station connected to the coordinator. The base station shown in doted lines in Figure 2 and detailed in Figure 3 is an extended block which comprises of software designed to analyze and deliver soil information and suggesting corrective measures for specific region of farm to the specific farmer in real time.

In this framework data collected from the soil is analysed in the local server in the field and is sent to the remote farmer directly to their mobile phones through

cellular network in real time. Internet service is used to send information to the remote server in which farmer can still receive information even if the internet connection is not available.

The conflict between intra-communication and cellular communication is avoided because Coordinator node is equipped with ZigBee only for intra communication while GPRS modules is installed in the base station which initiate all external communications through cellular network and Internet Connections.

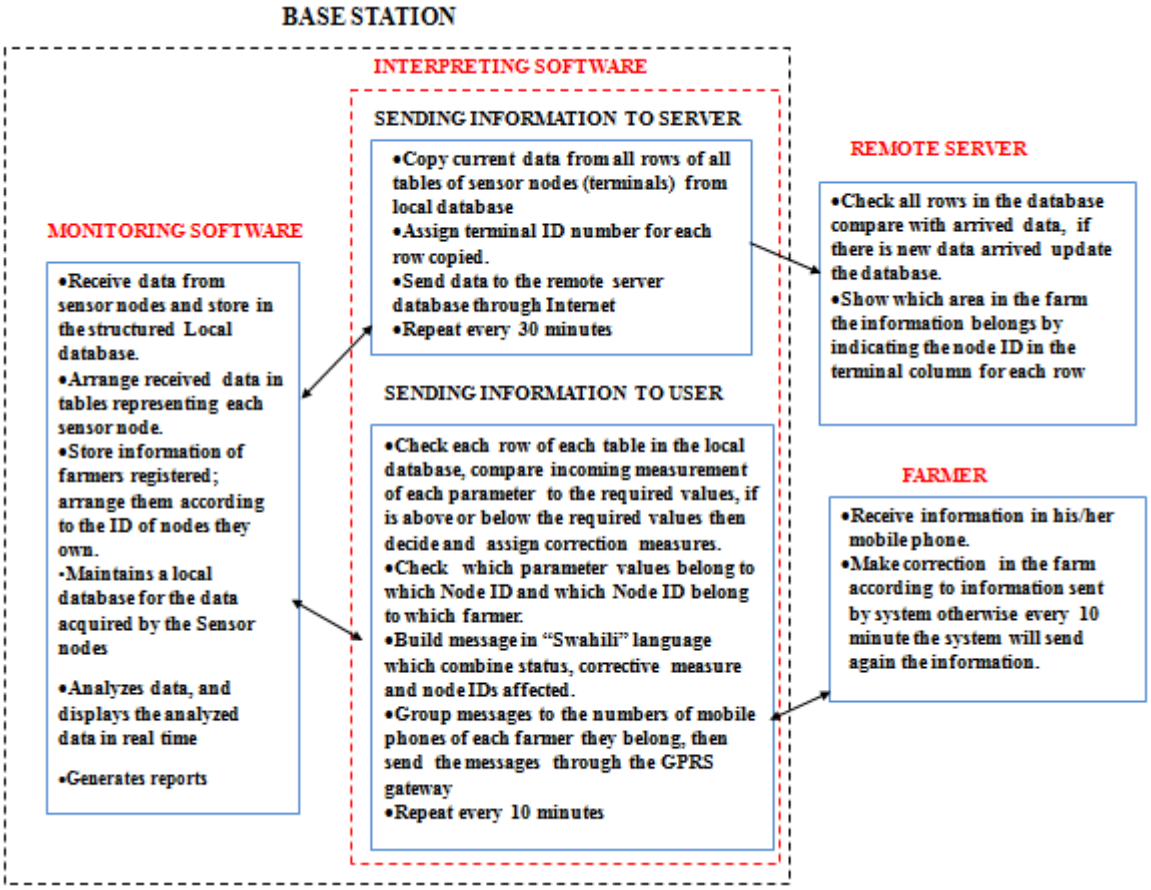


Figure 3. Algorithms in Base Station

The summary of the best technologies identified and deployed in this study is shown in the Table 1.

Table 1. WSN Technologies for Designed Framework

SN	COMPONENT	DESCRIPTION
1.	**Topology**	Simple Star Topology single hop Communication It includes the sensor nodes, a sink node and pc as a BS and the Server for data storage
2.	**Microcontroller**	ATmega8A
3.	**Radio transceiver**	IEEE 802.15.4 (ZigBee) compliant RF transceivers are used for intra-network communications between sensor nodes and sink node

4.	**Actuators**	The actuators will be placed somewhere in between the controlled sensor nodes.
5.	**Sending measurements to BS**	Time-based every 30min
6.	**Number of sensors**	The number of sensors depends on number of critical parameters to be monitored.
7.	**Number of nodes**	The number of nodes in the farm depended on the size of the farm, soil characteristics and the ground nature of the farm.
8.	**Communication within the LAN**	▪ In the physical layer IEEE 802.15.4 (ZigBee) with (DSSS, CSMA/CA) modulation scheme ▪ For media access layer, Fixed Assignment CDMA protocol will be used. ▪ The routing protocol is Sensor Protocols for Information via Negotiation Point to Point (SPIN-PP)
9.	**Communication to remote users**	• GPRS between BS and Mobile phones. • TCP/IP to remote server
10.	**Data Aggregation**	Decentralized based architecture is used as aggregation type where data aggregation occurs locally at each sensor node on the basis of local observations and the information obtained from sensors.
11.	**User Interface**	Menu driven Interface for mobile user and GUI for monitoring user interface
12.	**Monitoring Software**	Lab Windows/CVI is a software development environment chosen for this study. It maintains a database for the data acquired by the wireless sensor nodes. The program provides a facility to save the fundamental data, generates reports. This program can handle 256 nodes
13.	**Interpreting Software**	• The software interprets data stored in local database and integrates control measures information before sending the information to a farmer. It advises the farmer what to be done due to the current observed parameter values • It reads the local database and send updates to the remote server
14.	**Gateway Software**	• Ozeki NG used as SMS Gateway to sends information to user mobile phones. It allows sending of SMS messages toward the GSM/GPRS networks.

3. DEVELOPMENTS

By using the designed framework, the WSN model was developed. This model developed is for monitoring and control horticulture crop farm parameters as per user and crop requirements. In the development of the system to represent the framework, the Rapid Application Development method was used. This method is used in order to quickly produce high quality system. Active user involvement, and computerized development tools were used. These tools included Graphical User Interface (GUI) builders, Computer Aided Software

Engineering (CASE) tools, Database Management System (DBMS), fourth generation programming languages, code generators, and object-orient techniques.

The wireless sensor network for horticulture crops in this study consists of individual nodes that are able to interact with the environment by sensing physical parameters like soil temperature, moisture, humidity and nutrients (potassium, nitrogen and phosphorus). The nodes communicate by using wireless links and each node is able to communicate and collaborate with each other. In this design the nodes are equipped with a display screen so that the farmers without monitoring devices like mobile phones can directly see the measured parameter values of these nodes in their farms.

As shown in Figure 4, the wireless sensor network comprises of the standard components like sensor nodes used as (source), gateway/coordinator and control computer. The sensor nodes in this design send sensed data to the gateway using ZigBee while at the same time; these nodes display the measured data in the display screen. In this architecture the computer is used as a base station which analyses and sends information to the server through internet and mobile phone through cellular network.

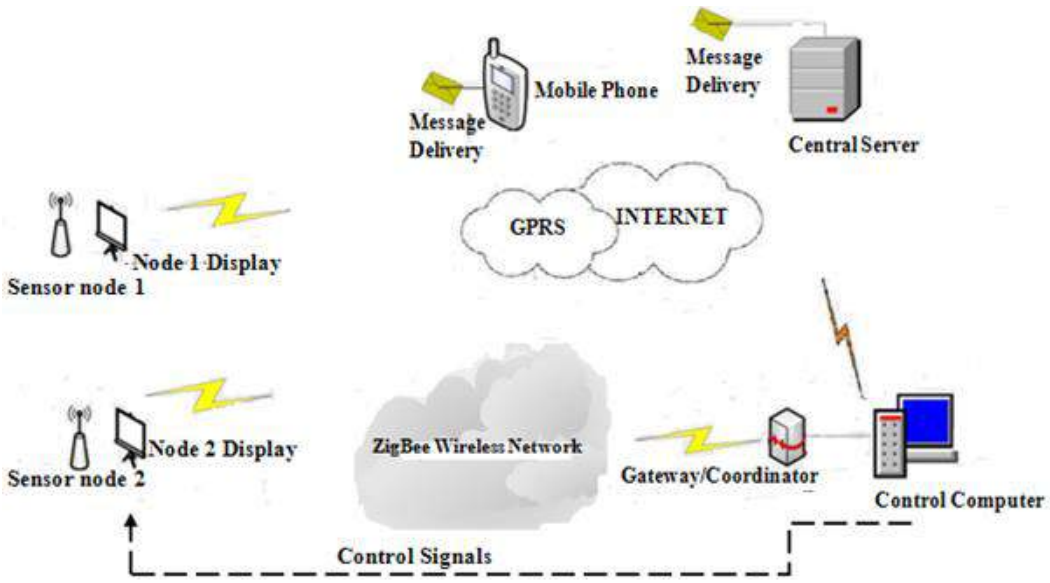


Figure 4. System Architecture of Proposed WSN system

Software that is used to analyze data stored in database in Base station was developed; this software interprets data and integrates control measures information before sending the information to a farmer. This makes it possible for the system to advise the farmer on what needs to be done due to the current observed parameter values. The software also sends and updates the data to the server every 3 hours.

Ozeki NG was used as SMS Gateway to send information to mobile phones; it is a software that allows connection from the system to a mobile network. It allows sending of SMS messages toward the GSM/GPRS network using GSM Modem [2]. The software was downloaded, installed and configured to send SMS to predefined users.

The communication technology used in this Gateway/coordinator node is the same with that in the sensor node (ZigBee (SZ05). This node receives power from the computer through the SPI bus, which ensures that the node is online all the time. Since many computers do not have COM ports, this system uses universal serial bus (USB) interface instead of serial port.

Under normal communication process the computer commands the coordinator/

gateway node via serial port. The physical address is used to collect commands, read the network status, locate the network address and send the data commands to the sensor nodes through ZigBee wireless communication modules[14]. After the sensor nodes collect the measurement or control command, they return the sensor data to coordinator/ through ZigBee wireless communication modules. The communication algorithms between sensor node and Coordinator node is shown in figure 5a and figure 5b respectively.

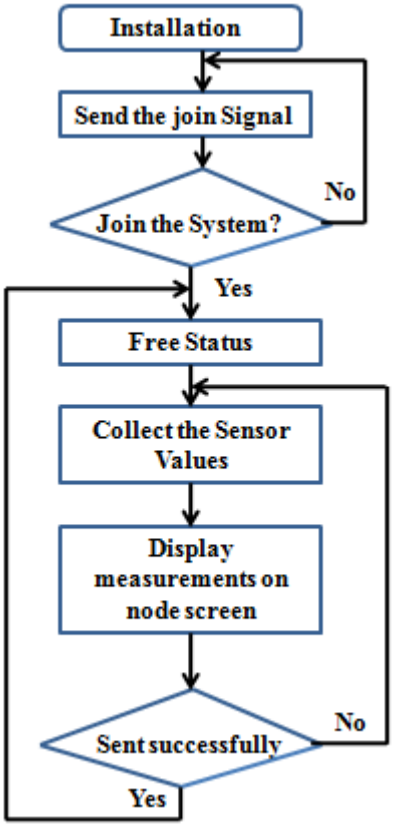


Figure 5a. Coordinator Process Flow Chart

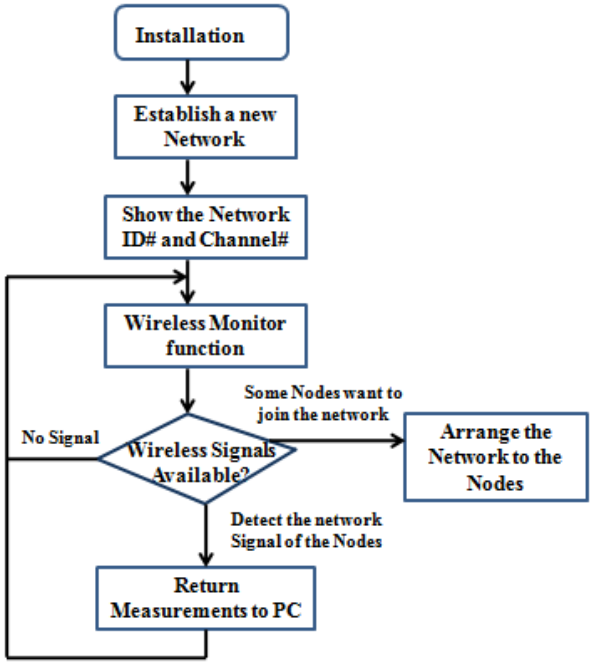


Figure 5b. Sensor Node Flow Chart

4. RESULTS

In this work, the Wireless Sensor Network was implemented and tested at Ruvu area Pwani Region Tanzania. The system measured environmental parameters of ambient temperature (Ta), soil temperature (Ts), relative humidity (RH), soil moisture (SM), and soil macronutrients of nitrogen (N), phosphorus (P) and potassium (K), sensors of light intensity (Li) and carbon dioxide (CO2). Each node processed the raw data and displayed the results on LCD screen continuously.

