QUEST RJ

ISSN 1605-8607
Bi-Annual

Quaid -e- Awam University Research Journal of Engineering, Science & Technology

MISSION STATEMENT OF QUEST

"To provide quality and state-of-the- art education (coursework, practical training and research) in prescribed branches of Engineering and Science to the enrolled students in order to make them better professionals and better human being; so that they become capable of contributing amicably towards national development".

Volume 11 No. 2 JULY - DEC, 2012

Bi - Annual

ISSN 1605-8607

Quaid-e-Awam University Research Journal of Engineering, Science & Technology

EDITORIAL BOARD

ANNUAL SUBSCRIPTION..... RS. 200.00 (INLAND0, US$20.00 (FOREIGN, BY SURFACE MAIL)
SINGLE ISSUE ..... RS. 100.00 (INLAND0, US$10.00 (FOREIGN, BY SURFACE MAIL)

The Bi-Annual Quaid-e-Awam University Research Journal of Engineering, Science & Technology, Shall be supplied free of cost in exchange of Research Journal (s) from other Universities/ Institutions and Research Centers.

Bi - Annual

ISSN 1605-8607

ACKNOWLEDGEMENT

The members of Editorial Board, Quaid -e- Awam University Research Journal of Engineering, Science & Technology are grateful for valuable and critical suggestions on various Research paper(s) sent to following Researchers/Experts for Volume 11 No. 2 July - December 2012 issue. The members also appreciate the Referees/Experts for sharing their knowledge and expertise towards improvement of standard of this research journal.

Prof. Dr. Muhammad Hanif Durad
PIEAS, Islamabad

Prof. Dr. Attaullah Khawaja
Department of Electronic Engineering
NED, Karachi

Prof. Dr. Muhammad Akaram Shaikh
PASTIC, Islamabad

Prof. Dr. Muhammad Ayaz Keerio
Department of Computer Science,
University of Sindh, Jamshoro

Prof. Dr. Ghulam Ali Mallah
Department of Computer Science
Shah Abdul Latif University, Khairpur

Prof. Dr. Imtiaz Korejo
Department of Computer Science,
University of Sindh, Jamshoro

Prof. Ghulam Murtaza Mastoi
Dr. M. A. Kazi Institute of Chemistry,
Universty of Sindh, Jamshoro

Prof. Dr. Ghulam Qadir Shar
Department of Chemistry,
Shah Abdul Latif University, Khairpur

ISSN 1605-8607

A STUDY RELATED TO PROTOCOLS AND PORTS USAGE FOR THE ISA SERVER

Pardeep Kumar*, Iftikhar A. Koondhar**, Akhtar H. Jalbani***, Agha Shiraz A. Khan****, M. Aamir Bhutto*****

ABSTRACT

Managing and monitoring networks have never been so important as today. Although there are variety of tools available in the market to secure and monitor the network from malicious attacks but these tools do not provide detailed analysis capability of network usage patterns and trends. Analysis of network usage logs to uncover the activities of internal users is a good way of managing the network security and bandwidth related issues. ISA Server generates very detailed access and security logs for all the web requests made through it, but it doesn't provide efficient and interactive analysis capability of the same. This paper aims to provide multidimensional and interactive (providing drill-down, roll-up and slicing & dicing) analysis capability of proxy logs generated by ISA Server.

The solution suggested in this paper can generate very customizable and interactive multidimensional reports, which can help in better understanding the network usage patterns and trends of the users including the details about the website, ports, protocols, bandwidth usage and malicious activities attempted by the users while accessing the network. Moreover, the analysis of ISA Server logs has revealed the ports and protocols commonly being used at the Server. In addition it has been found that the unassigned ports by IANA are never used for any useful browsing, hence they may be blocked.

1. INTRODUCTION

ISA (Internet Security and Acceleration) Server from the Microsoft Corporation is the follow-on release of their Proxy Server 2.0 and is also a part of the .Net Family. Note the all the abbreviations that have been used in this paper are given in Table 1. Web log mining is an important application in the Web Mining research fields. The aim of Web log mining is to find out user access patterns of Web sites. The process mainly includes four steps: data collection, log preprocess, pattern recognition and pattern analysis. First, we extract available information from raw access logs; insert them into databases, certain each user based on heuristic rules, and dig out each user's access sequence [12].

The basic services of ISA Server include an enterprise firewall and a Web proxy (a.k.a. cache server) [1, 2, 10]. ISA Server's firewall monitors all the network traffic that flows through it whereas the Web cache provides storage and faster access to the frequently accessed web pages in order to lessen the network rush. The ISA Server does this by downloading the updates of frequently accessed web pages when the server is in idle state.

Beside this, ISA Server provides very comprehensive security and access logs for all the traffic that passes through it. These logs contain tons of data which can reveal hidden patterns if analyzed properly including:

Table 1: Abbreviations used in the paper

ELT	Extract Load Transform
ETL	Extract Transform Load
FTP	File Transfer Protocol
HTTP	Hyper Text Transform Protocol
IANA	Internet Assigned Network Authority
ICPC	Information Communication Processing Center
IP	Internet Protocol
ISA	Internet Security and Acceleration
LAN	Local Area Network
RDBMS	Relational Database Management System
RPC	Remote Procedure Call
SMB	Server Message Block
SQL	Structured Query Language
SSIS	SQL Server Integration Services
TCP	Transmission Control Protocol
URL	Uniform Resource Locators

*,** Assistant Professor, Department of CSE, Quaid-e-Awam University of Engineering, Science & Technology, Nawabshah, Sindh
*** Associate Professor, Department of IT, Quaid-e-Awam University of Engineering, Science & Technology, Nawabshah, Sindh
****,***** Lecturer, Department of CSE, Quaid-e-Awam University of Engineering, Science & Technology, Nawabshah, Sindh

- URL(s) visited by the users
- Protocol(s) used for communication
- Port(s) used for accessing hosts
- Browser used to made request
- On-Time and Off-Time internet usage details
- Get notified on excessive internet usage

These results can help network administrator in securing the ISA server from unauthorized use through LAN service ports, saving the network bandwidth which may be consumed heavily by the unidentified protocols and providing caches for the most often visited pages.

Logs Collection is the very first phase of this research. Detailed security and network usage logs are generated for all the traffic that passes through the firewall service and the web-caching service of the ISA Server. These logs are in the form of flat text files, which should be parsed and formatted before they may be analyzed. Logs can be generated on daily, weekly monthly or yearly basis [10]. ISA Server generates three types of logs depending on the type of installation and configuration, namely *packet filter*, *firewall* and *proxy logs*.

1.1 PACKET FILTER LOGS

ISA Server logs each and every packet that falls against the packet filtering rules. Packet filtering logs contain only the dropped packet however it may be configured for logging allowed packets.

1.2 FIREWALL LOGS

These logs contain detailed information regarding the data sent through the ISA server's firewall service. Firewall logs include user and host IP address, ports and protocol used for communication, bytes sent and received for any request, time taken to fulfill any request and the date and time of browsing. Beside this firewall service log properties can also be adjusted.

1.3 PROXY LOGS

Proxy logs contain almost the same attributes as that of firewall logs but these logs are specific to the configuration of the web proxy server.
All the above mentioned types of logs can be obtained in two formats, i.e., (1) flat text file and (2) ISA Server can be configured to move the logs to any RDBMS including Oracle, SQL Server.
In this research only the details and usage of proxy logs for Internet are analyzed. These logs are available in the form of flat text files; however ISA Server may also be configured to directly move the logs to some predefined database of any RDBMS. The flat files are generated on daily basis i.e., a single flat text file is generated daily that contains the details of all the hits made on ISA Server on that particular day. The following screen shot depicts a typical flat text proxy log generated by ISA Server.

2. RESEARCH METHODOLOGY

The overall research methodology consists of following four phases as shown in Figure 1.

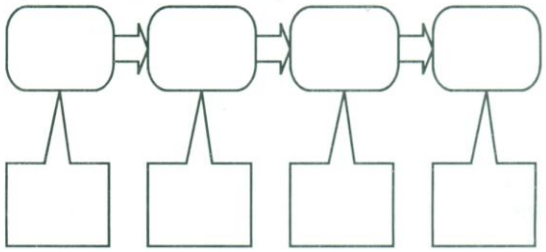


Figure 1: Research Methodology

A. Log collection
B. Data loading
C. Transformation
D. Analysis

2. 1 LOG COLLECTION

During the first phase, log information about the web usage is collected from the ISA server in order to process it further for the data loading phase. Figure 2 shows a snap shot of the ISA log file.

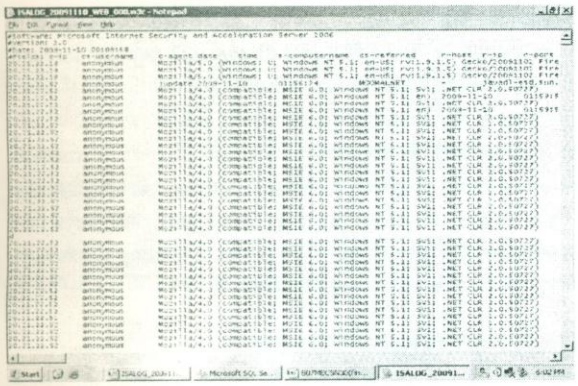


Figure 2: A snap shot of the ISA log file

2. 2 DATA LOADING

The flat text logs generated by ISA Server need to be loaded into database of any RDBMS so that it may be parsed, transformed and analyzed. In this research we have used SQL Server 2005 as the target RDBMS. SQL Server 2005 provides a lot of services for the loading and transformation of data. In this research we have used SQL Server Integration Services (SSIS) for the loading and transformation of ISA Server logs.

For loading the logs, a thorough study of the same is required. The logs need to be studied for the attributes available, their field width and the data types. Since every log file contains millions of records, the loading phase is one of the most time consuming phase of the whole ELT cycle. The logs are initially loaded into the staging area and then transformed and moved to the final database. Figure 3 shows the loading phase of ELT cycle using SQL Server 2005. Here the source is the flat text file and the destination is SQL Server 2005.

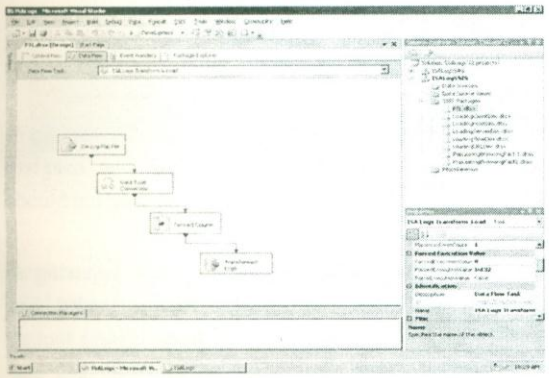


Figure 3: Loading data through SQL server integration service

2. 3 DATA TRANSFORMATION

Once the data is in SQL Server staging database (temporary database), it needs to be transformed for the following reasons.

The data loaded is in a single table in the staging area and it needs to be transformed according to the multi-dimensional schema. Some of the attributes in the staging area are useless, so they need to be dropped. Date and time formats conversion is needed. Aggregates of numerical attributes are needed for fast and quick analysis etc.

Figure 4 depicts the transformation phase used in this process. This transformation will convert the schema of the database and is going to populate the fact and dimension tables.

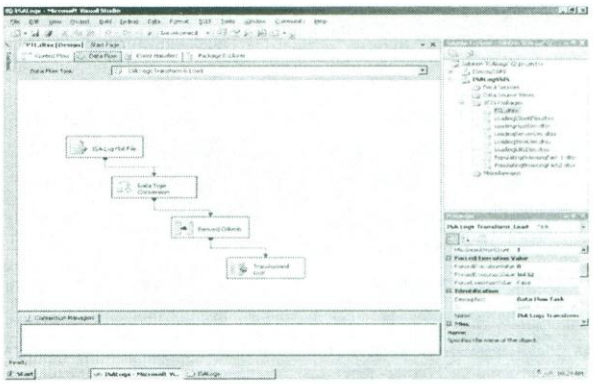


Figure 4: Transformation of Logs using the SQL server integration services

2. 4 ANALYSIS

The final phase of the research methodology is the data analysis, which is achieved by using one of the most powerful data analysis tool named COGNOS. COGNOS is a multi - dimensional reporting tool that delivers managed reporting for consistent, fact-based decision-making. Managed reporting enables report authors to create reports drawn from any data source. These reports can then be delivered to report consumers.

Data Analysis is one of the most important phases. A lot of textual and graphical reports are generated while analyzing the logs which discovered many hidden network usage pattern and trends.

3. SCHEMA TRANSFORMATION

As stated earlier, complete ISA logs are loaded into a single table, means there is no logical schema of the loaded data. Moreover absence of logical model will also result in data redundancy and analysis chaos.

For the analysis of huge amount of data it is always recommended to use dimensional model. Since dimensional model is denormalized, it results in faster analysis of data. The dimensional model used in this paper is as follows:

3. 1 ISA LOG'S DIMENSIONAL MODEL

In the schema above, there is a central fact table containing the foreign keys and numerical attributes surrounded by four dimension tables namely Client, Host, Server and Time dimension. Customarily, the central fact table contains numerical measures or the attributes on which we want to perform calculations. In my case the factual attributes include Bytes Sent, Bytes Received and Time Taken in fulfilling any request. Fact table is always surrounded by dimension tables which contain non-numeric or textual attributes. As the ISA log contains attributes related to client, server and host; hence there are four dimensions client, server, host and time dimension. Time dimension is considered as the mandatory dimension of all dimensional models. Figure 5 depicts the dimensional model of the ISA log.