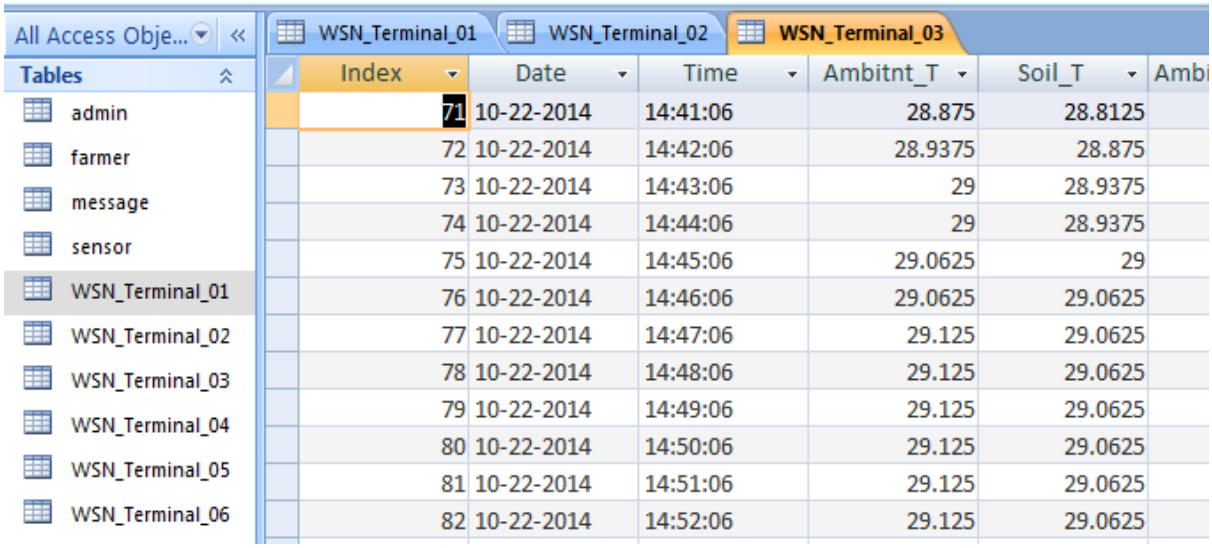
Figure 6 shows the sensor node displaying acquired data of ambient temperature (Ta), Relative humidity (RH), Soil temperature (Ts). All sensor nodes were capable of displaying the status of irrigation pumps and humidifies on the current state. The node was able to display other parameters in the next page through menu functions. The user was comfortable to use this user-friendly interface with keypads to calibrate different parameters of interest, set the optimum values range, and reset.

Figure 6. Sensor Node

The smart nodes sent measured data through the coordinator node to monitoring software in base station (Laptop) in real time. Through a comparison with standard parameters, control signals were issued to drive relays for driver motor for irrigation in case temperature measurement is high or soil moisture is below the standard threshold value. The actuators were placed in between the controlled sensor nodes, while the sink node (Coordinator) was placed indoors connected to the base station computer near to the field.

The software graphical user interface (GUI) platform enabled user to observe and modify the related values of soil environment easily. Its function is to manage the network and send signals to the control routine which switches on an actuator to correct the relevant condition such as water valve, aerator pump and heater pump.

The monitoring software receives data from sensor nodes and stores it automatically in the structured local database. Received data was arranged in tables representing each sensor node according to their IDs as shown in figure 7. Database also stored information for registered farmers, administrators and sent messages.



Index	Date	Time	Ambitnt_T	Soil_T	Ambi
71	10-22-2014	14:41:06	28.875	28.8125	
72	10-22-2014	14:42:06	28.9375	28.875	
73	10-22-2014	14:43:06	29	28.9375	
74	10-22-2014	14:44:06	29	28.9375	
75	10-22-2014	14:45:06	29.0625	29	
76	10-22-2014	14:46:06	29.0625	29.0625	
77	10-22-2014	14:47:06	29.125	29.0625	
78	10-22-2014	14:48:06	29.125	29.0625	
79	10-22-2014	14:49:06	29.125	29.0625	
80	10-22-2014	14:50:06	29.125	29.0625	
81	10-22-2014	14:51:06	29.125	29.0625	
82	10-22-2014	14:52:06	29.125	29.0625	

Figure 7. *Deployed data at Local Database*

The user-interface provides features for the complete real-time measurement, control, data processing, display result output, report printing and other functions. The main features include real-time measurement and display of environment values. The implementation of the environmental parameters are real-time displayed in a variety of experimental

curve display and can always switch observation and zoom, and save operation curve very conveniently. In figure 8, the system was configured to display three graphs for Nitrogen, Phosphorus and Potassium in real time.

The system evaluates the collected data and makes the decision based on the requirement of the horticulture crop. It managed to build a message in "Swahili" language which combined status, corrective measure and node IDs affected. The messages were sent to farmers through cellular network (see Appendix A). Ozek Short Message Services (SMS) gateway application was used to send messages every 10 minutes to the farmers in their language "Swahili".

The information collected from sensor nodes was sent to the remote server every 3 hours. The current data from all rows of all tables of sensor nodes (terminals) from local database was copied, assigned terminal ID number (for each row copied) and then sent to the remote server through Internet.

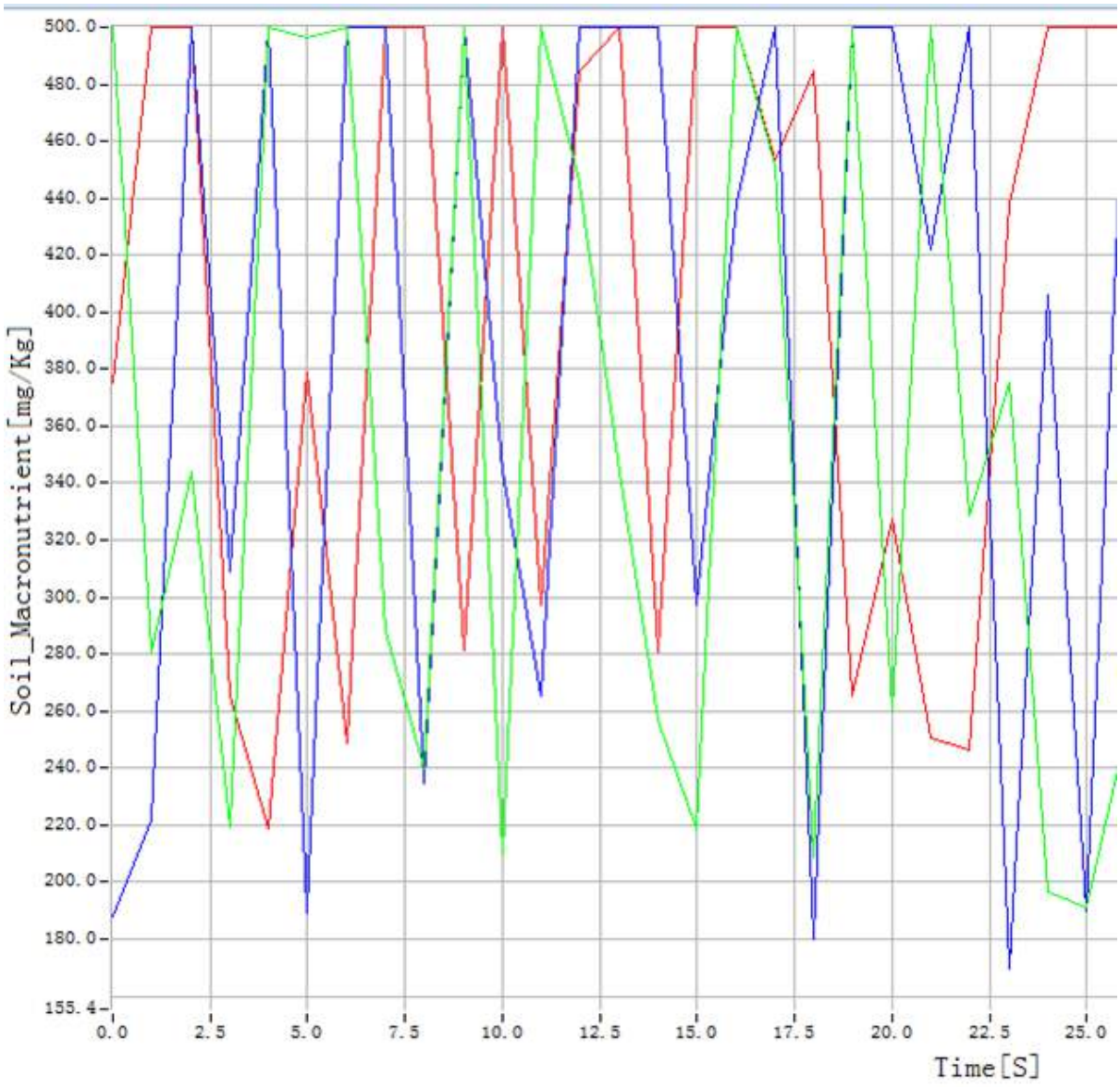


Figure 8. Nitrogen, Phosphorus and Potassium *Measurements*

5. BUSINESS BENEFITS

This study is very important for horticulture farmers and the growth of Tanzanian economy in general. The proposed model will help farmers produce crops which meet the quality standards and

requirements so as to effectively compete and access the regional and international markets. The outcome of this research will also benefit farmers of all education levels and even those who are ICT illiterate because it will result into a system which can be used by farmers to get necessary information in their own language through their normal mobile phone.

The horticulture farmers will be able to get information on the status of their farms and what needs to be done in a timely manner and take required measures in time. This will help them monitor their horticulture fields and prevent damages which will eventually increase crop production. In addition, the outcome of this study will help farmers monitor, control and increase the size of their farms for maximum agricultural productivity..

6. CONTRIBUTION

The contribution of this work is to develop a technical guideline/framework which includes all necessary technologies which are simple, reliable and easy to use. The framework can be easily understood and used to develop a WSN for monitoring agriculture parameters for horticulture crops. This framework as compared to others, it can be modified easily to fit development of WSN for other applications. Furthermore, in comparison with other frameworks where data is transmitted to remote server through the Internet for analysis and then sent to users who eventually encounter problems when the internet is down, this framework solves this problem by first analyzing data in the field. Data is analyzed in a computer at the field, sent to the farmer's mobile phones through a cellular network and to the remote server through the internet. This ensures reliability and timely delivery of information to the farmers while avoiding communication conflict within intra-communication.

7. CONCLUSIONS

In this work we presented a framework/guideline for deployment of easy to use wireless sensor networks (WSN) for the Horticultural farm soil monitoring and control. In this study both the requirements for horticultural crop and the farmers were identified. The technologies related to WSN frameworks were surveyed and the best technologies were identified. The technologies for the WSN framework namely sensors, smart nodes, sink nodes, network topologies, algorithms, communication protocols, and type of suitable user interfaces are discussed in detail. In this paper, we have demonstrated how the WSN model was developed from the framework/guideline and tested. We found that simple star topology and decentralized aggregation architecture when used together produce the system which is easy to use as a farmer can easily move a node from one region to another a farm or add new nodes in different areas in the farm without the need to reconfigure the system manually. The node placement metric presented was found to control the number of nodes to be deployed. Our experimental results show that using sink node equipped with ZigBee module, while GPRS modules incorporated separately in the base station, real time communication were maintained without conflicts.

Future development efforts should involve enhancing the WSN by integrating WSN of different application to a single server, and to make the system web-based so that users can access the data through web interface. Also there a need of linking the data to the cloud or smart grid so that they can be timely accessible statistical purposes.

8. REFERENCES

1. Abdullah, A. Identification of the Type of Agriculture suited for Application of Wireless Sensor Networks, Russian Journal of Agricultural and Social- Economic Sciences,Vol 12. pp. 19-36. (2012).