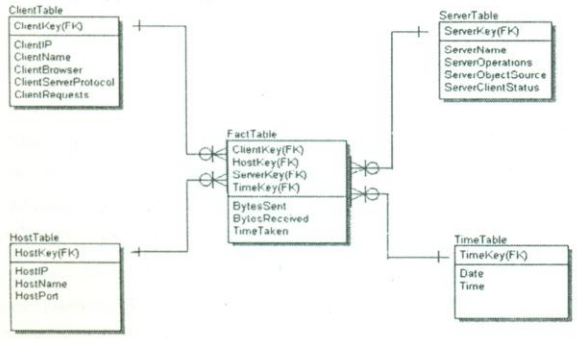


Figure 5: Dimensional model of the ISA log

3. 2 SURROGATE KEY GENERATION

In case of dimensional model, fact and dimension tables are linked by a special type of key known as surrogate key. Surrogate keys are simple sequential keys used by dimension tables. This key helps in linking fact and dimension tables. Due to the simplistic nature of surrogate key, joining between fact and dimensions become fast and data retrieval efficiency increases.

Surrogate key generation is one of the transformation stages. In this stage an additional attribute is added to the dimensions. In this project four surrogate keys are generated for four dimension tables including ClientKey, ServerKey, HostKey and TimeKey

3. 3 DATA TYPE CONVERSION

The data type of the data loaded in the loading stage is usually character type. However for different attributes there need to be different data types, e.g. for numerical attributes there need to be integer data type, for date time related attributes there need to be date data type.
As mentioned above, SQL Server Integration Services (SSIS) may be utilized for all of the above discussed transformation including schema conversion, surrogate key generation and data type conversion. The screenshot shown in Figure 6 shows the schema conversion and surrogate key generation.

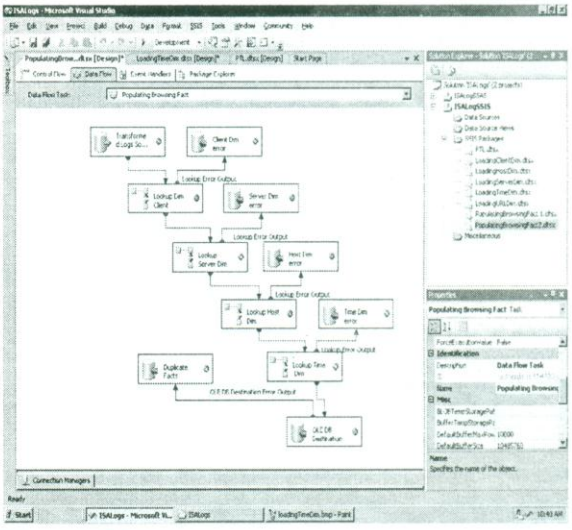


Figure 6: Data type conversion

In the above screenshot, data staging area is behaving as the source. From the staging area data is loaded into different dimensions and fact tables. In each stage lookup table is used for the generation of surrogate key and finally the data is loaded into the dimensional model.

4. LOG ANALYSIS

ISA Server Proxy logs are analyzed for network/internet usage patterns. Some of the analysis includes Bandwidth usage with respect to ports and protocols Sites accessed through unassigned ports popular hosts on popular ports etc.

4. 1 PORT TRAFFIC SUMMARY

The report shown in Table 2 summarizes the network traffic on different ports. Moreover ports are ranked with respect to the bytes received on them while browsing. Port numbers 80, 443 and 1935 are the most traffic bearing ports respectively.

4. 2 PORT CLASSIFICATION

The IANA has divided the port numbers into three ranges well-known ports, registered ports and dynamic (or private) ports.

Table 2: Port summary

Port Number	Bytes Sent	Bytes Received
21	343515399	2872
80	1853719555	43074735
436	0	0
443	8277698	1671041
1935	6392606	316322
8080	220405	11134
8090	0	0
8101	0	0
9816	0	0
Total	2212125663	45076104

Well-known ports: The ports ranging from 0 to 1023 are assigned and controlled by IANA; Well-known ports should not be used without IANA registration.

Registered ports: The Registered ports are those from 1024 through 49151; registered ports should not be used without IANA registration.

Dynamic (or private) ports: The Dynamic and/or Private Ports are those from 49152 through 65535. Dynamic ports are neither controlled nor registered. They can be used by any process [11].

As listed in Table 2, port numbers 9816, 8090 and 8101 are the 4th, 5th and 6th most heavily used ports respectively and all three of these ports are unassigned by IANA. It is interesting to note that unassigned ports are normally used by useless sites e.g. port number 9816 is used exclusively by host '*streamsolutions.co.uk*' for audio and video streaming, port number 8090 was used by '*videos.asiantsunamivideos.com.nyud.net*'.
Table 3 draws top 15 heavily ports used by the port 80. Out of the 15 ports, 7 ports are unassigned which may be blocked to save some of the network bandwidth.

Table 3: Top 15 hosts for the port 80

Port 80	Received Bytes	Sent Bytes
www.facebook.com	296	2107
www.rapidshare.com	602	33806
www.google.com	536	7932
www.youtube.com	284	117222
www.orkut.com	287	49872
www.wikipedia.com	370	14440
www.onlinewatchmovies.net	415	978
www.islamonline.com	297	6277
www.tagged.com	409	119838
www.manjam.com	359	1239
www.9adultsexgames.com	545	48972
www.hentaicake.com	611	8609
www.hardcartoon.com	507	19034
www.myxxxtoon.com	406	22493
www.thumbparty.com	428	17078
Total	6352	470097

4. 3 POPULAR HOSTS ON PORT 80

The list report presented in Table 3 summarizes the top 15 hosts with respect to bandwidth consumption in a week on port number 80, which concludes that the almost 97% of the whole traffic travels through this port.

4. 4 HOST ON PORT 21

Table 4 presents the list of hosts and their traffic conditions on port 21. This port has specifically been used

only for the FTP communication between the server and clients.

Table 4: Top hosts for the ftp use on port 21

Port 21	Bytes Received	Bytes Sent
ftp://20.21.22.3/indian%movies	371	3686
ftp://20.21.22.3/indian%songs	420	1575
ftp://20.21.22.3/english%movies	371	3668
ftp://20.21.22.3/english%songs	355	3358
ftp://20.21.22.3/ring%tones	402	783
ftp://20.21.22.3/mobile%videos	204	915567
ftp://20.21.22.3/softwares	525	6532
ftp://20.21.22.3/mobile%converters	426	5849
ftp://20.21.22.3/mobile%softwares	356	6954
ftp://20.21.22.3/sindhi%songs	254	2587
ftp://20.21.22.3/sindhi%videos	110	6589
Total	3794	957103

4. 5 HOSTS ON PORT 443

The traffic usage on port 443 is depicted in Table 5.

Table 5: Top hosts for port 443

Port 443	Received Bytes	Sent Bytes
www.urs.microsoft.com	284	4665
www.secure.tagged.com	2164	3447
www.google.com	8682	10904
www.login.wetpaint.com	3590	19358
www.images.wetpaint.com	1416	6842
www.wetpaint.login.rpxnow.com	1230	15073
www.ajax.googleapis.com	836	22260
www.s3.amzonws.com	915	5583
www.sstats.wetpaint.com	2212	4353
www.rpxnow.com	895	361
www.facebook.com	1542	2544
www.client4google.com	9310	1633
www.beliscity.com	859	54599
www.loging.yahoo.com	995	16229
www.s.yimg.com	886	3227
www.login.live.com	2688	5274
Total	3770	179604

4. 6 PROTOCOL SUMMARY

Normally HTTP protocol is used for browsing. Beside http; SSL, HTTPS and FTP are also being used but their usage is not quite significant. Figure 7 presents the overall usage of the protocols that clearly shows that the HTTP protocol is mostly used. Moreover, no any anonymous protocol was found in the two weeks logs.

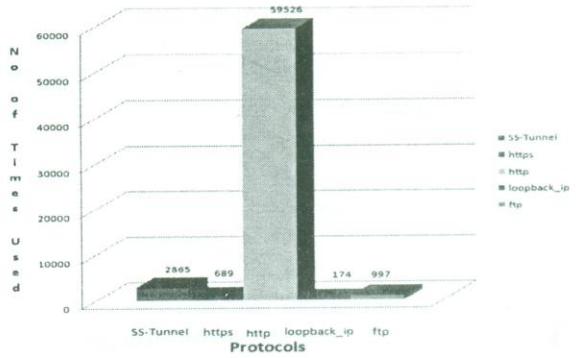


Figure 7: Protocol summary

5. FINDINGS OF THIS RESEARCH

As a result of analysis of list reports and graphs generated from the ISA Logs, a lot of patterns and trends have been discovered. Some of the findings of the analysis are discussed below.

5. 1 PORTS

In TCP/IP networks, a port is a transport layer address that the client program must specify in order to communicate with a specific server application on a computer in the network. In order to communicate on the web, one should specify the port number of the remote host. Normally port number 80 is used as the default port for World Wide Web; however some hosts make use of other assigned and unassigned ports to communicate on TCP/IP network.

5. 2 DATA ON ASSIGNED V/S UNASSIGNED PORTS

From the analysis of data traffic on different ports, it has been discovered that 80, 21 and 443 are the top three most heavily used assigned ports. Port numbers 80 and 8080 are used by a well known protocol HTTP for accessing the World Wide Web, approximately 97% of the whole network traffic travels through these ports. Port numbe

21 is the second most heavily used ports used by FTP. FTP is a file transfer protocol employed for uploading, downloading and manipulating files over a TCP network. Moreover port number 443 is employed by HTTPS-SSL protocol for encrypted communication of data over TCP/IP network for security reasons.

Port numbers 1935 and 8080 are the 4th and 5th most heavily used ports respectively and all four remain of these ports are unassigned by Internet Assigned Numbers Authority (IANA). It is interesting to note that unassigned ports are normally used by useless sites e.g. Port number 9816 is used exclusively by host 'streamsolutions.co.uk' for the audio and video streaming and port number 8090 was used by 'videos.asiantsunamivideos.com.nyud.net' for downloading tsunami videos.

5. 3 PORT NUMBER 80

Port 80 is the default port employed for browsing the World Wide Web. As port 80 is the most heavily used port, data on this port is analyzed for possible bandwidth threats. As a result of careful analysis of port 80 for a week, it has been revealed that approx 66% of bandwidth on this port is wasted on useless surfing and Most of porn sites or links are visited via this port.

5. 4 PORT NUMBER 21

Port number 21 is the second most used port and this is should also be monitored carefully so that any prohibited or restricted link or URL might be accessed via server.

5. 5 PORT NUMBER 9816

9816 is the unassigned port by Internet Assigned Numbers Authority (IANA) and this port is used exclusively by 'streamsolutions.co.uk' for audio and video streaming and which causes the use of bandwidth uselessly.

5. 6 PORT NUMBER 443

Port number 443 is another port by which a very large number of users are accessing the server. The traffic at this port is observed less as compare to port 80 and port 21.

5. 7 PORT 8090

8090 is another unassigned port like port number 8101, 9816 and 436, and these ports might be very dangerous for the attackers and which may be waste of bandwidth usage.

5. 8 PORT NUMBER 8080

Port 8088 is the unassigned port, sometimes used for accessing or downloading the audio and video files or movies.

5. 9 BANDWIDTH HUNGRY HOSTS

From the analysis of data movement on various ports, it has been found that a lot of hosts on different unassigned and assigned ports are useless. Top 15 most heavily visited hosts include:

- *stream.wmlivesvc.vitalstreamcdn.com*
- download.microsoft.com
- *www151.megaupload.com*
- *dl.search-download.org*
- *www.nakedwebtv.com*
- *dal-v10.dal.youtube.com*
- *www.chaltatv.com*
- software-files.download.com
- *www132.megaupload.com*
- *akmccvideos.metacafe.com*
- ardownload.adobe.com
- ejang.jang.com.pk
- pagead2.googlesyndication.com
- definitions.symantec.com
- *www.videosnstuff.com*

In the 15 hosts listed above, almost 9 hosts (*in italic*) are useless / filthy and are used for audio/video streaming and for uploading/downloading similar stuff. These hosts may be blocked to save the bandwidth.

5. 10 PROTOCOLS

In computer networks, a protocol is a set of rules that governs data communications. As a result of analysis of ISA logs, it has been found that the following protocols are commonly used for accessing the network. HTTP is

the mostly used protocol. FTP is the protocol used less than http SSL-Tunnel is the protocol which is used very rarely. Beside these, HTTPS is also being used, but its usage is negligible. Whereas, a hyphen "-" in the protocol field signifies the loopback address 127.0.0.1.

6. RELATED WORK

In [14] similar type of work have been done where the web server log of an academic institute have been analyzed. The obtained results can be helpful for variety of public or private applications. The usage of sever logs have also been analyzed in [16] for the decision making of the organization and the monitoring of user access modes. The research work in [15] describes the regular patterns of system features that describe program and the user behavior, large-scale log processing for analyzing the log records.

However, the study of log files of the ISA server for the ports, protocols and bandwidth usage factors is mainly missing in the literature. We, through this paper, have attempted to fill this gap.

7. RECOMMENDATIONS

Based on the analysis presented in this paper, following recommendations are drawn:

- Heavily used ports should be analyzed regularly for the identification of useless hosts. E.g. approximately 94% of the whole network traffic travels through port number 80 and out of this 94% traffic, 55% of data traffic is generated by useless hosts. If the port 80 is analyzed on regular basis, it would help us in saving a lot of network bandwidth.
- According to Internet Assigned Number Authority, "unassigned port numbers should not be used" [11].
- From the analysis of different ports traffic, it has been discovered that unassigned ports are normally used by streaming sites, which may be blocked. e.g., Port 436 & 8080.
- Some of the clients are given direct access to the ISA Server or in other words they are allowed to bypass server security. These users have no restriction on the type and amount of content they are browsing, hence may choke the network bandwidth and would remain unnoticed. Therefore it is strongly recommended that no one should be allowed to bypass the Symantec security.