2. Munsuri, A., Arzuaga, J. and Zamalloa A. GPRS Technology and Application, uSysCom, Spain, pp 39-43, (2010).

3. Byerlee, D. de Janvry, A. and Sadoulet, E. "Agriculture for Development: Toward a New Paradigm", Annual Review of Resource Economics, Vol. 1: pp 15-35. (2009).

4. Mafuta, M. and Bagula Successful Deployment of A Wireless Sensor Network for Precision Agriculture in Malawi, Electrical Engineering Department, University of Malawi. vol. 47, no. 2, pp. 97-106. (2012),

5. Mashindano, O., Kayunze, K., da Corta, L. and Maro, F. Agricultural growth and poverty reduction in Tanzania 2000-2010: where has agriculture worked for the poor and what can we learn from this?, Chronic Poverty Research Centre, pp 2-39. (2011).

6. Mwakalukwa, E. Meilby, H. and Treue, T. Floristic Composition, Structure, and Species Associations of Dry Miombo Woodland in Tanzania. Department of Forest Biology, Faculty of Forestry and Nature Conservation, Sokoine University of Agriculture, vol 2, pp 1-15. (2014).

7. Shah. N and Das. I, Precision Irrigation: Sensor Network Based Irrigation, Problems, (2012).

8. Divya S, Sandeep V, and Kanika S. Network Topologies in Wireless Sensor Networks: Dept. of ECE, NITTTR, Chandigarh, UT, India Vol 4, pp 93-97. (2013).

9. Sherine, M., El-kader, A., Basma M. and Mohammad, E, Precision farming solution in Egypt using the wireless sensor network technology, Egyptian Informatics Journal Vol.14, pp 221–233, (2013),

10. Abdullah, A., Identification of the Type of Agriculture suited for Application of Wireless Sensor Networks, Russian Journal of Agricultural and Social- Economic Sciences, Vol.12. pp. 19-36, (2012).

11. Mampentzidou, I. Karapistoli, E. and Anastasios, A. Basic Guidelines for Deploying Wireless Sensor Networks in Agriculture. Deptment of Information `Systems University of Macedonia Thessaloniki: Greece, pp 978-4673, (2012).

12. Hans-Henrik, H., Steger-Jenseir, K. and Falster, P. Architectural Frameworks for Business Information System Analysis and Design, Department of Social Sciences, Wageningen University: The Netherlands, pp 414-416, (2008)

13. Gurwinder, K. and Rachit, G. Energy Efficient Topologies for Wireless Sensor Networks, International Journal of Distributed and Parallel System (UPDS) Vol.3, No.5, pp 179-192. (2012).

14. Harneet K, Sukesha S, A Comparative Study of Wireless Technologies: Zigbee, Bluetooth LE, Enocean, Wavenis, Insteon and UWB, Trends In Computing and Communication Engineering, UIET, PU Chandigarh India, pp 273-276, (2013),

International Journal of New Computer Architectures and their Applications (IJNCAA) 5(3): 119-126
The Society of Digital Information and Wireless Communications, 2015 ISSN 2220-9085 (Online); ISSN 2412-3587 (Print)

Opportunities for Employing IGBT in Photo-switch based on Silicon Avalanche LEDs

Kaikai Xu[*a], Siyang Liu[b], Jianming Zhao[a], Qi Yu[a], Weifeng Sun[b], and Guannpyng Li[c]
[a]State Key Laboratory of Electronic Thin Films and Integrated Devices,
University of Electronic Science and Technology of China, Chengdu, Sichuan, China 610054
[b]National ASIC System Engineering Research Center, Southeast University, Nanjing, China 210096
[c]California Institute for Telecommunications and Information Technology, Irvine, California 92697
[*]Email address: (kaikaix@uestc.edu.cn)

ABSTRACT

A comprehensive approach for the practical realization of silicon light-emitting devices (Si-LEDs) with emitting visible light in the 400 to 900 nm wavelength region is discussed. Prototype Si-LEDs are fabricated in the standard CMOS technology, using the same processing procedures with other components. Since fully integrated silicon photon-receivers with Si-LED on the same chip will largely improve the overall system performance, monolithic integration leads to lower cost and smaller size. Some structural details and performances of several practical two and three-terminal Si-LEDs are presented. In this paper, we report on further progress that has been made with regard to modeling of the physical processes in realizing an increase in the optical emission power, as well with regard to higher frequency modulation capability of such device. The theory of silicon optical modulation based on p-n junction in reverse bias is primarily discussed. Initial investigations indicate that the Si-LEDs have a very fast inherent modulation bandwidth capability, and the upper limit derived value for the expected maximum modulation of the device could be in the range of a few hundred GHz. According to the best of our knowledge, despite the low efficiency, the Si-LEDs show potential for on-chip electro-optical communication.

KEYWORDS

Silicon, optoelectronics, micro-optical devices, electro-optical modulation, PN junction

1 INTRODUCTION

Modern MOSFET technology has advanced continuously since its beginning in the 1950s. The complementary nature of p-type FETs and n-type FETs makes it possible to design low-power circuits called CMOS or complementary MOS circuits, as presented in Fig. 1. Because of the advantages of low power dissipation, short propagation delay, controlled rise and fall times, and noise immunity equal to 50% of the logic swing, the CMOS process is defined as the standard fabricating technology for semiconductor devices and electronic circuits in industry [1].

Silicon is well known as the core material of the electronic industry, but it has some difficulties with photonic application because silicon, which has an indirect band-gap of 1.12-eV, emits light only weakly by band-to-band transitions or by defect states to band transitions in the near infrared [2]. However, because of silicon's mature processing technology, low cost, CMOS compatibility and compactness, the development of silicon photonics is welcomed by many scientists [3].

Recently, a large variety of the attempts, such as dislocation loop light-emitting diode [4, 5], porous silicon [6], light emitting from rear locally diffused solar cell [7], silicon light-emitting diodes in silicon-on-insulator [8–11], silicon nano-particles in Si-in-SiN_x thin films [12], crystalline silicon LED [13, 14], and Si/SiO_2 super-lattices luminescence [15, 16], have been made to increase the efficiency of Si-based light emitters. These Si light sources have achieved relatively high efficiency [17], but all of these technologies are quite complicated and can not be easily integrated into the standard Si-CMOS process technology [18–21].

In contrast to the light emitters above, Si-diode LED can be easily realized using standard Si-

CMOS process technology without any additional process. Because of the full compatibility with the standard Si-CMOS process, the silicon light-emitting device (Si-LED) is capable of integrating with other silicon devices or circuits to realize monolithic integration in optoelectronics. It is widely known that most of the injected carriers in a forward biased silicon diode are recombined non-radiatively due to silicon's indirect band structure. Instead, a reverse-biased silicon-diode can emit visible light in the depletion region under avalanche breakdown [22]. Since the breakdown condition is generally regarded as being a solid state analog of a gas discharge plasma [23], Bremsstrahlung [24] (i.e., braking radiation) by hot electrons in the Coulomb field of charged impurities was previously treated as the major cause of the photon emission previously.

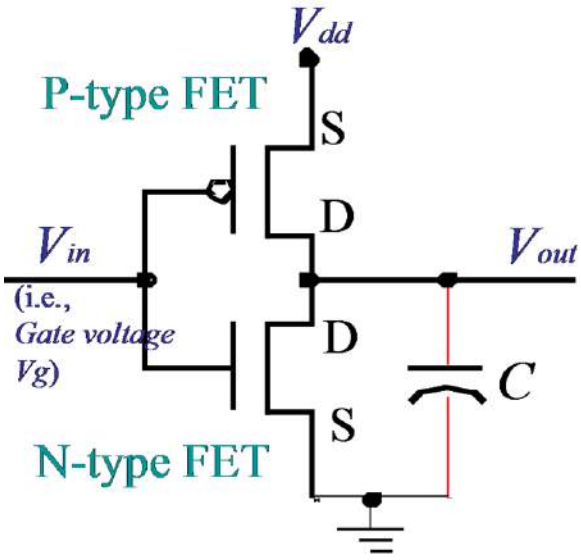


Figure 1. Schematic diagram of the CMOS inverter

A Si-PMOSFET device is fabricated to investigate the difference between silicon gate-controlled-diode LED and silicon-diode LED. In contrast to silicon-diode LED, the major advantage of silicon gate-controlled-diode LED is the existence of the insulated-gate terminal which can make the Si-PMOSFET device work as two identical parallel connected gate-controlled-diodes (i.e., the "P^+ Source/Drain to N-Substrate" junction with varying gate voltage). Generally speaking, in silicon-diode LED the increase in light intensity is achieved by increasing the electric field through increasing the reverse-bias voltage; whereas in silicon gate-controlled-diode LED the increase in light intensity is achieved by increasing the electric field through an increase in the gate voltage. In addition, it is noted that, in the case of silicon gate-controlled-diode LED, the reverse-bias of the "P^+ Source/Drain to N-Substrate" junction is fixed. It was reported that, at the same reverse current (i.e., the current flowing through the reverse-biased p-n junction), the emitted light intensity in gate-controlled-diode structure is much higher than that in the diode structure. In fact, simulated results in this paper indicate that the field in the gate-controlled-diode is one order of magnitude higher than that in the diode, and it is shown that the gate-terminal produce light intensity enhancement in reverse-biased silicon p-n junctions.

The paper is will attempt to give some flavor of the history, current status, and future prospects of the research field of silicon light-emitting devices. Section 2 will introduce the structure and fabrication of the Si-PMOSFET device. Section 3 will consider technologies for the modulation of light intensity in the silicon-diode LED case and in the silicon gate-controlled-diode LED case. Section 4 will describe applications to the insulator gate bipolar transistor (IGBT) in the power ICs. Section 5 will present the opto-coupler applications. Finally, Section 6 will make a conclusion.

2 DEVICE STRUCTURE AND CONFIGURATIONS

Standard 3-μm CMOS process with self-aligned technology is utilized for device fabrication. The device consists of MOS capacitor fabricated on a lightly doped <100> Si n-type substrate (525 ± 25 μm thick with a resistivity of 0.8–1.2 Ω × cm). Oxidation is performed at 1200℃ for 5 hrs in a N_2 ambient and 2 hrs in an O_2 ambient. The gate dielectric of the device consists of 450 Å thick thermally grown silicon oxide.

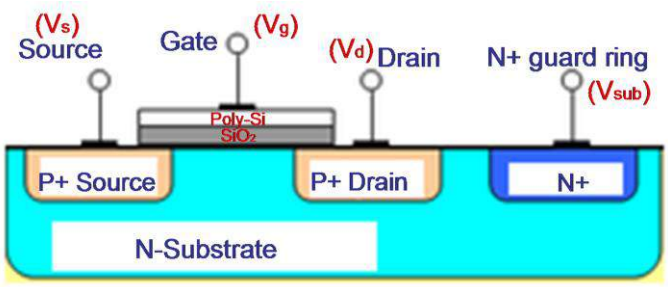


Figure 2. Schematic cross section of the light-emitting MOS device: gate width is 6 μm, gate length is 175.5 μm, N-Substrate surface doping concentration N_d ~10^{16}cm^{-3}, P^+ Source/Drain diffusion region doping concentration N_a ~10^{19}cm^{-3}, P^+ Source/Drain junction depth X_j ~0.5 μm. V_s is the source voltage, V_g is the gate voltage, V_d is the drain voltage, and V_{sub} is the substrate voltage

The PMOSFET sample was implanted with boron at 60 keV to a total fluency of 10^{13} ions/cm^2. The implantation energy (80 keV) has been chosen to have the projected range of the ion distribution roughly the same as in the N^+ guard ring which is for the ohmic-contact between substrate and electrode. After implantation, annealing at 850℃ for 30 minutes was performed in a N_2 flux to eliminate implantation defects and obtain the growth and agglomeration of Si residuals in thin films. The structure of the PMOSFET device is completed by the CVD deposition of a 4000Å thick n^+ poly-silicon layer. Finally, a metal ring consisting of an Al-Si-Cu layer (3 μm thick) completes the device structure, thus allowing device bonding to a standard TO_3 package. Fig. 2 shown a schematic cross section of the device.

3 RESULTS AND ANALYSIS

The key elements of silicon photonic systems are optical source capable of fast modulation, suitable transmission media, and fast optical detector or optically coupled power semiconductor devices.

The switching characteristics as associated with P^+N gated MOSFET silicon LED are reviewed. By employing the insulated-gate terminal that allows adjustment of "P^+ Source/Drain to N-Substrate" junction breakdown voltage (*BV*), it is demonstrated that the electro-optical modulation in the Si-PMOSFET device can be achieved using gate-controlled diodes. The PMOSFET device can operate as a Si-diode LED or a Si gate-controlled diode LED. The main features of switching transitions of Si-diode LED and Si gate-controlled diode LED are characterized, and a model developed to explain the modulation speed is then reviewed. The upper limit derived value for the expected maximum modulation of the device can be in the range of a few hundred GHz. Despite of its low efficiency, the Si-PMOSFET light-emitting device (Si-PMOSFET LED) will be a potentially key component for silicon photonic integrated circuits for future computing I/O applications.