The solution given in this research if implemented can be very helpful in analyzing different ports & protocols and identifying bandwidth hungry hosts & clients responsible for generating lot of network.

8. SUMMARY & CONCLUSIONS

ISA Server generates very detailed logs for all the web requests made through it, these logs are in the form of semi structured flat text files. Although these logs contain a lot of data which can be mined or analyzed to gain insight into the network usage trends and patterns but this precious data goes useless. Moreover ISA Server provides built-in logs analysis and reporting features, but with minimal customization options for users.

The solution suggested in this paper is capable of parsing and transforming these semi-structured logs and is capable of generating more customizable multidimensional reports that helps in better understanding and analyzing the network usage patterns & trends.

The paper analyses the ports and protocols of the ISA server being used by the users for accessing the network. As outlined in the recommendation section, some of the bandwidth hungry hosts consuming approximately 55% of the total bandwidth have been identified. Additionally it has been found that the unassigned ports are never used for any useful browsing, hence they may be blocked. Beside this, every user should not be blindly allowed to bypass the server security.

REFERENCES

[1] Arun Sen, Peter A. Dacin and Christos Pattichis, "Current trends in Web Data Analysis", Communications of the ACM, Vol.49, Issue 11, pp. 85 – 91, Nov 2006.

[2] Jia Hu; Ning Zhong; "Clickstram Log Acquisition with Web Farming", The 2005 IEEE/WIC/ACM International Conference on Web Intelligence; pp. 257 – 263; 19-22 Sept. 2005.

[3] I-Hsien Ting, Chris Kimble & Daniel Kudenko, "UBB Mining: Finding Unexpected Browsing

Behaviour in Clickstream Data to Improve a Web Site's Design", Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence, 2005.

[4] Dong-Ho Kim, Il Im & Vijayalakshmi Atluri, "A Clickstream-based Collaborative Filtering Recommendation Model for E-Commerce", Proceedings of the Seventh IEEE International Conference on E-Commerce Technology, 2005.

[5] Olfa Nasraoui, Osmar R. Zaïane, Myra Spiliopoulou, Bamshad Mobasher, Brij Masand & Philip S. Y, "Web Mining and Web Usage Analysis" 11th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2005.

[6] Montgomery, A. L. "Using Clickstream Data to Predict WWW Usage", WebShop, University of Maryland, June 2003.

[7] Alan L. Montgomery and Kannan Srinivasan, "Learning About Customers Without Asking", Nirmal Pal and Arvind Rangawamy (eds.), The Power of One - Leverage Value from Personalizaton Technologies, eBRC Press, Penn State University, 2003.

[8] SANS Institute, "Using ISA Server Logs to Interpret Network Traffic", SANS Institute 2002.

[9] Xiaohua Hu, Nick Cercone; "An OLAM framework for Web Usage Mining and Buiness Intelligence Reporting", IEEE, 2002.

[10] Jesper Anderson, Anders Giversen, Allan H. Jensen, Rune S. Larsen, Torben Bach Pedersen; "Analyzing Clickstreams using Subsessions", Proceedings of the 3rd ACM international workshop on Data warehousing and OLAP; pp. 25 – 32; 2000.

[11] http://www.iana.org/assignments/port-numbers.

[12] Behzad Mortazavi, "Discovering and Mining User WebPage Traversal Patterns," The Requirements for the Degree of Maser of Science in the School of Computing Science, Simon Fraser University, 2001.

[13] Amiya Kumar Tripathy, Siddharth Nimkar, Meenakshi Srivastava "SCAzer - Analyzing Web Content through Users' Access Patterns" International Journal of Recent Trends in Engineering,Vol. 1, Issue 1, 2009.

[14] Dilip Singh Sisodia, Shrish Verma, "Web usage Pattern Analysis Through Web Logs: A review", Ninth International Joint Conference on Computer Science and Software Engineering (JCSSE), 2012.

[15] Kazimierz Kowalski, Mohsen Beheshti, "Analysis of Log Files Intersections for Security Enhancement", Third International Conference on Information Technology: New Generations (ITNG'06), 2006.

[16] DeMin Dong, "Exploration on Web Usage Mining and Its Applications", 978-1-4244-3894-5, IEEE, 2009.

ISSN 1605-8607

DESIGN, CHALLENGES AND FUTURE RESEARCH ASPECTS OF MAC PROTOCOLS FOR WIRELESS SENSOR NETWORKS

Pardeep Kumar*, Akhtar H. Jalbani**, M. Ibrahim Channa***, Anees A. Soomro****, Iftikhar A. Koondhar*****

ABSTRACT

The most substantial challenge facing Wireless Sensor Networks (WSNs) is the requirement of significant reductions in energy consumption of sensor nodes. Size and cost limitations result in a very limited on-board power capacity and the dense, remote and inaccessible deployment of sensor nodes prevents recharging their batteries, making energy a scarce resource for WSNs. A medium access control (MAC) protocol for WSNs, apart from other responsibilities, can substantially conserve the energy of a node by controlling the functionality of radio, which is the major energy consuming component of the sensor node. In this paper, we first discuss the basics of MAC design for WSNs and present a set of important MAC attributes. Subsequently, we present the main categories of MAC protocols proposed for WSNs and highlight their strong and weak points. After briefly outlining different MAC protocols falling in each category, we provide a substantial comparison of these protocols for several parameters. Lastly, we envision future research directions on open issues in this field that have mostly been overlooked.

1. INTRODUCTION

The pervasiveness, self-autonomy and self-organization of low-cost, low-power and small-sized WSNs [1 - 6] have brought a new perspective to the world of wireless communication. They are emerging as an ideal candidate for several daily-life applications, especially in monitoring and controlling domains. The demands placed on such type of networks are expending exponentially with the increase in their dimensions. The development of new hard-ware, software and communication technology and continuous refinements of current approaches is also pushing this domain even further. However, all these unique characteristics along with the limited resources have made WSN a challenging network. Integrating sensing, processing and communication functionalities into a sensor node has added a lot of complexities. Moving from sensors with only few hours of life time to one with many years of life time demands several iterations of energy efficient techniques.

The MAC protocol for WSNs can save a substantial amount of energy by judiciously controlling the functionalities of radio, which is the most power-depleting component of resource-scarce nodes. The nodes of WSNs share a wireless medium and communicate with each other usually via multi-hop routes in a scattered, dense and rough sensor field. The MAC protocol also manages the shared-medium and creates a basic network infrastructure for nodes to communicate with each other. Thus it provides a self-organizing capability to the nodes and tries to enforce singularity in the network by letting the sender and receiver communicate with each other in a collision-free fashion. Therefore, the MAC protocol helps fulfilling important design objectives of WSNs by specifying how nodes employ the radio to conserve energy, share the channel and avoid collision in correlated and broadcasting environments.

1. 1 MAC RESPONSIBILITIES

The primal task of any MAC protocol, in general, is to govern the fair access of nodes on a shared medium in an efficient way in order to achieve good individual throughput and better channel utilization [7]. However, dense deployment of nodes, collaboration among nodes

*, ****, *****Assistant Professor, Department of CSE, Quaid-e-Awam University of Engineering, Science and Technology, Nawabshah, Sindh

**Associate Professor, Department of IT, Quaid-e-Awam University of Engineering, Science and Technology, Nawabshah, Sindh

*** Assistant Professor, Department of IT, Quaid-e-Awam University of Engineering, Science and Technology, Nawabshah, Sindh

rather than competition, dispersed applications, often decentralized control and volatile communication links make WSNs quite a different type of network. The responsibilities of a MAC protocol for WSNs therefore differ considerably as of the other wired or wireless networks. Fairness, higher channel utilization and individual throughput are not as important for a WSN MAC protocol as the energy conservation, collaboration and collective information among nodes.

Additionally and as per application requirement, provision of timeliness, reliability, scalability and non-synchronized operation may also play an important role in designing a MAC protocol for WSNs. Furthermore, an ideal MAC protocol ensures self-stabilization, graceful adaptation to topology and traffic changes, an acceptable delivery ratio and low overhead and error rate for a WSN.

2. MAC CHALLENGES FOR WSNs

The design of a MAC protocol for WSNs is challenging due to limited energy resources, low transmission ranges, compact hardware, event based network behavior, self-configuration and high redundancy factors. Moreover, by virtue of a wireless broadcast medium, WSNs inherit all the problems like interference, fading, path loss, attenuations, noise, and high error-rates [8 - 9]. Evidently, a lot of research work related to the MAC designing for WSNs revolves around energy efficiency. Other goals like latency, packet delivery ratio, and adaptation to traffic conditions and scalability are often traded-off in favor of energy efficiency. The main sources of energy waste in WSNs that a MAC protocol has to dealt with are idle listening, collision, overhearing, over-emitting and control packet overhead [10].

Along with energy efficiency, several WSN applications may need delay bound operations. However, unlike traditional distributed systems, the timeliness guarantee for WSNs is more challenging. They interact directly with the real world, where the physical events occur in an unpredictable manner with different traffic and delay requirements. Duty cycling, dynamic topology and limited memory and computation power also restrict the design space we could trade off.

As a result, many existing architectures and protocols for traditional wireless networks such as IEEE 802.11 and Blue-tooth are not suitable for WSNs as they usually target higher data rates with less emphasis on energy issues. Therefore, novel MAC approaches supporting special requirements of such challenging network are being developed.

3. COMMON MAC APPROACHES FOR WSNs

There is not any generic best MAC protocol for WSNs; the design choice mainly depends on the nature of the application. The stringent design requirements of a MAC protocol for WSNs can be met by a plethora of approaches. The most widely used approaches in designing such MAC protocols with their implications are outlined below.

3. 1 DUTY CYCLING

A WSN generally generates much less data traffic and sends very small data frames as compared to traditional wireless or wired networks. The sensor nodes therefore remain idle most of the time either waiting for their periodic turn to generate data or listening the idle channel for an event to occur. Since the radio consumes as much energy during idle listening as in receiving data packets, switching it in low-power sleep mode and waking up shortly to listen the medium can significantly conserve the energy of nodes. The sum of the sleep period and the listen period is called wake-up period, whereas the ratio of the listen period to the wake-up period is called duty cycle of the node.

Duty cycling significantly increases system lifetime of a dense WSN by reducing idle-listening and over-hearing. However, the selection of an optimal duty cycle value is not an easy task. Long sleep period may induce extra per-hop delay since a sending node needs to wait until the receiver wakes up and accepts the packets. More frequent switching of radio between sleep and awake modes, i.e., lower sleep phases also outweighs the benefits of duty cycling. Hence, the optimal selection of duty cycle value is a critical step towards achieving the desired system performance.

3.2 TOPOLOGY CONTROL

The aim of topology control is to build a reduced topology by dynamically changing transmitting range of

nodes. In this way energy of nodes can be saved while keeping the network connectivity and coverage intact [11]. As transceivers are one of the primary sources of energy consumption in WSNs, topology control reduces energy consumption by forcing packets to travel through multiple hops. The topology control mechanism reduces energy consumption by reducing collisions, contentions and exposed terminal problems. However, idle listening, overall latency, complexity and increased packet loss probability remain core issues with this mechanism.

3.3 SCHEDULING AND SYNCHRONIZATION

Several WSN MAC protocols assume that nodes follow a fix schedule to switch their radios between wake-up and sleep modes. They also assume that nodes are timely synchronized. However, in reality such time synchronization in dynamic and resource-limited WSNs is very difficult to achieve as it induces a lot of overhead and may need extra hardware and/or complex algorithms. Collisions and retransmissions increase dramatically if all nodes wake up simultaneously resulting in higher energy waste and packet delay. Therefore, it is wise to use the random and non-synchronized wake-up and sleep schedules for WSN nodes.

3.4 CROSS-LAYERING

Most of the time, the conventional layered architecture has been followed by the researchers to design WSN MAC protocols. With very limited resources available for WSNs, a trend of cross-layer designing is emerging in order to achieve aggregate optimization among different layers. Unlike layered networks, WSNs cannot afford significant layered overhead due to their limited energy, storage and processing capabilities. Moreover, application-aware communication and low-power radio considerations motivate for the cross-layer architecture for WSNs. Recent studies in [12 - 16] affirm improvement in WSN performance by using cross-layering.

3.5 TIMELINESS SUPPORT

No doubt, energy efficiency in WSNs has always remained a prominent objective of researcher. However, with ever increasing applications of WSNs in many diverse fields, new concept for offering timeliness related Quality of Service (QoS) is inevitable. Nevertheless, limited resources, low node reliability, dynamic network topology and direct interaction with the physical world make hard timeliness in WSNs very difficult. As a result, the probabilistic based soft timeliness guarantee in many applications of WSNs is mostly permissible.

4. CLASSIFICATION OF WSN MAC PROTOCOLS

The WSN MAC protocols can be classified depending on how nodes access the channel into the general categories of contention based, scheduling based, channel polling based, and hybrid protocols [17].

4.1 CONTENTION BASED MAC PROTOCOLS

With the contention-based Carrier Sense Multiple Access (CSMA) method, the transmitting node, before any transmission, first senses the carrier. If the carrier is found idle, it starts with its transmission, otherwise defers the transmission for some random time, which is usually determined by the back-off algorithm. Such MAC protocols consume less processing resources and are suitable for event-driven WSN applications. They are flexible to network scale and dynamics as no clustering and topology information is required. However with this approach, the transmission is purely handled by the sender and the problem of hidden- and exposed-terminal may occur resulting in collisions, overhearing, idle listening and less throughput. Moreover, in many proposals, authors consider that contention time of nodes is synchronized and based on a schedule, i.e., at each periodic interval, all neighboring nodes wake up simultaneously [18], [19], [20]. That would incur all the mentioned problems of synchronization in a WSN.

The well-known protocols working under this scheme are briefly outlined below.

(a) S-MAC: The design of the Sensor-MAC (S-MAC) [18] was one of the first attempts to significantly reduce idle listening; collisions and overhearing in WSNs by putting nodes in listen and sleep periods. S-MAC organizes the schedules of neighboring nodes by letting nodes share common listen periods according to a schedule. This requires formation and maintenance of synchronization among nodes. In order to reduce the hidden-terminal effect, S-MAC uses the RTS/CTS

handshake scheme. To minimize costly retransmissions, S-MAC fragments long messages into short frames and sent them in a burst. The RTS/CTS is only required before transmitting the first short frame.

S-MAC is rigid for a predefined set of workloads as there is no mean to adapt the length of listen and sleep periods with changing traffic conditions. As discussed earlier, the formation, maintenance and compliance of synchronization has serious consequences in WSNs. Longer and fixed sleep periods of S-MAC have serious impact on system latency. Moreover, the S-MAC nodes may follow more than one schedule, which results in higher energy consumption via idle listening and overhearing. With fragmentation in S-MAC, overhead and retransmission can be reduced, but it comes at the expense of unfairness since a node reserves the channel for a whole burst duration.

(b) T-MAC: The Time-out MAC (T-MAC) [19] protocol improves the energy efficiency of S-MAC by adaptively cutting down the listen period, which ends immediately if no activation event has occurred for a threshold period *TA*. The comparison between S-MAC and T-MAC shown in Figure 1 confirms this improvement. However, T-MAC could result in the early sleep for nodes, i.e., a node, specially the third hop one, goes to sleep mode when a neighbor still has messages for it. Though T-MAC saves more energy than S-MAC, it comes at the cost of reduced throughput, packet loss and higher latency. T-MAC also suffers from synchronization and scaling problems.

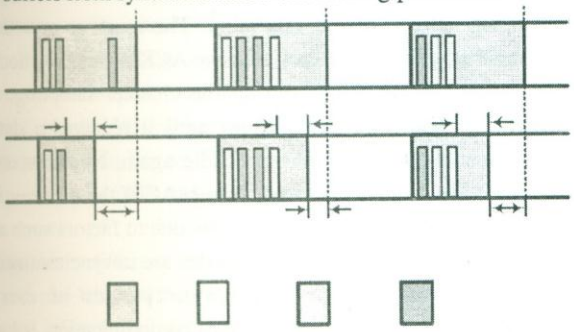


Figure 1: T-MAC vs. S-MAC

4.2 SCHEDULING BASED MAC PROTOCOLS

Scheduling based schemes assign collision-free links to each node in the neighborhood. The links may be assigned as time slots (TDMA), frequency bands (FDMA), or spread spectrum codes (CDMA). However, due to the complexities incurred with FDMA and CDMA schemes, TDMA schemes are preferred as scheduling methods for WSNs. A schedule in such schemes regulates which participant may use which resource at what time. A node can only access its allocated time slot and does not need any contention with its neighbors.

The cardinal advantages of scheduling based schemes include minimum collisions, less overhearing and implicitly avoidance of idle listening. They also provide a bounded and predictable end-to-end delay. However, the average queuing delay is much higher as a node has to wait for its allocated slot before accessing the channel. Overhead and extra traffic required in setting up and maintaining synchronization among nodes, no mean to adapt with varying traffic and topology conditions, reduced scalability, low throughput and no peer-to-peer communication are the major concerns with these schemes.

(a) Leach: A mostly scheduled based Low-Energy Adaptive Clustering Hierarchy (LEACH) protocol [21] divides a dense and homogeneous WSN into clusters each supervised by a cluster head. The cluster head is responsible for creating and maintaining TDMA schedules, communicating with its cluster members and forwarding the received messages to the sink node. LEACH also uses a randomized rotation mechanism for selecting a cluster head and tries to distribute the energy among nodes in an evenly manner. The cluster head of LEACH has to perform highly computational and energy consuming tasks and remains always awake, therefore the chances of a cluster head to die earlier are high. LEACH considers that all the cluster heads are within the range of the sink node. The lack of such multi-hop communication capabilities severely limits the network scalability. The channel under-utilization occurs with LEACH as it considers that nodes always have data to send during their allotted time.

(b) Trama: The TRaffic-Adaptive Medium Access (TRAMA) protocol [22] is mostly a TDMA based protocol. It schedules nodes on a distributive manner based on some traffic information. Each node computes its own priority and the priority of all its two-hop neighbors for each time slot. TRAMA uses neighborhood as well as schedule information to select the sender and

receivers for the current time slot, letting all other nodes to go to sleep mode. TRAMA also attempts to reuse slots that are not used by the selected transmitter.

The reuse of time slots and the utilization of neighborhood and traffic information are the positive features of TRAMA. The simulation results presented in [22] show higher percent-age of sleep time, less collision probability, and better data delivery with TRAMA as compared to S-MAC and IEEE 802.11. However, higher computation, large queuing delays and ineffective channel and memory utilization are the major concerns for a TRAMA based WSN.

4.3 CHANNEL POLLING BASED MAC PROTOCOLS

With the channel polling scheme, also known as preamble sampling or Low Power Listening (LPL), a sending node prefixes data packets with extra bytes called a preamble and sends it over the channel to ensure that the destination node would detect the radio activity and wake up before the actual payload is sent. On a wake-up, if radio activity is detected, the receiver turns on its radio, otherwise goes back to sleep mode until the next polling interval. To avoid deafness, a sender prefixes preamble at least as long as the check interval of the receiver to ensure that the receiver wakes up and performs channel sampling at least once while the preamble is being sent [23]. Figure 2

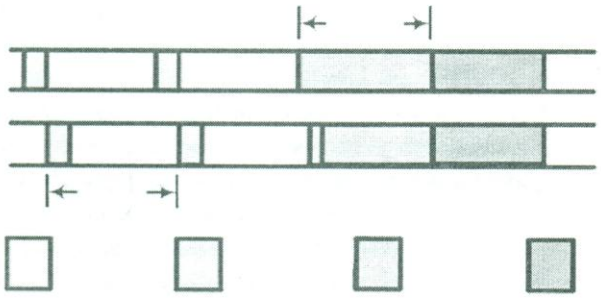


Figure 2: Channel Polling in WSNs

shows an example how channel polling works in WSNs. Channel polling based protocols do not need synchronization or clustering. Receivers consume significantly less energy as they wake up for very short period of time. However, the sender needs to send out a long and extended preamble before each data packet. Another issue is the limitation of the duty cycle value. Lowering the duty cycle extends the check interval. That is good from the receiver point of view but it significantly increases the transmission cost in the shape of long and extended preambles at the sender side. Therefore, this long preamble scheme results in unnecessary energy consumption both at the receiver and sender ends, overhearing at non-target receivers and excessive latency at each hop. These issues can be tackled by using short preambles, adaptive duty cycle values and by minimizing redundancy [24].

(a) B-MAC: Berkeley MAC (B-MAC) protocol [23] is one of the initial MAC protocols working on the traditional long preamble scheme, where the transmitting node precedes the data packet with a preamble that is slightly longer than the check interval of the receiver. On wake-up, if the node detects a preamble, it remains awake to receive the whole preamble. If the preamble is destined to this node, it further extends its wake-up time to receive the data packet; otherwise goes back to sleep mode. B-MAC uses an improved version of the Clear Channel Assessment (CCA) to determine whether the channel is clear. With an extended preamble, B-MAC reduces duty cycle and minimizes idle listening, particularly when there are no packet exchanges. It supports on-the-fly tuning of services by providing bidirectional interfaces to enable or disable services. However, B-MAC incurs all the mentioned problems of the long preamble scheme.

The WiseMAC [25] protocol tries to minimize the preamble length by letting a node learn about the awake periods of its neighbors. A sending node sends a preamble just before the receiving node wakes up and hence keeps the preamble length at minimum. The receiver puts the time of its next awake period in the ACK frame. If a node finds the medium busy during the channel sampling, it continues listening the medium until it receives a data packet or the medium becomes idle again. Nevertheless, over-emitting can occur with WiseMAC if the receiver is not ready at the end of the preamble due to factors such as interference or collision. Since nodes are unsynchronized, the transmitter has to keep awake periods of every neighbor and in case of a broadcast communication, it has to deliver the same packet many times to each neighbor. This redundant transmission leads to higher latency and energy consumption for nodes.

(b) AREA-MAC: The Asynchronous Real-time Energy-efficient and Adaptive MAC (AREA-MAC) protocol [24] re-visits the energy, timeliness, adaptability,

synchronization and redundancy problems of WSNs by using the short preamble technique. The sender sends out a stream of short preambles with a short spacing and turns its radio to receive mode in between in order to receive the pre-ACK from the intended receiver. The sender sends the data packet as soon as it receives the pre-ACK. This minimizes the chances of the data packet being dropped as the sender and receiver communicate via a 'pre-established' link. Each short preamble contains information such as the source address, destination address, sequence number and message type of the forthcoming data packet and the data-to-follow bit to inform the receiving node whether the sender has more data packet(s) waiting in its queue to be transmitted. Figure 3 shows the basic working of AREA-MAC protocol.

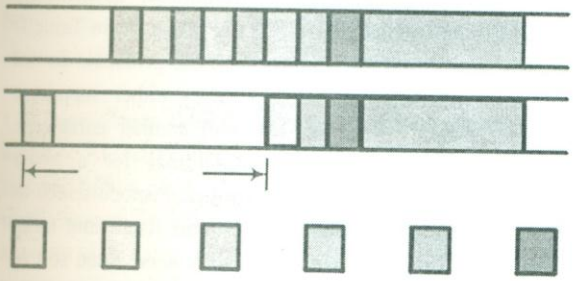


Figure 3: Basic working of AREA-MAC protocol

Short preambles unified with the pre-ACK mechanism (by using the source and destination addresses) impede several problems of the long preamble technique such as unnecessary energy consumption both at the receiver and sender ends, excessive latency at each hop, increased collision probabilities with the increased transmission and reception lengths, over-hearing at non-targeted receivers and bandwidth waste on the broadcast medium. The sequence number of the forthcoming data packet present in the preamble helps sensor nodes in reducing the number of redundant data packets. AREA-MAC considers different message types to support both normal and real-time traffic, which is indicated by the message type information of the preamble. The receiving node can also adapt its wake-up schedule in accordance with the sender that has enabled its data-to-follow bit, which further decreases packet latency and energy consumption of sensor nodes.

4.4 HYBRID MAC PROTOCOLS

Hybrid MAC protocols combine the strength of two or more different MAC schemes in order to achieve a joint improvement. They usually combine a synchronized protocol with an asynchronous one. Though such protocols aggregate the advantages of two or more schemes, they also carry, among other, scaling and increased complexity problems as they have to maintain two or more different working modes.

(a) IEEE 802.15.4: The IEEE 8021.5.4 standard [26] defines the PHY and MAC layers for a low cost, low power and low rate wireless personal area network (LR-WPAN). The standard supports both star and peer-to-peer operation with Fully and Reduced Function Devices (FFDs and RFDs). Together with ZigBee, which provides the upper (network and application) layers, 802.15.4 defines a full protocol stack suitable for several surveillance, home automation, health care, industrial and agricultural related applications of WSNs.

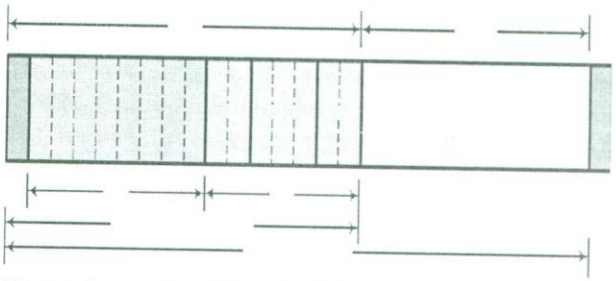


Figure 4: IEEE 802.15.4 MAC Superframe

The format of the MAC superframe is defined by the PAN coordinator (FFD) and is shown in Figure 4. The superframe is divided into 16 equally sized slots and the beacon is sent in the first slot to synchronize attached devices, to identify the PAN and to describe the overall structure of the superframe. The superframe can be further divided in two parts; Contention Access Period (CAP) and Contention Free Period (CFP). Any device wishing to communicate during the CAP period competes with other devices using a slotted CSMA/CA mechanism. For time and/or bandwidth critical applications, the PAN coordinator may dedicate contention-free TDMA-like Guaranteed Time Slots (GTS) portion of the CFP to devices. The combination of CSMA and TDMA makes 802.15.4 MAC a hybrid protocol.