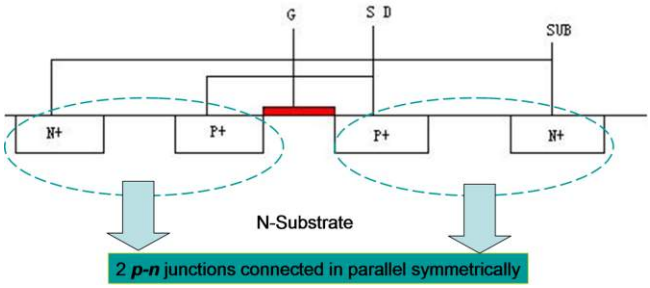


Figure 3. Schematic of the Si-PMOSFET device used in this study

This section especially focuses on the comparison in light intensity modulation between the Si-diode LED mode and the Si gate-controlled diode LED mode. Fig. 3 shows a schematic diagram of the Si-PMOSFET device.

3.1 Si-diode LED

As shown in Fig. 3, without the function of gate voltage V_g, the Si-PMOSFET device will act as two p-n junction diodes in parallel and the reverse bias of the "P^+ Source/Drain to N-Substrate" junction will be equal to the substrate voltage V_{sub} if source and drain are both grounded. To realize electro-optical modulation, the small-signal is electrically input from the two terminals of the p-n junction using the V_{sub} as the DC voltage carrier, and then the small-signal as an optical signal is output from the light emitting region of the device.

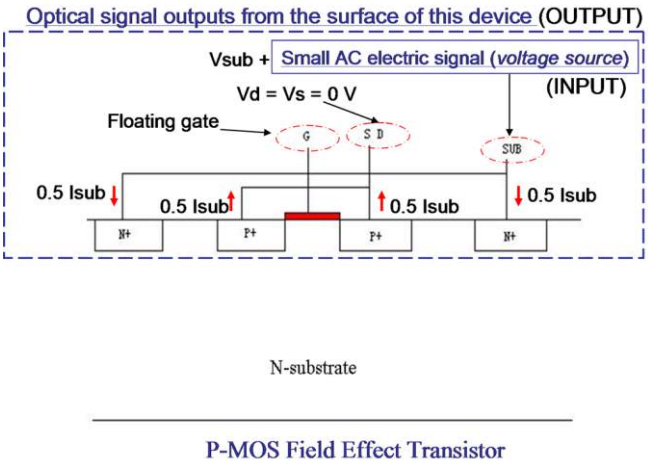


Figure 4. V_{sub}, a dc source, is the reverse bias across the "P^+ S/D to N-Sub" junction; V_{sub} is varied to realize optical modulation in the two-terminal device

From the schematics shown in Fig. 4, it can be observed that the prerequisite for realizing the electric-optical modulation is that the p-n junction is in avalanche breakdown. At the same time, because of the unwelcome capacitive load to devices and circuits, the speed of modulation which is closely related to the depletion-layer capacitance is described by

$$C_{dep} = A\frac{\varepsilon_s}{W_{dep}} \qquad (1)$$

where A is the cross-section area, ε_s is the permittivity of silicon, and the depletion width is given by

$$W_{dep} = \sqrt{\frac{2\varepsilon_s (V_{bi} + V_{sub})}{qN_d}} \qquad (2)$$

where V_{bi} is the built-in potential of the "P^+ Source/Drain to N-Substrate" junction, V_{sub} is the reverse bias of this junction, N_d is the background concentration (i.e., doping concentration of the N-Substrate), and q denotes the elementary charge. Substituting Eq. (1) into Eq. (2), it becomes

$$\frac{1}{C_{dep}^2} = \frac{W_{dep}^2}{A^2\varepsilon_s^2} = \frac{2(V_{bi} + V_{sub})}{qN_d\varepsilon_s A^2} \qquad (3)$$

which implies the capacitance is inversely proportional to the reverse-biased voltage. On the other hand, the 3-dB frequency is

$$f \sim \frac{1}{2\pi RC_{dep}} \qquad (4)$$

Substituting Eq. (3) into Eq. (4), the modulation speed of the PMOSFET device working as two p-n junction diodes will be obtained as

$$f \sim \sqrt{V_{sub}} \qquad (5)$$

It has been shown by 2-D device simulation that the p-n junction based silicon modulator has a fast intrinsic response time of ~7 ps [25]. Since the total capacitance of the reverse-biased silicon diode is less than 3 pF, it is found that silicon diode has an intrinsic frequency capability of GHz in theory by taking the dynamic series resistance into account [26].

However, the current Si-diode is by no means optimized for performance, and the device can be improved by optimizing the doping profile and p-n junction placement to increase phase efficiency. By applying the gate voltage, a field induced junction optimizes the silicon p-n junction, thus making the device a Si gate-controlled diode LED.

3.2 Si gate-controlled diode LED

As shown in Fig. 3, by applying a gate voltage V_g, the Si-PMOSFET device will act as two gate-controlled diodes in parallel and the reverse bias of the "P^+ Source/Drain to N-Substrate" junction, which is equal to the substrate voltage V_{sub} if source and drain are both grounded, has a certain value in the mode of gate-controlled-diode.

Due to the variation in gate voltage V_g, the breakdown voltage BV of the "P^+ Source/Drain to N-Substrate" junction will be changed, thus resulting in the modulation of breakdown current and its corresponding light intensity. Since the reverse-bias V_{sub} is fixed to function as a DC voltage source, the small electric signal will be input from the gate terminal.

From the schematics shown in Fig. 5, the DC section of gate voltage V_g will be the carrier for the electric signal input while the optical signal output is constituted by photons emitted from the light emitting active region in the Si-PMOSFET device. In contrast to the mode of Si-diode LED, the modulation speed of the three-terminal Si gate-controlled-diode LED is determined by

$$C_{ox} = A_G \frac{\varepsilon_{ox}}{t_{ox}} \quad (6)$$

where C_{ox} denotes the capacitance of the metal-oxide-semiconductor (MIS) capacitor of the MOSFET, A_G is the area of gate, ε_{ox} is the permittivity of SiO_2, and t_{ox} is the thickness of the SiO_2 layer. Accordingly, the speed of modulation in this case is expressed as

$$f \sim \frac{1}{2\pi R C_{ox}} \quad (7)$$

In addition, both avalanche and Zener breakdowns are inherently fast process, and operation of silicon-based LED at a frequency of 10 GHz was reported [27]. A clear and better understanding of the modulation phenomena in the three-terminal Si gate-controlled diode LED operating in the depletion mode is presented in Ref. 28.

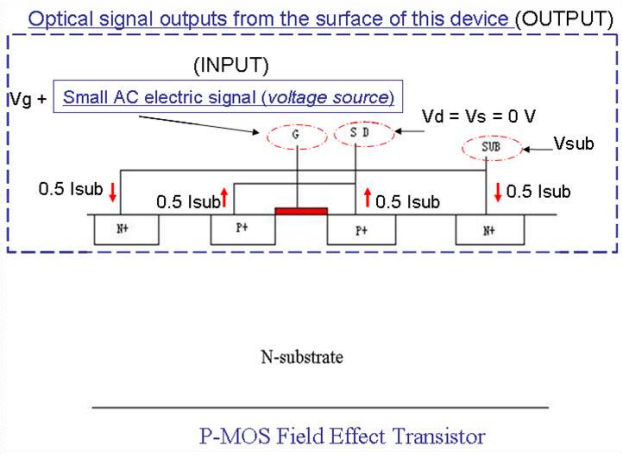


Figure 5. V_g , a dc source, is the gate voltage. V_{sub} , a dc source, is the reverse bias across the "P^+ S/D to N-Sub" junction, and V_g is varied to realize optical modulation in the three-terminal device

In order to bridge the theoretical formulas above and experimental data, the dynamic behavior of the Si-PMOSFET device should be further tested in future. Especially, in Si-diode LED we increase the avalanching current to increase the light intensity, whereas in Si gate-controlled diode LED an additional field is applied to increase the light intensity. In other words, the light intensity modulation in Si-diode LED requires direct modulation of the avalanching reverse current I_{sub}, whereas in Si gate-controlled diode LED the gate voltage V_g is varied to realize modulation via changing the electric field distribution. Overall, Si-diode LED is a conventional current driving device based on avalanche breakdown, but Si gate-controlled diode LED can be defined as a field-emission device in which both avalanche and tunneling processes occur together [29].

4 POTENTIAL APPLICATIONS AS OPTICAL SWITCH FOR POWER IGBTS

The power device is known as one of the most important components for driving HDTV Plasma Display Panel (PDP). It is noted that the PDP driver IC always requires that the integrated power devices have high off-state breakdown voltages (BV) and large current capability in order to make the system operate.

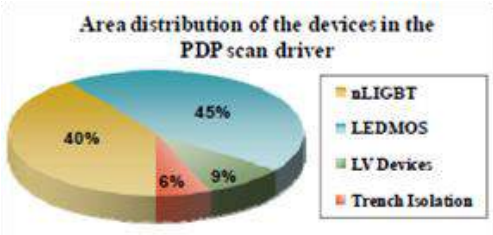


Fig. 6. Area distribution of the devices for the PDP scan driver IC

According to the area distribution of the devices in the PDP scan driver IC shown in Fig. 6, it is seen that the power devices, including the n-type lateral insulator gate bipolar transistor (nLIGBT) and the lateral extended drain MOS (LEDMOS), take up 85% area of the IC.

200 V nLIGBT, nLEDMOS and pLEDMOS devices are fabricated using the SOI technology [30]. The measured electrical results, observing larger current drivability and higher off-state BV in the three devices, are shown in Fig. 7.

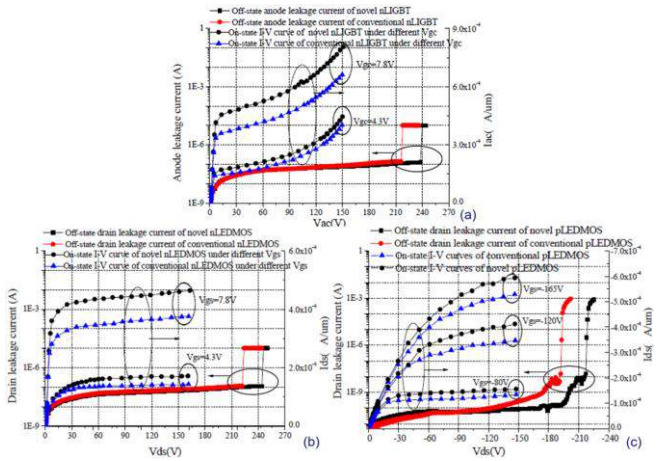


Figure 7. The experimental off-state and on-state I-V curves for: (a) the two nLIGBT devices with 3×1012cm-2 N-drift dose (W = 80 μm); (b) the two nLEDMOS devices (W = 80 μm); (c) the two pLEDMOS devices (W = 90 μm);

An opto-coupler is a device consisting of a LED and a photodiode coupled to input and output amplifiers that transforms an electric signal into an optical one in order to transfer data that is galvanically isolated. Several advances have been made in opto-coupler technology that improve performance and current input. The desirable features are low manufacturing cost, compact packaging and resistance to rugged. The potential of high modulation speed satisfies the data transmission speed advantages. Since the opto-coupler requires CMOS chip as a receiver, the Si-LED is the primary choice as the optical transmitter.

5 OPTO-COUPLER APPLICATIONS

Optical coupling is increasingly used in systems where complex communications must occur across a galvanic isolated boundary. Specially, in factory automation there exists a need for various pieces of equipment to communicate with each other and with a central computer via a bus. Because many of these pieces of equipment involve power devices, disruptive transient electrical signals are generated. To prevent these signals from corrupting the communication bus, optical isolation is used.

Industrial control applications also require opto-coupler. An example is the motor controller which includes sophisticated DSP based functions. Optical isolation is used by the motor controller to control the electrical power to the motor and to monitor the motor's response both electrically and mechanically. In addition, the so-called Internet of Thing (IoT) requires the control of many actuators using internet communication. There are cases were the internet controller will need to be isolated from the actuator because of severe noise generated in the actuator such as inductive spikes. Typically, today's opto-coupler manufactures place only a single GaAs LED chip on the sending side of the opto-coupler. It is left up to the user to provide an interface to the electrical requirements of the LED. This typically entails using discrete components to convert logic level signals or other types of signals to a signal suitable for driving the LED. With a silicon based LED, however, interface circuits can be built on the same chip as the LED thus eliminating the need for external, discrete components. Thus, the ability to place circuits on the LED side reduces component count.