Though 802.15.4 MAC protocol fulfils many of the WSN

Table 1: Comparison of different WSN MAC protocols for their support to different parameters

Protocol	Type	Energy eff.	Delay eff.	Async.	Adaptive	Scalable
S-MAC	CSMA	yes[1]	no	yes[2]	yes[1]	yes[1]
T-MAC	CSMA	yes[1]	no	yes[2]	yes	yes[1]
LEACH	TDMA	yes	no	no	no	no
TRAMA	TDMA	yes[1]	no	no	yes	no
B-MAC	Ch. Polling	yes[1]	no	yes	yes[1]	yes
WiseMAC	Ch. Polling	yes[1]	yes[1]	yes	yes	yes
AREA-MAC	Ch. Polling	yes	yes	yes	yes	yes
802.15.4	Hybrid	yes[1]	yes[1]	yes[3]	no	no
Z-MAC	Hybrid	yes	no	no	yes	no

[1] Suboptimal, can be improved [2] Uses common listen periods [3] Topology dependent

requirements, it endures the limitations, especially for timeliness, energy and bandwidth critical applications, and its performance can certainly be improved. All these limitations and their proposed solutions, along with the detailed working of 802.15.4 are elaborated in [28].

(b) Z-MAC: The Zebra MAC (Z-MAC) [29] protocol is a hybrid scheme that aggregates the positive aspects of TDMA and CSMA while offsetting their weaknesses. Z-MAC is a traffic adaptive protocol as it switches to CSMA under low traffic conditions in order to achieve high channel utilization and low delays. Under high traffic conditions it switches back to TDMA to achieve high channel utilization, fairness and fewer collisions. Unlike the traditional TDMA schemes, a node using Z-MAC can also utilize slots assigned to other nodes. Z-MAC provides a simple two-hop synchronization scheme where the transmitting node adapts the frequency based on its current data rate and resources. Hence, Z-MAC offers better performance under different time-varying channel/slot conditions and failures. However, during the start up phase, Z-MAC requires global time-synchronization, which certainly is a heavy burden for light-weight nodes. Complexity in maintaining both CSMA and TDMA modes, contention among nodes to gain access of the slots owned by other nodes, collisions and bandwidth under-utilizations are the other issues with Z-MAC.

5. COMPARISON OF DIFFERENT MAC PROTOCOLS

Having discussed main categories of accessing the channel in WSNs and different MAC protocols falling in each category, a comparison of these protocols for their support to energy efficiency, timeliness, asynchrony, adaptability and scalability factors is presented in Table I. It is clear from the table that most of the protocols do not support all of these parameters. They either support a parameter partially or trade-off with another parameter. Some of the protocols are suboptimal for a given parameter and can be improved further. Protocols such as S-MAC and T-MAC assume common listen and sleep periods for all their nodes, hence they need some sort of synchronization among nodes. As discussed, to create, manage and maintain synchronization in a resource scarce WSN is always a challenging task. The AREA-MAC protocol intends to address almost all the considered parameters in an efficient manner.

6. FURTURE RESEARCH DIRECTIONS

The research community has witnessed intense research related to the MAC design for WSNs over the last years. Various MAC protocols target different objectives and have different performance priorities. The most visible and prime consideration of almost all of them is the issue of energy efficiency. Other typical performance metrics like latency, adaptability to traffic and topology conditions, scalability, fairness, throughput and bandwidth utilization are mostly overlooked or dealt as secondary objectives due to the school of thought that nodes are densely deployed and they collaborate with each other rather than compete. However, with the advent of new Micro Electro-Mechanical Systems (MEMS) technologies and energy harvesting techniques and with the dynamic increase in several WSN related applications, the research community is set to experience many diverse

directions in this domain. For numerous applications of WSNs, some of these metrics may temporarily outweigh energy efficiency. Hence, energy consumption, which no doubt is the most critical parameter for WSNs performance, should not be the only focal point in designing a MAC protocol.

For medical urgency, surveillance, security, terrorist attacks, home automation, flood, fire and seismic detection applications, the provision of timeliness is as crucial as saving energy. For example; a sensor node embedded in an e-textile worn by patients should automatically but timely alert doctors or emergency services when a patient suffers from a severe disease. In this case, the importance of delivery ratio may also be increased as one disease could immediately result in other diseases and body sensors deployed on or near different body parts need to urgently inform medical personnel about the relative body part.

Last but not least, selecting a proper hardware scheme has a lasting impact on the performance of a MAC protocol and ultimately on a whole WSN system. Support for multi-channel hardware, balancing an appropriate memory size, usage of wake-up radio and selection of a packet- or bit-based radio are the important factors [10].

7. CONCLUSIONS

Although a lot of MAC protocols for WSNs have been proposed over the last years, the selection of a proper MAC protocol usually depends on the application. This paper manifests different factors and techniques that need definite consideration while designing a MAC protocol for WSNs. To cope with most of the energy related issues such as idle listening, collisions, overhearing, control packets and over-emitting, an eminent amount of research work supports duty cycling in WSNs. However, this duty cycling results in high latency and low throughput and a deep consideration is required to select a proper duty cycle value.

Along with the detailed working of each category and protocols falling in, their advantages as well as disadvantages are also discussed in detail. The channel polling based mechanism of MAC designing provides better energy consumption but the traditional long preamble schemes have several limitations that can be handled by using short preamble scheme. A substantial comparison of these protocols for different parameters is also presented in the paper. At the end, future research directions in designing a MAC protocol are envisioned, where we conclude that energy efficiency even though being the most critical metric is not sufficient to address. Factors like timeliness, scalability, asynchrony, delivery ratio and adaptability to traffic and topology changes may also need important considerations as per requirement of the application.

REFERENCES

[1] H. Karl and A. Willig, Protocols and Architectures for Wireless Sensor Networks. John Wiley and Sons Ltd, 2006.

[2] Pardeep Kumar & Mesut Günes, "Medium Access Control Protocols in Wireless Sensor Networks: Design Space, Challenges, and Future Directions" in 'Wireless Sensor Networks and Energy Efficiency: Protocols, Routing and Management' book, IGI Global Publisher, 2012.

[3] Pardeep Kumar & Mesut Günes, "MAC Protocols for Low-latency and Energy-efficient WSN Applications", Technical Rep. TR-B-08-18, Free University of Berlin, Germany, December 2008.

[4] S. Misra, I. Woungang, and S. C. Misra, Guide to Wireless Sensor Networks. Springer-Verlag London Limited, 2009.

[5] Pardeep Kumar, Mesut Günes, Qasim Mushtaq & Bastian Blywis, "A Real-Time and Energy-Efficient MAC Protocol for Wireless Sensor Networks", Sixth IEEE International Conference on Wireless and Optical Communications Networks, Cairo, Egypt, April 2009.

[6] M. Tubaishat and S. Madria, "Sensor networks: an overview," IEEE Potentials, Vol.22, Issue2, pp. 20–23, April 2003.

[7] M. A. Labrador and P. M. Wightman, Topology Control in Wireless Sensor Networks. Springer Science and Business Media B.V., 2009.

[8] Tse and P. Viswanath, Fundamentals of Wireless Communication. Cambridge University Press, 2005.

[9] V. Garg, Wireless Communications and Networking. Elsevier - Morgan Kaufmann Publishers, 2007.

[10] Bachir, M. Dohler, T. Watteyne, and K. Leung, "MAC essentials for wireless sensor networks," IEEE Communications Surveys and Tutorials, vol. 12, Issue 2, pp. 222–248, 2010.

[11] P. Santi, "Topology control in wireless ad hoc and sensor networks," ACM Computing Surveys, vol. 37, pp. 164–194, June 2005.

[12] T. Melodia, M. C. Vuran, and D. Pompili, "The State of the Art in Cross-layer Design for Wireless Sensor Networks," EuroNGI Workshops on Wireless and Mobility. Springer Lecture Notes in Computer Science 3883, July 2005.

[13] V. Kawadia and P. R. Kumar, "A Cautinary Perspective On Cross-Layer Design," in IEEE Wireless Communication Magazine, February 2005.

[14] J. Haapola, Z. Shelby, C. Pomalaza-Raez, and P. Mahonen, "Cross-Layer Energy Analysis of Multi-hop Wireless Sensor Networks," in EWSN'05, 33–44, 2005.

[15] L. V. Hoesel, T. Nieberg, J. Wu, and P. J. M. Havinga, "Prolonging the lifetime of wireless sensor networks by cross-layer interaction," in IEEE wireless communication, vol. 11, pp. 78–86, December 2004.

[16] P. Kumar, M. Gunes, Q. Mushtaq, and J. Schiller, "Performance evaluation of area-mac: A cross-layer perspective," in The Fifth International Conference on Mobile Computing and Ubiquitous Networking (ICMU 2010), (Seattle, USA), April 2010.

[17] K. Klues, G. Hackmann, O. Chipara, and C. Lu, "A component-based architecture for power-efficient media access control in wireless sensor networks," The 5th International Conference on Embedded Networked Sensor Systems (SenSys'07), November 2007.

[18] W. Ye, J. Heidemann, and D. Estrin, "Medium access control with coordinated adaptive sleeping for wireless sensor networks," IEEE/ACM Trans. Net., vol. 12, no. 3, pp. 493–506, June 2004.

[19] T. van Dam and K. Langendoen, "An adaptive energy-efficient mac protocol for wireless sensor networks," 1st ACM Conf. on Embedded Networked Sensor Systems (SenSys), pp. 171–180, November 2003.

[20] P. Lin, C. Qiao, and X. Wang, "Medium access control with a dynamic duty cycle for sensor networks," IEEE WCNC, vol. 3, pp. 1534–1539, March 2004.

[21] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor net-works," IEEE Transactions on Wireless Communications, vol. 1, no. 4, October 2002.

[22] V. Rajendran, K. Obraczka, and J. J. Garcia-Luna-Aceves, "Energy-efficient, collision-free medium access control for wireless sensor net-works," ACM SenSys, Los Angeles, CA, November 2003.

[23] J. Polastre, J. Hill, and D. Culler, "Versatile low power media access for wireless sensor networks," 2nd ACM Conf. on Embedded Networked Sensor Systems (SenSys 2004), Baltimore, MD, pp. 95–107, November 2004.

[24] P. Kumar, M. Günes, Q. Mushtaq, and B. Blywis, "A real-time and energy-efficient mac protocol for wireless sensor networks," Interna-tional Journal of Ultra Wideband Communications and Systems (IJUW-BCS), Vol. 1, Issue. 2, pp. 128 – 142, 2009.

[25] I.El-Hoiydi and J. Decotignie, "WiseMAC: An ultra low power mac protocol for the downlink of infrastructure wireless sensor networks," 9th IEEE Symposium on Computers and Communication, ISCC04, pp. 244–251, June 2004.

[26] "IEEE 802.15.4-2006: MAC and PHY specifications for LR-WPANs," http://ieee802.org/15/pub/TG4.html.

[27] "Zigbee" http://www.zigbee.org.

[28] P. Kumar, M. Günes, A. A. Almamou, and I. Hussain, "Enhancing IEEE 802.15.4 for low-latency, bandwidth and energy critical wsn applications," October 2008.

[29] Rhee, A. Warrier, M. Aia, and J. Min, "ZMAC:a hybrid MAC for wireless sensor networks," 3rd ACM Conf. on Embedded Networked Sensor Systems (SenSys'05), November 2005.

ISSN 1605-8607

COIN RECOGNITION BASED ON HEURISTIC SEGMENTATION AND IMAGE FUSION VIA BAYESIAN APPROACH

Akhtar H. Jalbani*, Pardeep Kumar**, M. Ibrahim Channa***, Chee Wai Mak****, Aness A.Soomro*****

Abstract

A simple algorithm that emulates the human ability and speed to learn and recognize coins and denominations of a foreign currency is presented. The use of color in hue, saturation and brightness indices to separate the coins from other objects / background and to determine the fusion of processed images results in a heuristic approach motivated by Bayesian learning framework. The complex process of coin detection and coin recognition reduces to a single threshold-limited image from which simple parameters are extracted and the value of the coins presented can be accurately determined. This method is applied to recognize Thai currency coins of 10-baht, 5-baht, 2-baht, 1-baht and 50-satang denominations and achieves 100% accuracy as compared to 44.29% using template matching in a trial of 14 test images acquired from a specific object distance.

Keywords: Bayesian approach, Bayesian learning framework, coin recognition, Hue-Saturation-Value, Histogram, Hough circle transform, image processing, pattern recognition, Thai coins, threshold.

1. INTRODUCTION

LECTRONIC payment methods may have increased in popularity in certain parts of the world in recent years, doing away with the need to carry "cash", but coins are still very much an integral part of life in many other parts of the world. In the metropolitan city of Bangkok, a foreign visitor often struggles with recognizing the value of the coins presented as some of the inscriptions on the coins are in Thai numerals (as shown in 0). While the conventional recognition by numeric is not possible, the human brain switches to visual differentiation methods. This subtle change often goes unnoticed but presents a powerful methodology that focuses on fundamental elements like color, shapes and size. The human brain quickly learns and remembers after a few occasions. Such methodologies are intriguing to study and emulate in machine learning.

Coin detection and coin recognition methods / systems are not new; many intelligent and complex systems have been researched and developed [1], using adaptive neural networks [3] and texton recognition methods [2]. The classification of coins based on parameters such as shape, size, surface design and weight was proposed in [1] for Indian coins. The use of color, particularly using special lighting matched to spatial properties of the camera was discussed in [10]. Yet another method that utilizes the HSV (hue, saturation and brightness) histograms to grade coins was presented in [11].

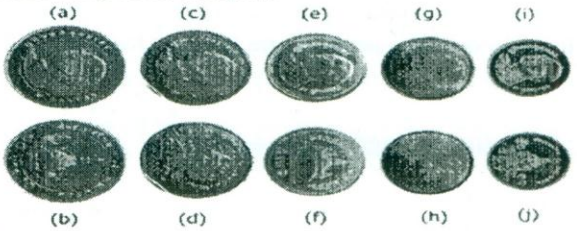


Figure 1: Thai coins with different denominations of interest; (a) head of 10-baht, (b) tail of 10-baht, (c) head of 5-baht, (d) tail of 5-baht, (e) head of 2-baht (type 1), (f) tail of 2-baht (type 1), (g) head of 1-baht, (h) tail of 1-baht, (i) head of 50-satang and (i) tail of 50-satang. Note: the 2-baht coin also comes in a silver colour version, while the 50-satang has a gold colour version.

*Associate Professor, Department of IT, Quaid-e-Awam University of Engineering, Science & Technology, Nawabshah, Sindh
,*** Assistant Professor, Department of CSE, Quaid-e-Awam University of Engineering, Science & Technology, Nawabshah, Sindh
***Assistant Professor, Department of IT, Quaid-e-Awam University of Engineering, Science & Technology, Nawabshah, Sindh
****Post graduate Student, Department of ISE, Asian Institute of Technology (AIT), Bangkok, Thailand

A multistage classifier based on eigenspaces capable of discriminating between hundreds of coin classes was presented in [7]. In the same, Bayesian fusion was also applied. A hybrid method of using eigenvalues of the covariance matrix, with circular hough transform and Bresenham's circle algorithm utilized with a raster scan algorithm that accurately pin-points position [12]. A gradient based method of recognizing coins using the discretized gradient direction went on to win the MUSCLE CIS Coin Competition 2006 [8]. Further methods for recognizing modern and Roman (ancient) coins were established in [9], where coin classification was accomplished with a simple k-Nearest Neighbor algorithm with k=5 after feature extraction. Many of these algorithms draw inspiration from more fundamental discussions on hough transform [5], histogram based methods [6], classification and segmentation techniques based on k-means [13 – 14]. A method of using R, G, B color together with area was proposed in [19] for single coin images.

With such a wealth of methods available, it is interesting to note that most of the methods focus on single coin, or extracted images of single coins at a time for detection and recognition. It is also noted that learning algorithms require hundreds of examples to allow the system to learn. This paper focuses on simplifying the learning process of differentiating the coins from the background (including other objects) so that processed images can be fused to form the optimum image where heuristic features can be extracted and evaluated to recognize coins accurately.

2. METHODOLOGY

A typical image used in this paper is shown in 0(a) below. Sample images do not necessarily have to contain all the denominations of interest. For ease of experimentation, the algorithms are implemented in OpenCV platform.

2.1 PRE-PROCESSING

The sample image of coins is first passed through a 7x7 Gaussian smoothing filter for noise reduction. The Gaussian smoothing operator is a 2-D convolution operator that is used to `blur' images and remove detail and noise. The weights are chosen according to the shape of a Gaussian function and the 2-D mask has the form:

$$G(x,y)=\frac{1}{2\pi\sigma^2}e^{\frac{x^2+y^2}{2\sigma^2}} \qquad (1)$$

In parallel, the grey-scale version of the image is passed through a Canny operation.

The smoothed image is converted from RGB space into HSV space. The H (hue), S (saturation) and V (brightness) planes are then extracted separately from the converted HSV image. The results obtained are shown in 0.

(a) Heuristic Threshold for Coin Segmentation

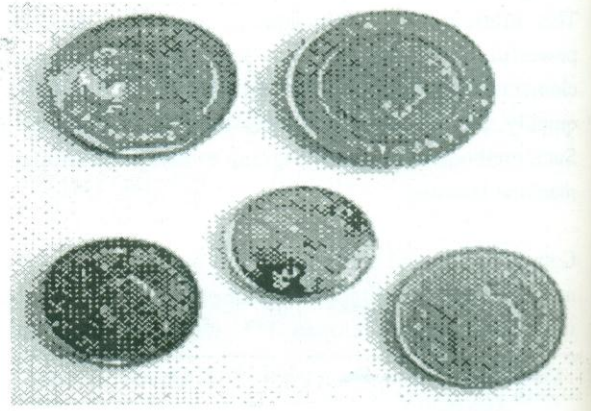
The primary task is to segment the image so that the coins become foreground and all other details become background. Motivated by the proposed method in [11], we turned to examining the histograms of the extracted H, S and V planes (0(d), (e) and (f) respectively). With an appropriate value selected (as marked approximately by the white arrow in 0), it is possible to set a threshold for each of the images to segment as many coins from the background as possible.

The apparent weakness of such a threshold scheme is over-reaction to "noise" in the histogram, which makes decisions difficult. The probability function of the coin in the H-plane is given by:

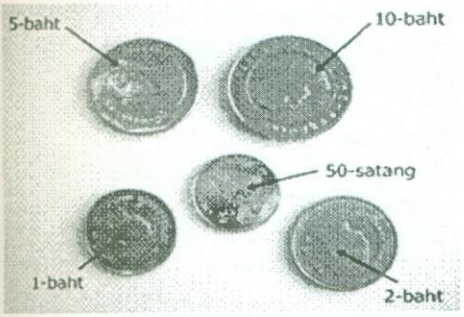
$$h(x;4,1)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-4)^2} \qquad (2)$$

The probability function of the background in the S-plane and V-planes are given by equations (3) and (4) respectively:

$$s(x;0,4)=\frac{1}{2\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x}{2}\right)^2} \qquad (3)$$



(a)



(b)

Figure 2: A typical image containing Thai coins of interest used in this paper; (a) shows the image of coins presented and (b) shows the image marked with denomination for reference.

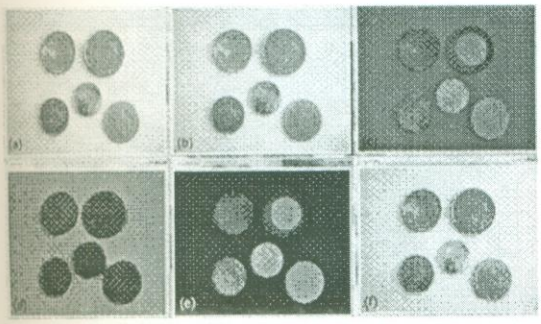


Figure 3: Pre-processing results obtained; (a) the original image as reference, (b) Gaussian smoothed image, (c) RGB image converted to HSV space and visualized, (d) extracted H (hue)-plane from HSV space, (e) extracted S (saturation)-plane from HSV space and (f) extracted V (brightness)-plane from HSV space.

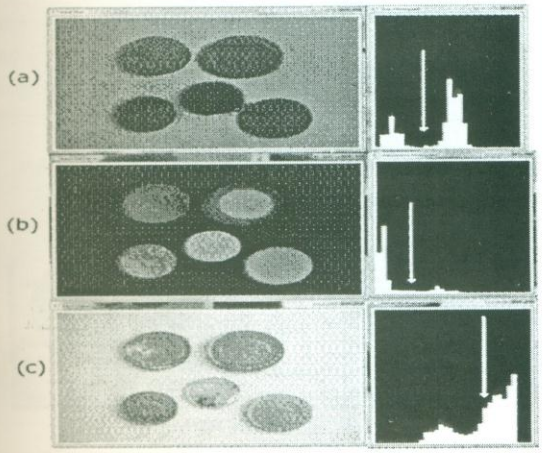


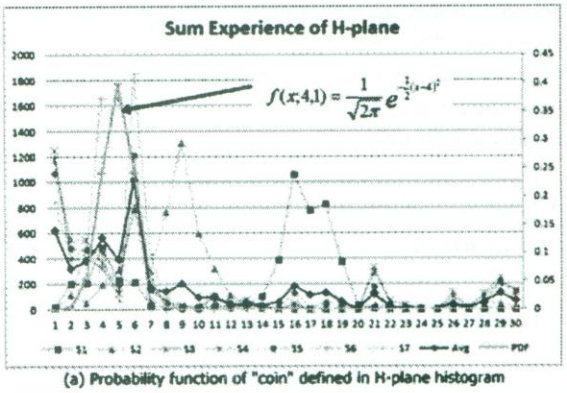
Figure 4: H-, S- and V-planes with respective histograms; (a) H-plane image with histogram, (b) S-plane image with histogram and (c) V-plane image with histogram. White arrow marks the approximate threshold position for each histogram to segment coins from background.

$$v(x;24,9) = \frac{1}{3\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-24}{3}\right)^2} \qquad (4)$$

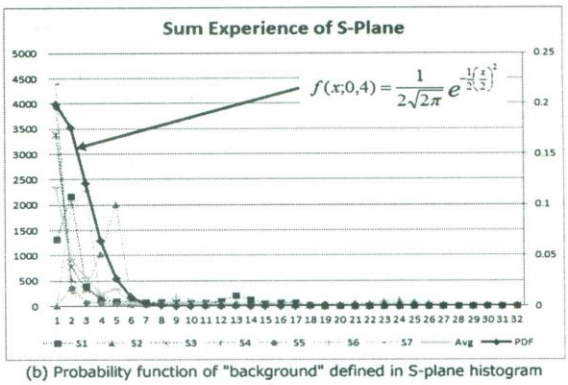
These probability functions can be derived from the heuristic behavior of histograms constructed from sample images in the H-, S- and V-planes, as shown in 0. Such histogram behavior is useful *prior knowledge* to the overall coin detection schema.

2.2 COIN DETECTION AND RECOGNITION

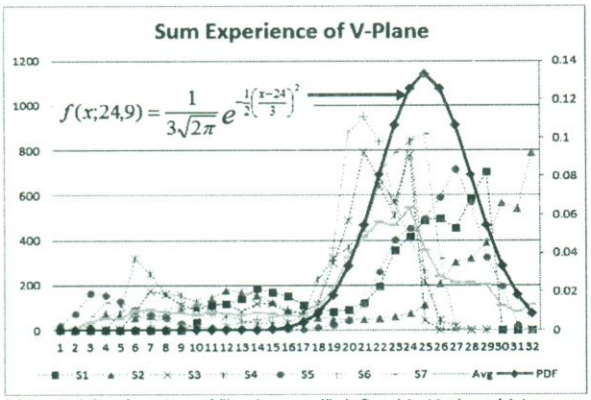
There are two important aspects of an object: shape and appearance [16]. We have defined Gaussian type probability density functions representing coins and background in H-, S- and V-planes in the prior section. It is important to recognize that a "coin" is a "round" object of certain "color" and "size", where "size" also corresponds to the "value". Where "round" is a property constituent of shape, "color" and "size" are property constituents of appearance.



(a) Probability function of "coin" defined in H-plane histogram



(b) Probability function of "background" defined in S-plane histogram



(c) Probability function of "background" defined in V-plane histogram

Figure 5: H-, S- and V-planes histograms characteristics; (a) probabilistic density function of "coin" casted in H-plane histogram, (b) probabilistic density function of "background" in S-plane histogram and (c) probabilistic density function of "background" in V-plane histogram.

(a) Shape

The simplest way to locate round shapes is to utilize the Hough Circle Transform [17]. The implementation via OpenCV first passes the image through a Canny edge detected. For every non-zero point, the local gradient is considered. The entire set of nonzero pixels in the edge image is considered for every candidate center. A circle with radius R and center (x_1, y_1) can be described with the parametric equations:

$$X = x_1 + R\cos\theta \quad (5)$$

$$Y = y_1 + R\sin\theta \quad (6)$$

When angle θ sweeps through 360°, the set of points (X,Y) traces the perimeter of the circle described.

The probability density function is given by:

$$P(\text{object is circle}) = \begin{cases} 1; & \text{iff object} \in \text{hough circle} \\ 0; & \text{otherwise} \end{cases} \quad (7)$$

The identification of circular objects via hough circle transform function is shown in 0.

(b) Size

Size can be determined either by the length of the perimeter of the circle traced by the hough circle transform function or determined by the area encircled by the perimeter of the object found. In this paper, the blob area function in OpenCV was used to determine the area of the object.

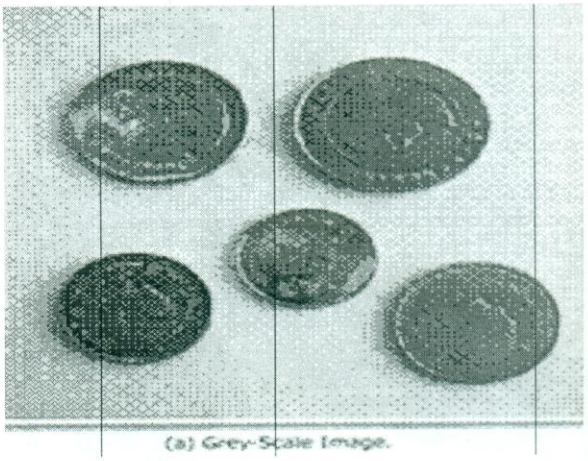
Intuitively, the "size" of the coin corresponds to the denomination or "value" of the coin. The distribution of size for different denominations as measured by the blob area function is examined, as shown in 0. A Gaussian type probability density function can be setup for each of the denominations of the coins of interest as the variation arises from the measurement itself with an error estimate that is Gaussian. The sizes of the coins are fixed; table 1 shows the relationship of sizes between denominations.