Integration of the light emitting function with silicon allows two distinct advantages over existing technology. One is the integration of a complex circuit function on the LED side of an opto-coupler. GaAs LED technology does not support circuit functions. The other is the potential to make multiple bi-directional transmit and receive channels using only two pieces of silicon in a single package. Using SOI offers the potential to integrate both the receive and the transmit functions onto a single chip. The resulting higher level of integration not only lowers cost, but simplifies manufacturing for both opto-coupler manufacturer and the customer. An added caveat is that many opto-coupler do not require high bandwidths with 1 MHz being adequate.

6 CONCLUSIONS

Research on efficient light emission from silicon devices that are compatible with standard CMOS fabrication technology have been investigated. The Si-LED, as a direct electro-optic modulator, will be used for massively parallel optical interconnects, with potential for on-chip and chip-

to-chip optical links. Anticipated applications include electro-optic isolator for power ICs [31].

ACKNOWLEDGMENT

This work is sponsored in part by the 1000-Talents Program of Sichuan Province, China, Scientific Research Foundation for the Returned Overseas Chinese Scholars. Dr. Kaikai Xu, would like to thank Mr. Eugene Worley of Qualcomm for some useful discussions. Encouragement from Prof. Bin Wu of the Ryerson University (Canada) is greatly appreciated. Owing to space restriction, it has only been possible to reference a few example from the large body of excellent published work consulted in preparing this paper. The authors would like to thank their colleagues in academe and industry for their many contributions to advancing the field of opto-coupler whether their work has been specifically cited here or not.

REFERENCES

1. CMOS, “the Ideal Logic Family,” Fairchild Semiconductor Application Note, 77, (1983)
2. B. Jalali, “Can silicon change photonics?,” Phys. Stat. Sol. (a), vol. 205, no. 2, 213-224, (2008)
3. L. Khriachtchev, “Silicon Nanophotonics: Basic Principles, Present Status and Perspectives,” Pan Stanford Publishing, (2009)
4. W. Ng, M. Lourenco, R. Gwilliam, S. Ledain, G. Shao, and K. Homewood, “An efficient room-temperature silicon-based light-emitting diode,” Nature, vol. 410, no. 6825, pp. 192-194, (2001)
5. M. Lourenco, M. Milosavljević, R. Gwilliam, K. Homewood, and G. Shao, “On the role of dislocation loops in silicon light emitting diodes,” Appl. Phys. Lett., vol. 87, no. 20, 201105, (2005)
6. Y. Kanemitsu, K. Suzuki, H. Uto, Y. Masumoto, T. Matsumoto, S. Kyushin, K. Higuchi, and H. Matsumoto, “Visible photoluminescence of silicon-based nanostructures: porous silicon and small silicon-based cluster,” Appl. Phys. Lett., vol. 61, no. 20, pp. 2446-2448, (1992)
7. M. Green, J. Zhao, A. Wang, P. Reece, and M. Gal, “Efficient silicon light-emitting diodes,” Nature, vol. 412, pp. 805-808, (2001)
8. T. Hoang, P. LeMinh, J. Holleman, and J. Schmitz, “Strong efficiency improvement of SOI-LEDs through carrier confinement,” IEEE Electron Dev. Lett., vol. 28, no. 5, pp. 383-385, (2007)
9. J. Zhao, G. Zhang, T. Trupke, A. Wang, F. Hudert, and M. Green, “Near-band edge light emission from silicon semiconductor on insulator diodes,” Appl. Phys. Lett., vol. 85, no. 14, pp. 2830-2832, (2004)
10. S. Selvaraja, W. Bogaerts, P. Dumon, D. Thourhout, and R. Baets, “Subnanometer linewidth uniformity in silicon nanophotonic waveguide devices using CMOS fabrication technology,” IEEE J. Sel. Topics Quantum Electron., vol. 16, no. 1, pp. 316- 324, (2010)
11. S. Pillai, K. Catchpole, T. Trupke, G. Zhang, J. Zhao, and M. Green, “Enhanced emission from Si-based light-emitting diodes using surface plasmons,” Appl. Phys. Lett., vol. 88, no. 16, 161102, (2006)
12. Y. Wang, Y. Wang, L. Cao, and Z. Cao, “High-efficiency visible photoluminescence from amorphous silicon nanoparticle embedded in silicon nitride,” Appl. Phys. Lett., vol. 83, no. 17, pp. 3474-3476, (2003)
13. J. Zhao, A. Wang, T. Trupke, and M. Green, “High efficiency bulk crystalline silicon light emitting diodes,” Mat. Res. Soc. Symp. Proc., vol. 744, M4.7.1-6, (2003)
14. L. Ding, M. Yu, X. Tu, G. Lo, S. Tripathy, and T. Chen, “Laterally-current-injected light-emitting diodes based on nanocrystalline-Si/SiO_2 superlattice,” Opt. Express, vol. 19, no. 3, pp. 2729-2738, (2011)
15. H. Song, and X. Bao, “Visible photoluminescence from silicon-ion-implanted SiO_2 film and its multiple mechanisms,” Phys. Rev. B, vol. 55, no. 11, pp. 6988-6993, (1997)
16. S. Novikov, J. Sinkkonen, O. Kilpelä, and S. Gastev, “Visible luminescence from Si/SiO_2 superlattices,” J. Vac. Sci. Technol. B, vol. 15, no. 4, pp. 1471-1473, (1997)
17. H. Finkelstein, M. Gross, Y. Lo, and S. Esener, “Analysis of hot-carrier luminescence for infrared single-photon upconversion and readout,” IEEE J. Sel. Topics Quantum Electron., vol. 13, no. 4, pp. 959- 968, (2007)
18. T. Koch and U. Koren, “Semiconductor photonic integrated circuits,” IEEE J. Quantum Electron., vol. 27, no. 3, pp. 641-653, (1991)
19. A. Krishnamoorthy and K. Goossen, “Optoelectronic-VLSI: Photonics Integrated with VLSI circuits,” IEEE J. Sel. Topics Quantum Electron., vol. 4, no. 6, pp. 899-912, (1998)
20. R. Nagarajan, C. Joyner, R. Schneider et al., “Large-scale photonic integrated circuits,” IEEE J. Sel. Topics Quantum Electron., vol. 11, no. 1, pp. 50-65, (2005)
21. A. Fang, M. Sysak, B. Koch, R. Jones, E. Lively, Y. Kuo, D. Liang, O. Raday, and J. Bowers, “Single-wavelength silicon evanescent laser,” IEEE J. Sel. Topics Quantum Electron., vol. 15, no. 3, pp. 535- 544, (2009)
22. J. Bude, N. Sano, A. Yoshii, “Hot-carrier luminescence in Si,” Phys. Rev., vol. 45, no. 11, pp. 5848-5856, (1992)
23. R. Zia, J. Schuller, A. Chandran, M. Brongersma, “Plasmonics: the next chip-scale technology,” Materials Today, vol. 9, no. 7-8, pp. 20-27, (2006)
24. A. Lacaita, F. Zappa, S. Bigliardi, and M. Manfredi, “On the bremsstrahlung origin of hot-carrier-induced

photons in silicon devices," IEEE Trans. Electron. Dev., vol. 40, no. 3, pp. 577-582, (1993)
25. F. Gardes, G. Reed, N. Emerson, and C. Png, "A sub-micron depletion-type photonic modulator in Silicon on Insulator," Opt. Express, vol. 13, no. 22, pp. 8845-8853, (2006)
26. A. Chatterjee, B. Bhuva, and R. Schrimpf, "High-speed light modulation in avalanche breakdown mode for Si diodes," IEEE Electron Dev. Lett., vol. 25, no. 9, pp. 628-630, (2004)
27. L. Snyman, M. du Plessis, E. Seevinck, H. Aharoni, "An efficient low voltage, high frequency silicon CMOS light emitting device and electro-optical interface," IEEE Electron Device Letters, vol. 20, no. 12, pp. 614-617, (1999)
28. K. Xu, S. Liu, J. Zhao, W. Sun, and G. Li, "Analysis of simulation of multi-terminal electro-optic modulator based on p-n junction in reverse bias," Opt. Eng., vol. 54, no. 5, 057104, (2015)
29. M. du Plessis, H. Aharoni and L. Snyman, "Two- and multi-terminal silicon light emitting devices in standard CMOS/BiCMOS IC technology," Physica Status Solidi (a), vol. 201, no. 10, pp. 2225-2233, (2004)
30. S. Liu, W. Sun, T. Huang, and C. Zhang, "Novel 200V power device with large current capability and high reliability by inverted HV-well SOI technology," 25th International Symposium on Power Semiconductor Devices and ICs (ISPSD), pp. 115-118, (2013)
31. B. van Drieënhuizen and R. Wolffenbuttel, "Optocoupler based on the avalanche light emission in silicon," Sens. Actuators A, Phys., vol. 31, iss. 1-3, pp. 229-240, (1992)

International Journal of New Computer Architectures and their Applications (IJNCAA) 5(3): 127-140
The Society of Digital Information and Wireless Communications, 2015 ISSN 2220-9085 (Online); ISSN 2412-3587 (Print)

Security of Composite Electronic Services

Jarosław Wilk
Military University of Technology / Comparex Poland Sp. z o.o.
Warsaw, Poland
jaroslaw.wilk@wat.edu.pl / jaroslaw.wilk@comparex.pl

ABSTRACT

This paper presents a new approach to the security of composite electronic services. They are defined as a part of integration platforms in the service oriented architecture. New security requirements are presented together with a short explanation why currently available models are not sufficient. Defined security model combines access and execution management with use of the lattice theory to allow automatic integration of electronic services originating from different information systems. Proposed solution is described using mathematical modeling with a short case study example. The direction of future work is indicated together with possible places of use for the author's model.

KEYWORDS

information security, integration platforms, electronic services, composite services, service oriented architecture, SOA, lattice theory

1 INTRODUCTION

Rapid development of digital information systems in the last decades enforces the use of data from many different sources. It becomes necessary to integrate independent systems, which historically were created as separate elements. Integration platforms are considered to be a solution to integrate systems which provide electronic services. However, this new solution has brought new challenges. One of them is a difficulty in providing information security for complex electronic services consisting of many atomic services (originating from different systems).

Currently available information security management models for integration platforms "are still not fully satisfactory (despite many recommendations and best practice guides)" [1]. There is a need for a new model that will be suitable for the "service oriented architecture (SOA)" [2] and will allow dynamic management of security levels for integrated services.

The article discusses the author's information security model for service oriented systems. It addresses security management for systems like electronic services portals, cloud environments and integration systems based on web services.

The paper is organized as follows: first it defines the service oriented architecture, a composite electronic service and an integration platform (section II). Section III focuses on the importance of security in integration process, pointing differences in security management between SOA based systems and data direct access systems. Section IV presents a mathematical model of defined systems with a proposition of the new security model based on the lattice theory [3]. Section V presents an example case study indicating the usefulness of the proposed solution. In Section VI, the conclusions of the paper are presented.

2 INTEGRATION PLATFORMS

2.1 Service Oriented Architecture

Accurate definition of electronic services and the service oriented architecture is crucial in understanding role of integration platforms in modern information systems.

According to European Union Council Implementing Regulation – "electronically supplied services' shall include services which are delivered over the Internet or an electronic network and the nature of which renders their supply essentially automated and involving minimal human

intervention, and impossible to ensure in the absence of information technology" [4].

Information systems which are designed these days are mostly focused on electronic services. This approach makes them user friendly, as typical end user is expecting specific services from IT systems. A service consumer is not interested in understanding what elements (servers, operating systems, databases, applications etc.) are involved in the whole process. He wants to achieve intended goal through the implementation of a service which can be supplied electronically.

The service oriented architecture is a reference model for IT systems based on e-services. "SOA is a paradigm for organizing and utilizing distributed capabilities that may be under the control of different ownership domains. It provides a uniform means to offer, discover, interact with and use capabilities to produce desired effects consistent with measurable preconditions and expectations." [2]

Summarizing Service Oriented Architecture enforces encapsulation and modularity by providing composite services to service consumers. Those composite services are created from atomic services originating from service components and operational systems.

2.2 Composite Electronic Services

Number of e-services is rapidly growing and they are covering different areas of the economic and public life. European Union Council Implementing Regulation [4] is pointing out some of them:

- e-government (public administration services using modern integrated digital techniques),
- e-banking services (access to bank accounts / trading systems using a computer or other electronic device via the Internet).
- e-insurance (services available on the Internet including: purchasing of insurance, contacting the agent, declaring a damage, interactive consulting, calculating the contribution),
- e-education (educational services available through electronic devices including: language courses, vocational and professional courses for employees, studies, private lessons without a teacher),
- e-culture (distribution, presentation, preservation and (re) use of cultural works using digital techniques).

Consumers are expecting composite services to perform their tasks with a minimum effort. That is why today's e-services are very complex and originate from many different systems. For example, if a citizen wants to register a company he should use only one composite e-service. However, this electronic public service has to use many information systems from many public administration units like Ministry of Economy, Social Insurance Institution, National Court Register, Ministry of Finance and many more depending on the national law.

Those independent systems can communicate in two different models: bilateral and multilateral.

2.3 Bilateral and Multilateral Interoperability Models

A bilateral interoperability model is based on two components communicating directly with each other. It is not allowed to mediate the communication through the third element. This type of solution used for information systems is often called "point to point". In case of a larger number of elements in order to ensure full communication, it is necessary to create and maintain $n \cdot (n-1) / 2$ connections (where n is the number of elements). Each connection is independent and can only support communication between the two systems. In case of one connection line failure, affected elements will lose contact with each other even though they are still physically connected by another network element (another path). Figure 1 below is the full mesh network of connections for only 6 elements (systems).

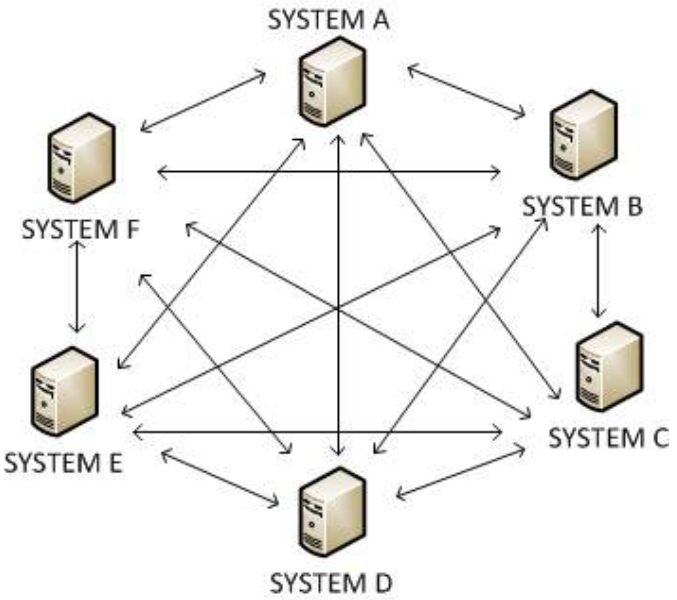


Figure 1. Bilateral interoperability model – full mesh network

A multilateral interoperability model is the answer to limitations of a bilateral model for complex systems rapidly growing with new elements. Managing a growing number of connections and security domains has become increasingly expensive and difficult. A multilateral model assumes the existence of a central element, which performs exchange of information between other systems. It reduces the number of connections to n-1 (where n is the number of elements) – adding at least one additional system when compared to the bilateral model.

Main advantages of this interoperability model are [1]:

- Fewer connections than in the bilateral model which ensures transparency and ease of adding new nodes (high adaptability and scalability).
- Thanks to the central point of integration it becomes easier to monitor and control performance of the whole environment.
- It forces data quality control and standardization (starting from the integration development processes).
- It ensures consistency in the data semantics.

Main disadvantages are:

- The difficulty of implementation in the already existing IT environment.
- Low level of reliability - a failure of the central system leads to a complete lack of communication in an integrated environment.
- It can lead to performance problems - the central system may become the bottleneck of the whole solution.

Figure 2 presents the multilateral interoperability model with full network of connections for 6 client systems (A - F) and one integrating system – hub.

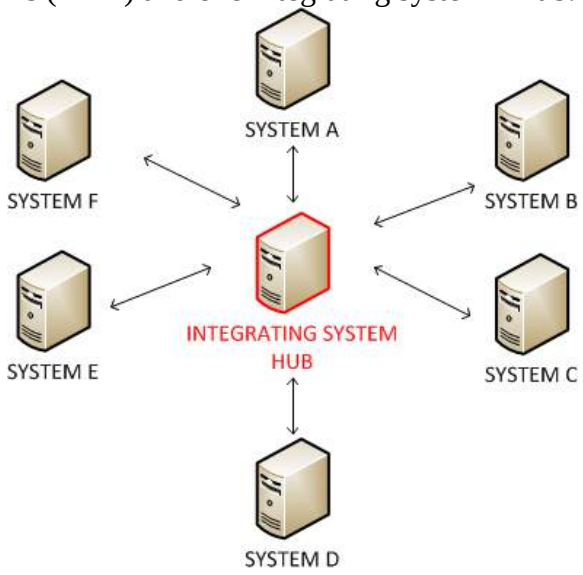


Figure 2. Multilateral interoperability model – full network

In order to eliminate performance and reliability problems, systems are integrated using ‘message bus architecture’. A single integration system – a ‘hub’ is replaced by the ESB (Enterprise Service Bus) mechanism, which provides a high level of reliability and does not significantly affect the performance.

A detailed comparison of bilateral and multilateral interoperability models has been presented in other author’s publication [5].

2.4 Integration Platform – Definition

An integration platform is defined as a “set of interrelated elements, which aims to create a cooperation environment for systems in order to deliver functions requested by the users of these systems” [1].

With the aid of discussed elements of an integration platform, it can be defined more precisely as a set of information systems that cooperate with each other. Cooperation is based on multilateral or mixed (multilateral with some external systems in bilateral) interoperability models. It gives access to multiple (more than one) electronic services provided at the individual request of a customer through a unified point of access (integrating system). However, a hub is not the owner of these services - the originating unit is still responsible for the service and its implementation (a hub does not replace integrated systems). An integration platform generates composite electronic services which encapsulate operational systems below (Service Oriented Architecture).

3 NEW SECURITY REQUIREMENTS

3.1 Service Oriented Security

Switching from data direct access systems to Service Oriented Architecture entails the need to change the security model.

In popular security models used these days access to data is performed through a specific set of elementary operations like reading or writing. The service architecture model adds service between the user and the data. Service processes the data and then transmit it to the user. The data processing can be a simple mechanism just to forward it in a unified format using the specified interface or can be more complex for example by combining it with other data and performing a set of statistical operations.

This means that in the service architecture we also have to protect “a service” as an object additionally to “a data objects” and its “access operations”.

Another important and new aspect is the security of a composite service. According to Gartner report [6] secure data exchange and secure communication management are the basis for aggregate (composite) public services. This enforces key requirements for integration platforms security model. It has to implement two operations:

- assessment of atomic (basic) electronic services originating from different sources in order to determine if integration into a composite service is possible (in terms of security),
- integration of atomic electronic services to one composite service which includes calculation method of new security attributes (level, category, class).

3.2 Existing Solutions

Currently there are many security models available starting from military models like the Bell-LaPadula model [7] and ending on business commercial models like the Brewer-Nash model (also known as the “Chinese Wall” model) [8] but according to author they don’t meet integration platforms needs.

Additionally information security requirements are well defined in the ISO 27001:2013 standard [9]. It describes steps which has to be taken like planning, operating, supporting and reviewing an information security management system but not indicating precisely which security model should be used in it. This is both advantage and disadvantage - giving more flexibility to security process stakeholders but making it more difficult to implement this standard. Usually a security model is already embedded in tools or systems available within organization. The problem begins when new architecture like SOA creates new objects and systems like integration platforms with their composite services.

In the previous author’s research [5] five assessment criteria were defined and 17 different security models were analyzed in terms of usefulness for composite electronic services security. Even if analyzed models had integration mechanisms like in the Smith-Winslett model [10] or the Jajodia-Sandhu model [11] they were strictly database models (almost impossible to implement in the service oriented architecture). On the other hand more general model like the Harrison-Ruzzo-Ullman (HRU) model [12] designed for protecting objects in operating systems can be easily converted to SOA environment but it doesn’t have any object security integration mechanisms.

3.3 Lattice Theory

According to the author's research [5] security models that are based on the lattice theory for example – Denning [13] and Szafrański [8] models have features which make them suitable as a basis for a new integration platforms security model.

Lattices "are ordered sets, for which the condition is satisfied, that for every pair of elements of the set, there is an upper and lower bound." [15] Lattice can be represented as:

$$(K, \leq, \oplus, \otimes) \quad (1)$$

where:

- K is a partially ordered set,
- $\leq$ is a partial order relation,
- $\oplus$ is the operation of determining upper bound (supremum) of its arguments,
- $\otimes$ is the operation of determining lower bound (infimum) of its arguments.

The lattice theory was used by Denning to create a secure information flow model [13] represented by:

$$MPD = (K, \leq, \oplus, \otimes, O, \rightarrow) \quad (2)$$

where partially ordered objects 'set O' and operation '$\rightarrow$' of allowed secure flow of information were added.

The lattice theory was also used by Szafranski [14] in his security model for data protection in distributed database systems. A database security is described with the data security lattice which is created from composition of flow and operation scope lattices. It can be used for distributed systems thanks to integration operation which transforms local security lattices (one for each database) to the 'security super-lattice' for the entire integrated environment.

This methodology allows to verify if integrated systems (in this case, distributed databases) can be integrated and allows to determine the security attributes for the entire integrated environment. A similar methodology can be used for integration platforms and composite electronic services.

Additionally to transformations presented by Szafranski [14] there are many new research [16, 17] on hierarchical lattices showing its usefulness not only in security but also in other fields.

A detailed analysis of lattice theory based security models has been presented in the other author's publication [18].

4 PROPOSITION OF A NEW SECURITY MODEL

4.1 Mathematical Model of an Integration Platform

An integration platform is a set of integrated electronic service systems (meeting the additional requirements specified in Section II) so it is important to define basic characteristics of such a system.

An electronic service - e_l is the smallest executable resource that processes (using algorithms which are part of e-service) data from data units. An electronic service consists of algorithms which are permanently tied up with it and will not be considered separately. Different combinations of algorithms generate various electronic services. Set of electronic services:

$$E = \{e_1, e_2, \dots, e_l, \dots, e_L\} \quad (3)$$

An entity - p_r is the customer of electronic services. Set of entities:

$$P = \{p_1, p_2, \dots, p_r, \dots, p_R\} \quad (4)$$

An electronic service is run by an enforcement w_n triggered by the entity or other electronic service. Set of enforcements:

$$W = \{w_1, w_2, \dots, w_n, \dots, w_N\} \quad (5)$$

An enforcement launches electronic services - function g.

$$g: W \to E \quad (6)$$

Entities P can launch enforcements within their permissions. Their permissions are defined by function u.

$$u: P \to 2^W \quad (7)$$

Enforcements launched by entities form a set of initiating enforcements (W_I).

$$W_I \subset W \quad (8)$$

Initial services (E_I) are registered on an integration platform and are the only services launched by entities (assumptions for the integrated environment).

$$E_I \subset E \quad (9)$$

$$\bigwedge_{w_n \in W_I} g(w_n) \in E_I \quad (10)$$

$$\bigwedge_{w_n \in W}(w_n \notin W_I \Longrightarrow g(w_n) \notin E_I) \quad (11)$$

$$\bigwedge_{e_l \in E_I} \bigvee_{w_n \in W_I} g(w_n) = e_l \quad (12)$$

Electronic services which can launch another services form a set of developmental services (E_w). Each initial service is also a developmental service.

$$E_I \subset E_w \subset E \quad (13)$$

Developmental services launch another services creating a services execution tree. Developmental services can't launch initial services to avoid loops on the integration platform.

$$z: E_W \to 2^W \quad (14)$$

$$\bigwedge_{e_l \in E_w}\big(w_n \in z(e_l)\big)\bigwedge(g(w_n) \in E_I) \Longrightarrow w_n \in W_I) \quad (15)$$

Finally any electronic service can execute operations (T) on data units (D).

$$h: E \to 2^{T \times D} \quad (16)$$

A data unit - d_i is an information object for storing data, for example: a file, a database record, a table, a database, a function parameter. Set of data units:

$$D = \{d_1, d_2, \dots, d_i, \dots, d_I\} \quad (17)$$

Data granulation depends on the level, at which access rights are considered for a particular system and may be different for individual units within the set of D.