Table 1: Size relationship between Coin Denominations

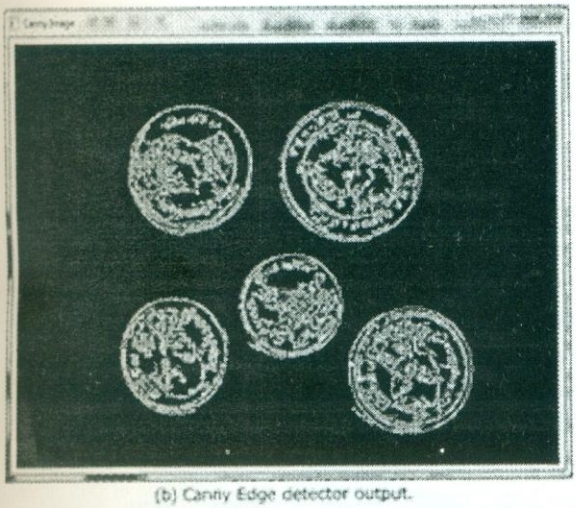
	10-baht	5-baht	2-baht	1-baht	50-satang
10 baht	1.00	1.22	1.40	1.70	2.08
5 baht	0.82	1.00	1.15	1.39	1.70
2 baht	0.71	0.87	1.00	1.21	1.48
1 baht	0.59	0.72	0.83	1.00	1.22
50 satang	0.48	0.59	0.68	0.82	1.00

(c) Coin Detection

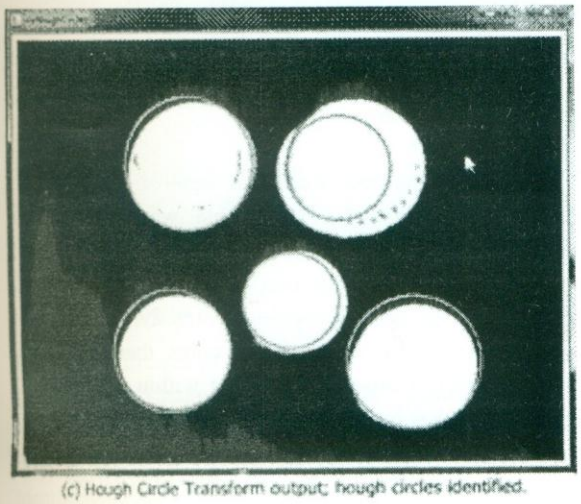
The motivation to construct probability density function for property descriptors of the coin (that is, a "round" object of a certain "color" and "size", where "size" corresponds to the denomination value) in shape and size, we explore the use of Bayesian methods. Bayesian methods allow us to use prior knowledge about the objects [15][16].



(a) Grey-Scale Image.

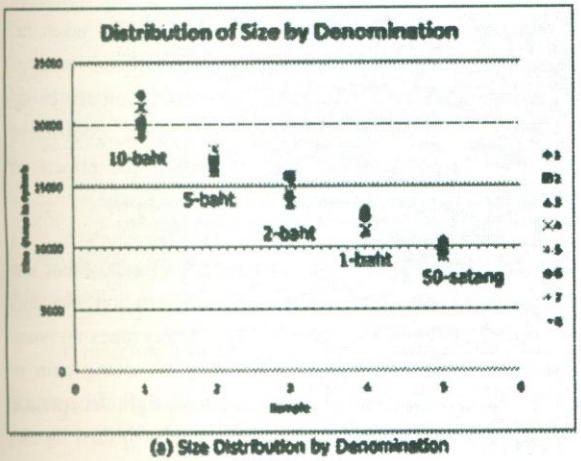


(b) Canny Edge detector output.

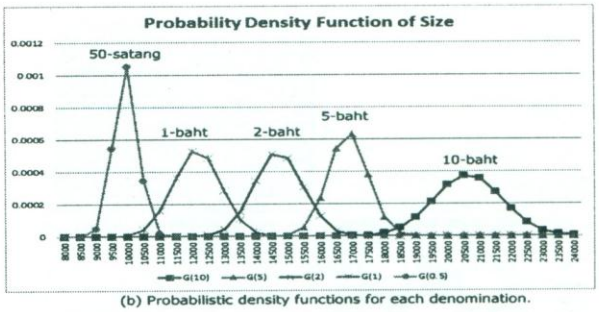


(c) Hough Circle Transform output; hough circles identified.

Figure 6: Identification of circular objects; (a) grey-scale version of original image, (b) Canny edge detector output and (c) identified circles from hough circle transform function.



(a) Size Distribution by Denomination



(b) Probabilistic density functions for each denomination.

Figure 7: Size distribution of different denominations studied in this paper; (a) distribution by sample and (b) probability function constructed from sample distribution.

The combination of the probabilistic evaluation of each of the object properties yields the final decision for the coin recognition. Implemented in a decision tree format, each of the probabilistic outcomes is accumulated sequentially in a manner well practiced by humans. For example, a "10-baht" coin is a "round object" of "a certain color that is different from the background" and "size that is between 19000 and 22000 pixels". Mathematically, it is given by:

$$P(10\text{-}baht) = P(object\ is\ circle) \bullet h(x) \bullet [1 - s(x)] \bullet [1 - v(x)] \bullet size(x) \qquad (8)$$

Using this Bayesian approach, the concept of "a certain color that is different color that is different from the background" is realized by the fusion of resultant binary images from apply appropriate thresholds previous derived from the probability density functions in the H-, S- and V-planes discussed previously.

3. Results & Discussions

The proposed method is applied to several images containing coin and non-coin objects and also of different backgrounds. In non-coin objects, we applied the method to round and irregular shaped objects. To test the robustness of the proposed method, we used round, non-coin objects of similar size as the coins.

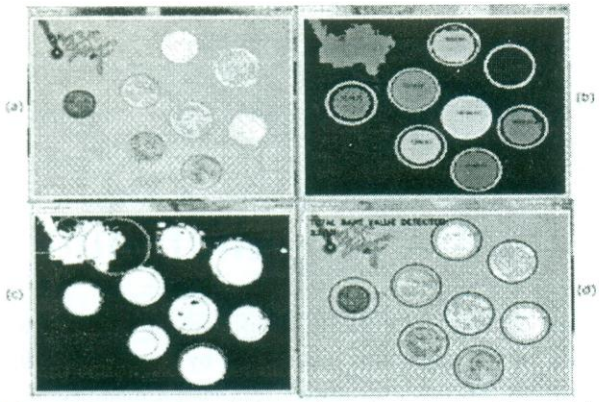


Figure 8: Image containing coin and non-coin objects; (a) original image, (b) blob area function, (c) hough circle transform and (d) coin identification and recognition mapped to original.

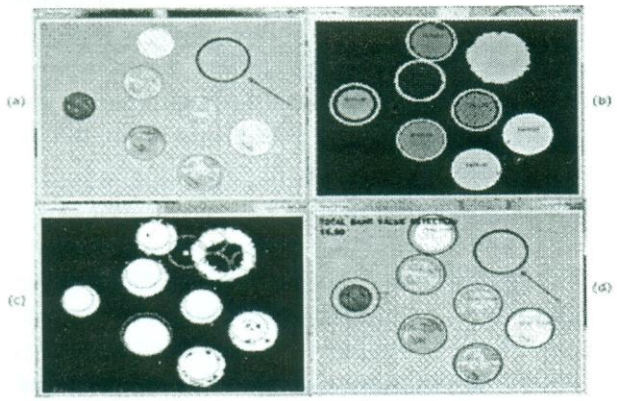


Figure 9: Image containing coin and round, non-coin object; (a) original image, (b) blob area function, (c) hough circle transform and (d) coin identification and recognition mapped to original.

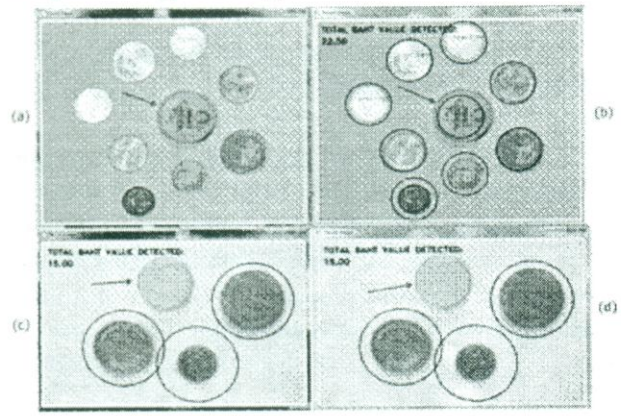


Figure 10: Images containing coin and round, non-coin object; (a) original image 1, (b) identification and recognition result of image 1, (c) original image 2 and (d) identification and recognition result of image 2.

3.1 COIN VS NON-COIN OBJECTS

As shown in 0(a)-(d), the proposed method is able to identify the coins of interest and also recognize the value correctly. The proposed method left out the irregular object placed in the image.

Another test image consisted of a round, non-coin object of similar size to the coins. Results are shown in 0; the round "rubber-band" was not selected by the algorithm. The values were correctly identified. Similarly, the same result is achieved with two other test images consisting of round, non-coin objects, as shown in Figure 10 In particular, the image in 0(c)-(d) contained a coin (of 25-satang denomination) that is not in the scope of this paper and the proposed method was able to discriminate this coin.

3.2 BACKGROUNDS

The proposed method was also applied to different backgrounds. Firstly a stone-table was chosen and secondly, the coins were placed on the palm of one's hand, as shown in 0. In both cases, the proposed method identified and recognized the coins correctly. Incidentally. Figure 11 (a) - (b) contained 3 coins there are of a denomination (25-satang) that is not within the scope of this paper and has been successfully discriminated.

3.3 COMPARISON WITH TEMPLATE MATCHING

The proposed method is compared with a template matching technique discussed in Chapter 7 of [18]. Template matching in OpenCV takes a template image and slides it over the image of interest and evaluates "correlation" with a specified method. We take the normalized versions, namely (i) normalized square-difference method, (ii) normalized correlation matching method and (iii) normalized correlation coefficient matching method as they help to reduce the effects of lighting differences between template and image.

The advantage of template matching method is that the matching techniques are invariant to rotation (0) and features (0). The invariance to features becomes an issue on accuracy, as shown in 0, where a 50-satang coin is found as a match to a 1-baht template; a tight acceptance criteria needs to be set.

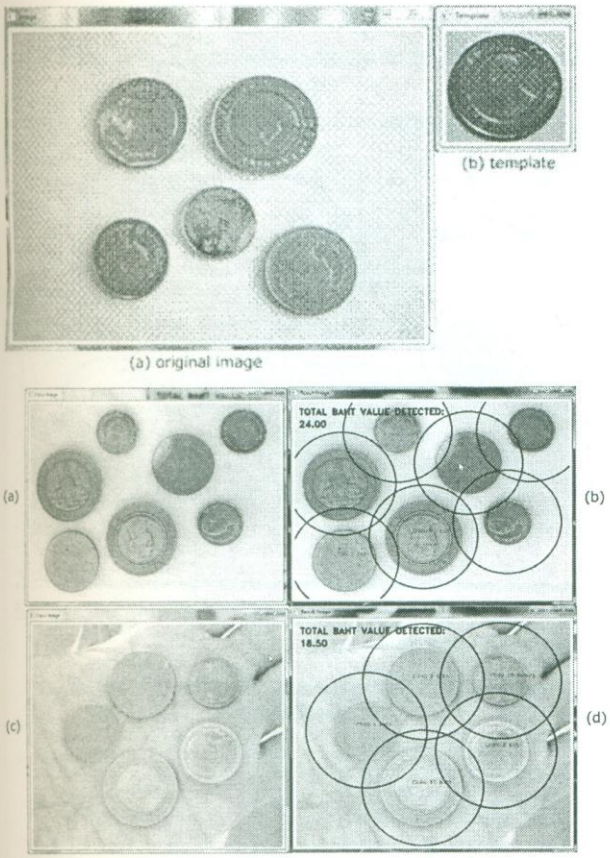


Figure 11: Images of coins on a stone-table background; (a) original image of coins on stone-bench, (b) identification and recognition result of image (a), (c) original image of coins on palm of hand and (d) identification and recognition result of image (c).

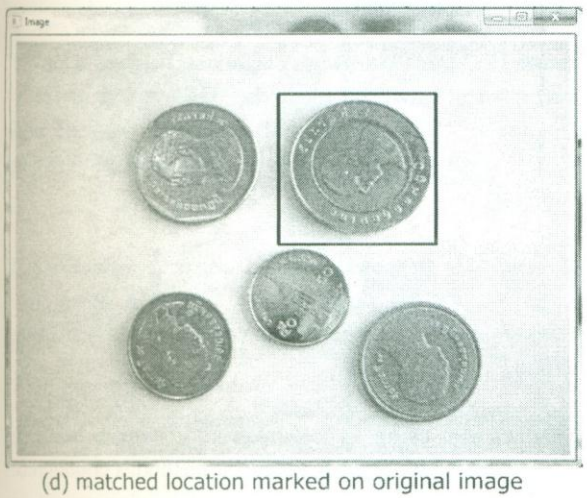


Figure 12: Results using template matching; (a) original image with (b) the template, (c) showing visualization of the matching results where the arrows mark the location of "best match" and (d) shows the "best match" marked on the original image.