A set of operations - T consists of elementary operations performed on data units d_i. Operations are universal and it is possible to specify set of them for every system. An example of operations are: write, read, search, delete, modify. Set of operations:

$$T = \{t_1, t_2, \dots, t_m, \dots, t_M\} \quad (18)$$

Operations can directly affect the lowest level data units (for example: the operation "read" the file) or data units which are higher level objects (for example: the operation "read" the database allowing to read any record from any table of the selected database). Function defining operation on data unit process:

$$f: T \times D \to D \quad (19)$$

A state of data unit s_j is being changed as a result of the operation t_m on the data unit d_i. A state change can be physical (for example: after "save" operation the data unit d_i has a new value) and formal (for example: after "read" operation the data unit d_i has still the same value, but it is in the new state "read" or even "compromised" if this operation was illegal). Set of all states:

$$S = \{s_1, s_2, \dots, s_j, \dots, s_J\} \quad (20)$$

States generator:

$$gst: D \to 2^S \quad (21)$$

$$gst(\mathrm{d}_i) = S_i \quad \text{for } \mathrm{d}_i \in D \text{ where } S_i \subset S;\ S_i \neq \emptyset \quad (22)$$

The state of the whole system is any possible subset of all data unit states of the given system. The state of the whole system:

$$S_{system} = S_1 \times S_2 \times \dots \times S_i \times \dots \times S_I \quad (23)$$

There are two possible ways to determine if specific system is still secure:

a) By controlling each system state after every operation. Impossible to implement in real systems because it requires a list of all states combinations with a description whether it is legal or illegal (from the security perspective).

b) By checking parameters of operations which lead to data unit state change. Implemented by functions (restrictions generators) and used in real systems.

The proposed security model is based on the second approach (b). It means that it has to consist of set of parameters and restriction functions. If result of a function is positive, it means that the system state after specific operation will be legal (from the security point of view). This allows to skip data unit states in further deliberations.

An integration platform consists of many electronic services which come from different systems and use different data sources. Figure 3 presents an example of a composite electronic service published in an integration platform environment. The entity (a service consumer) p_1 can invoke the composite e-service e_1 with the enforcement w_1. Selected service will trigger two data access operations t_1, t_2: saving some information to the data unit d_1 and reading from the data unit d_2. It will also invoke another e-service e_2 with the enforcement w_2. The electronic service e_2 will trigger one data access operation t_3: deleting the data unit d_3.

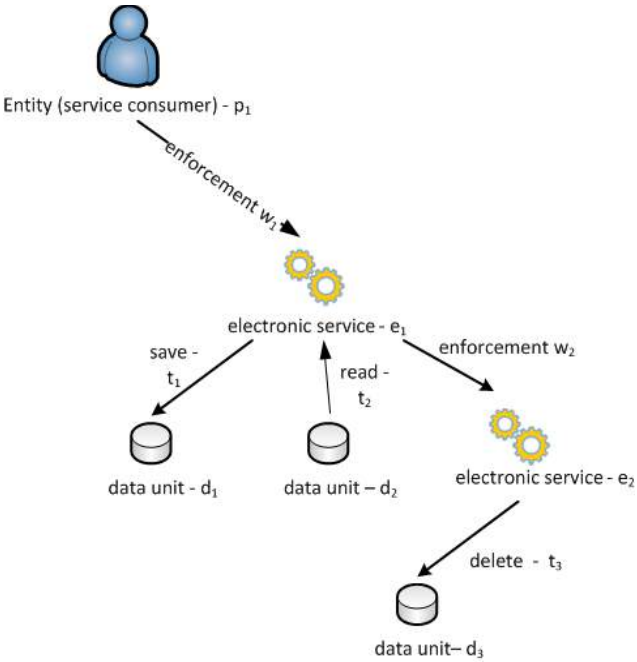


Figure 3. Example of a composite electronic service.

The composite electronic service is defined as:

$$u = (E_Z, e_I, g, z) \tag{24}$$

where:

- $\boldsymbol{E_Z \subset E}$ – electronic services (atomic services creating the composite service)
- $\boldsymbol{e_I \subset E_I}$ i $\boldsymbol{e_I \subset E_Z}$ – the initial service,
- $\boldsymbol{g: W \rightarrow E}$ – the function defining services launched by enforcements,
- $\boldsymbol{z: E_W \rightarrow 2^W}$ - the function defining enforcements launched by developmental services.

Additional assumptions for composite services:

$$\Lambda_{e_l \epsilon E_Z} \vee_{(e^* \epsilon E_Z) \wedge (e^* \epsilon E_W)} \vee_{\exists_{w^* \epsilon z(e^*)}} g(w^*) = e_l \tag{25}$$

$$\Lambda_{(e_l \epsilon E_Z) \wedge (e_l \epsilon E_W)} \Lambda_{w_n \epsilon z(e_l)} g(w_n) \in E_Z \tag{26}$$

A set of composite services:

$$U = \{u_1, u_2, \ldots, u_k, \ldots, u_K\} \tag{27}$$

An integration platform can be defined as a set of composite services, data units, operations, entities

and security management elements (a model with the security classification).

$$\langle U, D, T, P, SM \rangle \quad (28)$$

4.2 Security Management System

A integration platform security management system has to check entity (service consumer) parameters to allow or block execution of specified e-service. This means that all operations (data access and enforcements of next e-services) below have to be legal. There are two main problems that are solved by proposed model:

- Composite e-services can be very complex (a security model has to support process of triggering services by other services which is not present in currently available data access security models).
- Elements of composite e-services can originate from different systems (a security model has to support integration of security parameters from different systems).

The author's information security management model for integration platforms consists of two elements:

- data access management model - AM,
- e-service execution management model - EM.

Data units D are described by data protection classes, which are members of set K for example: public, confidential, secret, top secret. Electronic services are described by categories of an authorized execution, which are members of set B- for example: universal, special, restricted.

Each entity p_r from P set has the following parameters determining its access level:

- a protection class from the K set - for example: confidential
- a range of an authorized operation from T set – for example: reading,
- a category of an authorized execution from B set – for example: restricted.

4.3 Access Management Model

AM model determines, which access operations (T) are allowed on data units (D) when triggered by specific entity (P). Controlling functions are based on the protection class and range of authorized operations. AM model:

$$AM = \langle P, D, K, T, \rho, \tau, HF \rangle \quad (29)$$

where:

- P – a set of entities,
- D – a set of data units,
- K – a set of protection classes,
- T – a set of operations,
- ρ – an access relation,
- τ – a scope of operations relation,
- HF – a set of restrictions functions (restrictions generator for the access management model).

An access relation is built on pairs of protection classes: $\rho \subset K \times K$ and determines the hierarchy of allowed accesses to the data – for example: top secret > secret > confidential > public. A scope of authorized operations relation is built on pairs of operations: $\tau \subset T \times T$ and defines a hierarchy of operations – for example: update > delete > write > read > search. Both relations meet requirements of the lattice theory, which allows to build protection classes and operations lattices.

An entity is allowed to access a data unit using specific operations if its protection class is greater than or equal to protection class of a data unit and its scope of operations is greater than or equal to the operation being performed.

The restrictions generator for the access management model is defined as a set:

$$\mathrm{HF} = \{H_1, H_2, H_3, H_4, H_5, H_6, H_7\} \quad (30)$$

H_1- a function defining protection classes for entities (P) and data units (D):

$$H_1 : P \cup D \rightarrow K \quad (31)$$

H_2 - a function defining if a data flow between P and D objects is acceptable (1 – yes, 0 – no):

$$H_2: (P \cup D) \times (P \cup D) \rightarrow \{0,1\} \quad (32)$$

$$\bigwedge_{r_1,r_2 \in P \cup D} H_2(r_1, r_2) = \begin{cases} 1 \; if \; (H_1(r_1), H_1(r_2)) \epsilon \rho \\ 0 \; if \; (H_1(r_1), H_1(r_2)) \notin \rho \end{cases} \quad (33)$$

An access relation (ρ) is generating a hierarchical lattice from the protection classes set ($K \times K$). Fig. 4 presents an example of a protection classes lattice.

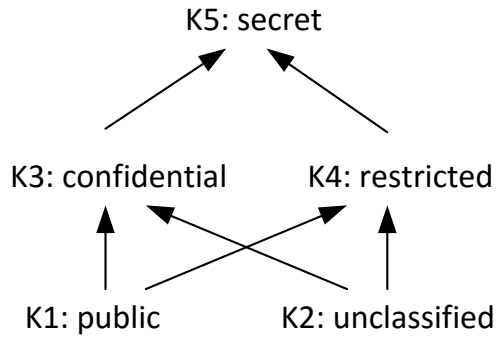


Figure 4. An example of a protection classes lattice

Further properties of an access relation and operations allowing integration of different K sets are not discussed in this paper.

H_3 - a function defining allowed operations which entity can perform on data units.

$$H_3: P \times D \rightarrow T \quad (34)$$

H_4 - a function defining an operation which is performed by an entity on a data unit object during an access request.

$$H_4: P \times D \rightarrow T \quad (35)$$

H_5 - a function defining if an operation execution is acceptable (1 – yes, 0 – no):

$$H_5: (P \cup D) \times (P \cup D) \rightarrow \{0,1\} \quad (36)$$

$$\bigwedge_{r_1,r_2 \in P \cup D} H_5(r_1, r_2) = \begin{cases} 1 \; if \; \left(H_4(r_1, r_2), H_3(r_1, r_2)\right) \epsilon \tau \\ 0 \; if \; \left(H_4(r_1, r_2), H_3(r_1, r_2)\right) \notin \tau \end{cases} \quad (37)$$

A scope of operations relation (τ) is generating a hierarchical lattice from the operations set ($T \times T$). Fig. 5 presents an example of a scope of operations lattice based on Linux permissions (R – read, W – write, X – execute / access directory contents). In the proposed model execution is extracted from operations and applied only to services, so X should be interpreted only as a 'access directory contents' operation.

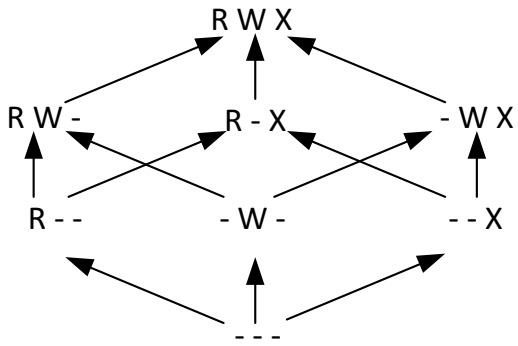


Figure 5. An example of a scope of operations lattice.

Further properties of a scope of operations lattice and different T sets integration mechanisms are not discussed in this paper.

H_6 – a function checking if an access request (a flow and an operation) is acceptable:

$$\bigwedge_{r_1,r_2 \in P \cup D} H_6(r_1, r_2) = H_2(r_1, r_2) \wedge H_5(r_1, r_2) \quad (38)$$

H_7 – a function checking if all access requests within a composite electronic service (E) are acceptable:

$$H_7(E) = \bigcap_{i \in I} H_6(\alpha_i, \beta_i) \quad where \quad \alpha_i, \beta_i \in P \cup D \quad (39)$$

An entity is allowed (according to the AM model) to execute a composite electronic service only if all access requests within this composite electronic service (E) are acceptable:

$$H_7(E) = 1 \quad (40)$$

4.4 Execution Management Model

EM model determines, which enforcements of electronic services (E) are allowed (which e-services can be executed by the entity (P) and other e-services (E)). EM model:

$$EM = \langle P, E, B, \delta, BF \rangle \quad (41)$$

where:

- P – a set of entities,
- E – a set of electronic services,
- B – a set of categories of an authorized execution,
- δ – a relation of an authorized service execution,
- BF – a set of restrictions functions (restrictions generator for the execution management model).

A relation of an authorized service execution is built on a pair of categories of an authorized execution: δ c B × B and defines execution permissions hierarchy – for example: restricted > special > universal. This relation meets requirements of the lattice theory, which allows to build authorized execution lattices.

An entity may execute an e-service if its 'category of an authorized execution' is greater than or equal to the 'category of e-service' (mandatory requirement). An e-service can invoke another e-service if its execution category is greater than or equal to the category of triggered e-service (strong optional requirement).

The restrictions generator for the execution management model is defined as a set:

$$\mathrm{BF} = \{G_1, G_2, G_3\} \quad (42)$$

G_1- a function defining categories of an authorized execution for entities (P) and electronic services (E):

$$G_1 : P \cup E \to B \quad (43)$$

G_2 - a function defining if an entity P is allowed to execute a electronic service E (1 – yes, 0 – no):

$$G_2 : (P \cup E) \times (P \cup E) \to \{0,1\} \quad (44)$$

$$\bigwedge_{q_1, q_2 \in P \cup E} G_2(q_1, q_2) = \begin{cases} 1 \; if \; (G_1(q_1), G_1(q_2)) \epsilon \delta \\ 0 \; if \; (G_1(q_1), G_1(q_2)) \notin \delta \end{cases} \quad (45)$$

An authorized service execution relation (δ) is generating a hierarchical lattice from the categories of an authorized execution set $(B \times B)$. An example of an execution categories lattice is presented in Section V on Fig. 7 as a super-lattice.

Further properties of an authorized service execution relation and operations allowing integration of different B sets are not discussed in this paper.

G_3 – a function checking if all service executions within a composite electronic service (E) are acceptable:

$$G_3(E) = \bigcap_{i \in I} G_2(\varphi_i, \omega_i) \quad where \quad \varphi_i, \omega_i \in P \cup E \quad (46)$$

An entity is allowed (according to the EM model) to execute a composite electronic service only if all service executions within this composite electronic service (E) are acceptable:

$$G_3(E) = 1 \quad (47)$$

In order to determine if a composite electronic service is allowed to be executed by an entity it has to be checked by the information security management system. It means it has to be checked in parallel or in series (order doesn't matter in this case) using access and execution management models. Electronic services can originate from different local systems and their protection classes, operations and categories of an authorized execution will be checked and integrated.

The integration process of local lattices consists of several steps. In simplification four can be pointed out:

1. Set of local lattices is checked for consistency (a decision if they can be integrated).
2. Construction of a super-lattice.

3. Reduction of a supper-lattice.
4. Transformation of a supper- lattice.

5 CASE STUDY

5.1 Public e-services integration – the current model

In order to demonstrate usefulness of the proposed model it is necessary to compare it to the current solution. This case study presents execution of simple but integrated public e-service: registration of a newborn child.

Without systems integration citizen had to visit two different offices to complete this public service:

- Register Office to receive a birth certificate,
- Municipal Office (department of Population Registry and ID Cards) corresponding to the registered address to generate and receive personal identification number (PESEL) for a newborn child.

When those offices and their systems are integrated it is enough to visit only Register Office and an official can do both operations for a citizen as a one integrated service. However it is not delivered electronically to the citizen but is an e-service for an official (it can be treated as a public electronic service for this case study purpose).

Each office has different IT systems, different security user groups and policies. Electronic services in accordance with the applicable Polish law should be integrated using integration platforms – like 'e-PUAP' (Electronic Platform of Public Administration Services) or 'Source' (original name in Polish is 'Źródło' - a system that is integrating public registers).

Figure 6 presents simplified model of 'registration of a newborn child' e-service (blue services are Register Office services, yellow service is a Municipal Office service, red is a initiator service).

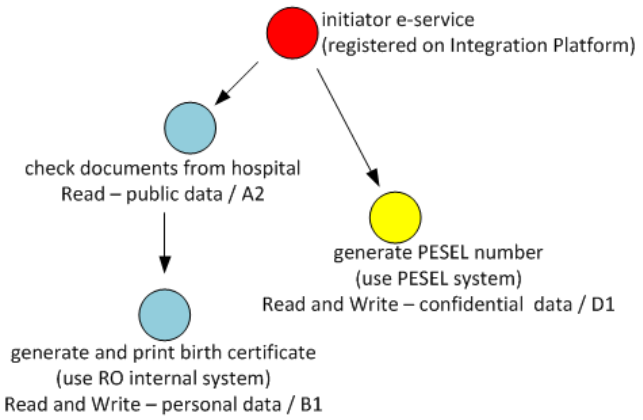


Figure 6. Simplified service example - 'registration of a newborn child'

For this case study purpose we can assume that each office has similar protection classes and authorized operation (for the data):

- K: public < personal < confidential < secret
- T: read < write < update < delete

Each office has different user groups which result in different authorized execution classes.

For Register Office it is:
B_{RO}: A1 (authenticated citizen) < A2 (authenticated official) < B1 (official with public documents authorization) < C1 (head of Register Office department)

For Municipal Office it is:
B_{MO}: A1 (authenticated citizen) < A2 (authenticated official) < D1 (official with PESEL system authorization) < E1 (head of Municipal Office)

New security level of integrated public e-services in usually a decision of an expert. Different security groups are used in each office so according to the actual model there will be a new security group created – F1. Users in the new group will be allowed to run selected e-service (more precisely right to execute a initiator service). In most cases all users from B1 security group will be added to F1 group to allow all officials in Register Office to run both services as an integrated solution. For complicated integrated services data access rights are not checked for every component originating from other system

so we can assume: 'read and write – personal data / F1' as the minimum security requirement for Register Office officials to run analysed integrated e-service (in the actual security model). In this case data access rights are checked locally and user has to be the member of security group created specifically for this new e-service.

This means that:

- there is no automation in e-service integration (need of expert every time new integrated e-service is created or changed),
- new security groups are often created specifically for new integrated e-services,
- security compliance is not checked for every component in every integrated system. In case study example proposed security level is too low as Municipal Office is processing confidential data in PESEL system so it should be 'read and write – confidential data / F1'. This would result in the correct situation that not all officials from Register Office can execute integrated e-service, even if they are in the proper group (F1). They need additional data access rights permissions.

5.2 Public E-services Integration – the New Model

The same example is discussed using mandatory requirements from the new proposed security model. The first step is system integration which requires security attributes integration. All security attributes of analyzed systems meet lattice requirements - parameters K, T, B_{RO} and B_{MO} are partially ordered sets so they can be treated as local lattices. Local 'protection classes – K' and 'authorized operations – T' are the same in both systems, so integrated super-lattice will be identical to local. The super-lattice of authorized execution classes B_L, presented in Fig. 7, is calculated using lattice transformations.

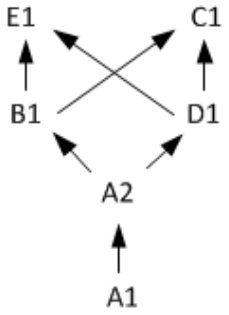


Figure 7. Super-lattice of authorized execution classes.

After completion of described procedure, every electronic service consisting of any atomic e-services from analyzed integrated systems can be automatically checked for resulting permissions. Every user (entity) trying to execute these electronic services will be checked against access and execution security policies.

Register Office official from the example needs at least 'read and write confidential data' as all data access rights are checked in K and T super-lattices. He also needs at least B1 'execution authorisation' as B1 is greater than or equal to D1 and A2 (from B_L super-lattice).

This means that:

- there is automation in e-service integration (possible need of expert only one time during systems integration, e-services can be created and changed with automatic security level recalculation),
- no need of new security groups (available security groups are integrated in super-lattices),
- security compliance can be checked for every component in every integrated system as they share common security schema defined with super-lattices.

Presented example is simple but integrated e-services can become very complex consisting of many electronic services from many different systems. Lack of well-defined security model for public integration platforms is a delaying factor in development of this area of public administration.

Long composite e-service creation time and need of experts makes integration projects very expensive.

Some complex calculations like lattices to super-lattice transformations have been deliberately omitted in this example, as purpose of this article is to present the idea of the new model, not its detailed mathematical proof. To achieve a complete secure solution in real environments, proposed security model has to be accompanied by additional functions like: authorization and authentication mechanisms, threat and antivirus protection, encryption and events logging.

6 CONCLUSIONS

This article presented only an introduction to the proper information security model described in the author's doctoral dissertation (in preparation). New security model for composite electronic services published on integration platforms, which was introduced in this paper meets all the requirements placed on complex heterogeneous service oriented systems. Proposed AM and EM models (thanks to the lattice theory) can be used for integrated e-services originating from different systems.

Further, new model will be tested on polish public administration integration platforms. It is an answer to requirements defined by polish government to provide composite public e-services, which "will be produced in process of integration (interoperability) of different systems" [19]. Presented model can also be easily adopted to private sector integration platforms for example: banking and insurance integration systems. It can also be used for military integration mentioned in "Multilateral Interoperability Programme strategy to achieve interoperability of cooperating national Command and Control Information Systems at all levels of command, in support of multinational, combined and joint operations" [20].

Service oriented architecture and integration of separate systems is the main trend in 21st century IT. However, a new systems architecture requires a new approach to information security. Direct access security models are not suitable for e-service systems, like integration platforms and increasingly popular cloud environments.

7 REFERENCES

1. T. Górski, „Integration platforms - selected issues", („Platformy integracyjne – zagadnienia wybrane"), PWN, Warsaw, 2012, pp. 12, 46, 189.
2. Oasis Open, "Reference Model for Service Oriented Architecture 1.0", Committee Specification 1, 2006, pp. 7-8.
3. G. Birkhoff, "Lattice theory", American Mathematical Society Colloquium Publications, XXV, New York 1940, II edition 1948.
4. European Commission, "Council Implementing Regulation (EU) No 282/2011", 23 March 2011, Article 7, paragraph 1 and II, apendix 1.
5. J. Wilk, "Security management in an distributed environment", ("Zarządzanie bezpieczeństwem w środowisku rozproszonym"), „Business development: multidimensionality of conditions and consequences", ("Rozwój działalności przedsiębiorstw: wielowymiarowość uwarunkowań i konsekwencji"), Warsaw, 2015, (in press).
6. Gartner Inc., "Preparation for Update European Interoperability Framework 2.0 – Final Report", 04-06.2007, chapter 4.5 "Generic Public Services Framework".
7. D. E. Bell, L. J. LaPadula, "Secure Computer Systems: Mathematical Foundations" – MTR-2547, Vol. I, MITRE Corporation, Bedford, MA, 1973.
8. D. F.C. Brewer, M. J. Nash, "The Chinese Wall Security Policy" Gamma Secure Systems Limited, Surrey, United Kingdom, 07.2001.
9. International Organization for Standardization, "ISO/IEC 27001:2013 - Information technology - Security techniques - Information security management systems - Requirements", 2003.
10. K. Smith, M. Winslett, "Multilevel secure rules: integrating the multilevel secure and active data models", Sixth Working Conference of IFIP Working Group 11.3 on Database Security on Database security, 1992, pp. 35-53.
11. S. Jajodia, R. Sandhu, "Toward A Multilevel Secure Relational Data Model,, Information Security: An Integrated Collection of Essays", IEEE Computer Society Press, 1994. pp. 3 – 6
12. M. A. Harrison, W. L. Ruzzo, J. D. Ullman, "Protection in Operating Systems". Communications of the ACM 19 (8), 1976, pp. 461–471.
13. D. E. Denning, P. J. Denning, "Certification of Programs for Secure Information Flow", Purdue University, 03.1976, pp. 6.
14. B. Szafrański, "Security process modeling for databases, with particular emphasis on their integration", („Modelowanie procesów ochrony baz danych ze szczególnym uwzględnieniem ich integracji"), Military Univesity of Technology, Warsaw, 1987.
15. H. Rasiowa, „Introduction to modern mathematics", („Wstęp do matematyki współczesnej"), PWN, edition XXIV, Warsaw, 2005, pp. 123.
16. Y. Shang, "Percolation in a Hierarchical Lattice", Zeitschrift für Naturforschung A, 2012, vol. 67, no. 5, pp. 225-229.
17. Y. Shang, "Phase Transition in Long-Range Percolation on Bipartite Hierarchical Lattices", The Scientific World Journal, 2013, vol. 2013.

18. J. Wilk, „The use of lattice theory in the modelling of safety management processes in public administration electronic services platforms", („Wykorzystanie teorii krat w modelowaniu procesów zarządzania bezpieczeństwem w platformach usług elektronicznych administracji publicznej"), The Collegium of Econimic Analysis, Warsaw, 2015.
19. M. Boni, A. Boboli, T. Jeruzalski, M. Olszewska, R. Paleń, A. Ręgowski, A. Siejda, "State 2.0. A new start for e-government" („Państwo 2.0. Nowy start dla e-administracji"), Ministry of Administration and Digitization, Warsaw 2012, pp. 7.
20. Multilateral Interoperability Programme (MIP), "MIP Vision & SCOPE (MV&S)", PMG Edition 6.3, Montebello, Canada, 2009.

International Journal of
NEW COMPUTER ARCHITECTURES AND THEIR APPLICATIONS

The *International Journal of New Computer Architectures and Their Applications* aims to provide a forum for scientists, engineers, and practitioners to present their latest research results, ideas, developments and applications in the field of computer architectures, information technology, and mobile technologies. The IJNCAA is published four times a year and accepts three types of papers as follows:

1. **Research papers**: that are presenting and discussing the latest, and the most profound research results in the scope of IJNCAA. Papers should describe new contributions in the scope of IJNCAA and support claims of novelty with citations to the relevant literature.
2. **Technical papers**: that are establishing meaningful forum between practitioners and researchers with useful solutions in various fields of digital security and forensics. It includes all kinds of practical applications, which covers principles, projects, missions, techniques, tools, methods, processes etc.
3. **Review papers**: that are critically analyzing past and current research trends in the field.

Manuscripts submitted to IJNCAA **should not be previously published or be under review** by any other publication. Plagiarism is a serious academic offense and will not be tolerated in any sort! Any case of plagiarism would lead to life-time abundance of all authors for publishing in any of our journals or conferences.

Original unpublished manuscripts are solicited in the following areas including but not limited to:

- Computer Architectures
- Parallel and Distributed Systems
- Storage Management
- Microprocessors and Microsystems
- Communications Management
- Reliability
- VLSI