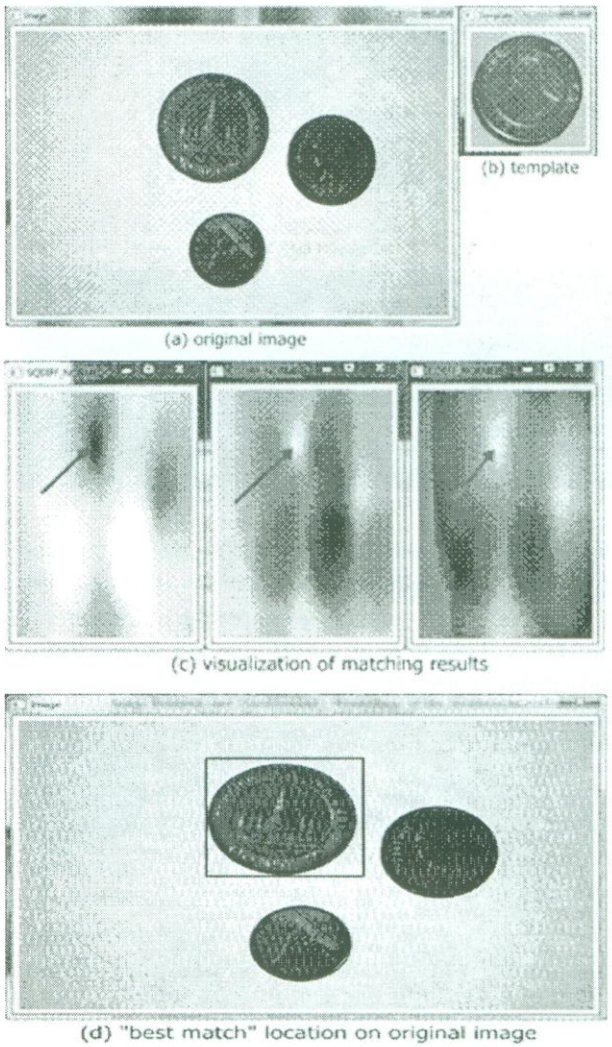
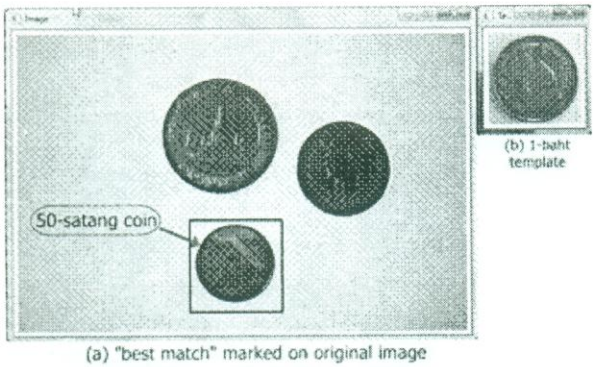
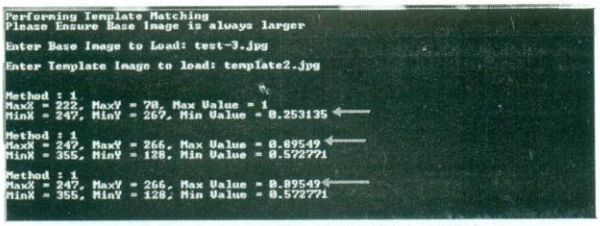


Figure 13: Results using template matching where template has different features than target in image; (a) original image with (b) the template, (c) showing visualization of the matching results where the arrows mark the location of "best match" and (d) shows the "best match" marked on the original image.

In another test image comparison, template matching method was unable to find a match due to size and/or color (0(a)); the best match was rejected as the values were outside the acceptance range. With a re-size template, the best match values showed an acceptance (99.99% confidence) but the match found the incorrect coin (0 (c)-(d)).

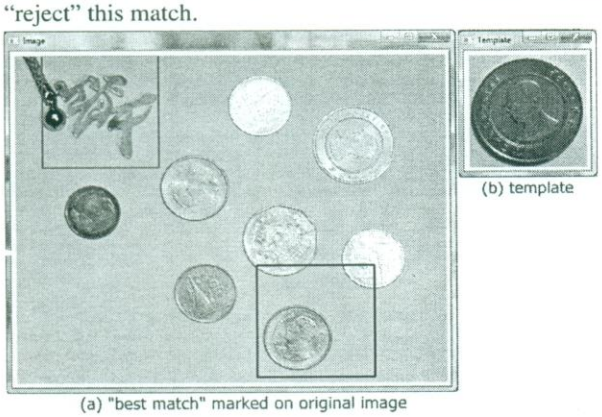


(a) "best match" marked on original image

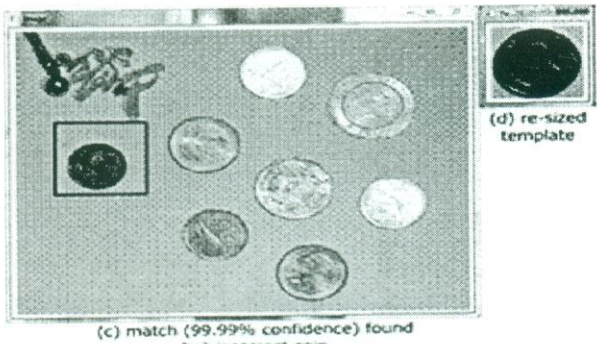


(c) acceptance criteria on min/max values need to be set

Figure 14: Incorrectly recognized coin; (a) a 50-satang coin in an image being found as a match to (b) a 1-baht template, therefore giving false results. The match criteria (c) needs to have a tight acceptance limit set to "reject" this match.



(a) "best match" marked on original image



(c) match (99.99% confidence) found but incorrect coin

Figure 15: Incorrectly recognized coin; (a) matches found that were rejected based on acceptance criteria with the template shown in (b). (c) shows the incorrect match still with (d) resized template (match accepted at 99.99% confidence).

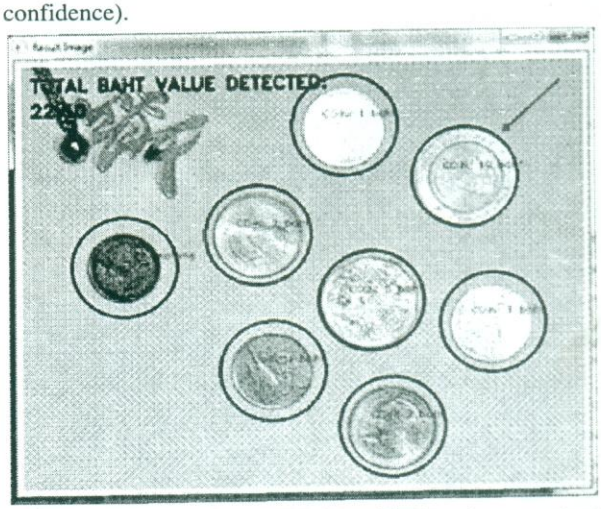


Figure 16: Coins correctly identified and recognized using our proposed method.

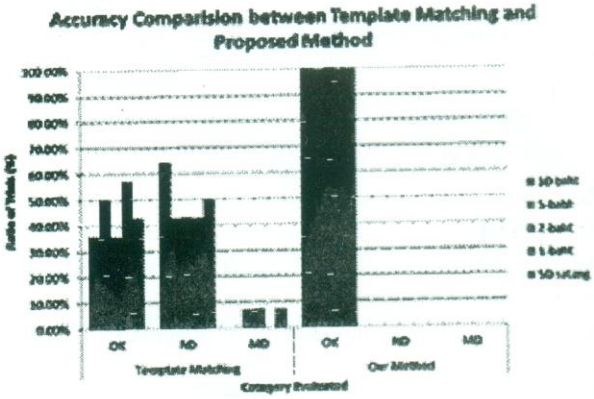


Figure 17: Graph showing comparison between template matching and proposed method. "OK" means correctly recognized, "ND" means not detected/recognized and "MD" means mis-recognized (recognized wrong denomination)

Whereas, our proposed method correctly detected and recognized all the coins in the same test image (0), showing superior performance over template matching method.

A further trial using 14 test images to compare between template matching and our proposed method is conducted. The brief algorithm for each method is summarized in table 2. A summary of accuracy comparison between template matching and our proposed method conducted

using the same test images is shown in 0. In the comparison trial, template matching method achieved only an overall accuracy of 44.29% (31/70 coin-presences detected and recognized correctly) while our method achieved 100% accuracy (70/70 coin-presences detected and recognized correctly). The trial confirms the superior performance of our method over template matching.

Table 2: Summary of Algorithm

Process	Template Matching	Proposed Method
Pre-Processing	1. Gaussian Smoothing (noise reduction)	1. Gaussian Smoothing (noise reduction) 2. Convert RGB image to HSV image 3. Split HSV image into 3 separate images; H-, S- and V-planes 4. H-, S- and V-planes threshold at limits based on Gaussian probability density function on histograms. 5. Fusion of H-, S- and V-images based on behavior of PDF determined in step (4). Result is a binary image.
Coin Detection & Recognition	1. Load 10-baht template. 2. Apply template matching using normalized square-difference, normalized correlation and normalized coefficient matching algorithms. 3. From each algorithm, determine location of "best match" and compare "confidence" with limit of 95%. If match found, go to step 4. If nothing accepted, load next and repeat. If nothing accepted after all templates used, skip to step 5. 4. Remove matched ROI from original image and repeat from Step 1. Accumulate count for coin detected and recognized. 5. Repeat for 5-baht, 2-baht, 1-baht and 50-satang template sets.	1. Subject fused image to blob detection and blob area function. 2. Subject fused image to hough circle transform. 3. For each blob detected, cross check with hough circle list. If matched, assign coin value based on area and go to step 4. If not matched go to step 5. 4. Accumulate count for coin identified and recognized. 5. Repeat step 3-4 for all blobs detected in image.

4. Conclusions

In this paper, a simple approach of presenting a fused image (consisting of binary H, S, and V representation of coins in the presence of background and other objects) to a Hough Circle Transform and Blob Detection algorithm for coin recognition is presented. The decision of threshold values employed and the choice of images selected for fusion with a Bayesian approach applied on heuristic trends exhibited by coins in the presence of other objects and background. This method achieved superior performance with an accuracy of 100% compared to 44.29% using template matching in a trial conducted with 14 test images.

Future work will focus on the ability to differentiate coin denominations from currencies mixed into the same image.

Acknowledgment

We wish to thank Western Digital for supporting the Master scholarship and AIT-School of Engineering and

Technology for providing the learning opportunity on this project.

REFERENCES

[1] M. Velu, P. Vivekanandan, "India Coin Recognition System of Image Segmentation by Heuristic Approach and Hough Transform (HT)", Int. J. Open Problems Compt. Math., Vol. 2, Issue. 2, June 2009.

[2] Dong-Shin Kim, Young Gyun Yun, You-Kyung Cha, "Coin Detector", CS7495/4495 Term Project, SSIP6, 14th Summer School on Image Processing, Szeged, Hungary 2006, [ONLINE] available at URL:www.inf.u-szeged.hu/ssip/2006/ projects/team6/ coin.pdf

[3] Shatrughan Modi, Seema Bawa, "Automated Coin Recognition System using ANN", International Journal of Computer Application (0975-8887), vol. 26, Issue 4,pp. 0975-8887, July 2011.

[4] M. Nölle, H. Penz, M. Rubik, K. Mayer, I. Holländer and R. Granec, "Dagobert – A new Coin Recognition and Sorting System", Proc. VIIth Digital Image Computing: Techniques and Applications, Sun C., Talbot H., Ourselin S. and Adriaansen T. (Eds.), 10-12 Dec. 2003, Sydney.

[5] H. Ballard, "Generalizing the hough transform to detect arbitrary shapes", Pattern Recognition, 31-2, 1981.

[6] S. Siggelkow and H. Burkhardt, "Improvement of Histogram-Based Image Retrieval and Classification", Proceedings of the International Conference on Pattern Recognition 3, 367-370, 2002.

[7] R. Haber, H. Rosmoser, K. Mayer, H. Penz, M. Rubik, "Classification of Coins using an eigenspace approach", Pattern Recognition Letters, Vol. 26, Issue 6, pp. 61-75, 2005.

[8] Marco Reisert, Olaf Ronneberger, and Hans Burkhardt, "An Efficient Gradient Based Registration Technique for Coin Recognition", Proceedings of the MUSCLE CIS Coin Competition Workshop, Berlin, Germany, 11 September 2006, [ONLINE] available at URL: http://muscle.prip.tuwien.ac.at/coin_workshop2006 _proceedings/reisert.pdf.

[9] Martin Kampel, Maia Zaharieva, "Optical Recognition of Modern and Roman Coins", Layers of Perception, Proceedings of the 35th International Conference on Computer Applications and Quantitative Methods in Archeology (CAA), Berlin, Germany, 2007.

[10] M. Ademeck, M. Hossfeld, M. Eich, "Three colour selective stereo gradient method for fast topography recognition of metallic surfaces", Proceedings of Electronic Imaging, Science and Technology (Martin A. Hunt and Jeffrey R. Price .eds), Machine Vision Applications in Industrial Inspection XI, vol. SPIE 5011, pp. 128-139, January 2003.

[11] Rick Bassett, Ping Gallivan, Xiang Guo, Eric Heinen, Akarsh Sakalaspur, "Development of an Automated Coin Grader: A Progress Report", Proceedings of MASPLAS'02, the Mid-Atlantic Student Workshop on Programming Languages and Systems, April 2002.

[12] J. Prakash, K. Rajesh, "A Novel Approach for Coin Identification using Eigenvalues of Covariance Matrix, Hough Transform and Raster Scan Algorithm", World Academy of Science, Engineering and Technology 44, 2008.

[13] R. M. Haralick and L. G. Shapiro, "Survey – Image segmentation techniques", Computer Vision Graphics and Image Processing vol. 29, pp. 100-132, 1985.

[14] M.K. Ng, "A note on K-means Algorithm", Pattern Recognition, vol. 33, pp. 515-519, 2000.

[15] Moshe Blank, Ita Lifshitz, "A Bayesian Approach to Recognition", Advanced Topics in Computer Vision, Faculty of Mathematics and Computer Science, [ONLINE] available at URL: http://www.wisdom.weizmann.ac.il/~vision/course

s/2004_2/files/bayessian_recognition.pps, April 2004.

[16] Fei Fei Li, Robert Fergus, Pietro Perona, "A Bayesian Approach to Unsupervised One-Shot Learning of Object Categories", Proceedings, 9th IEEE International Conference on Computer Vision (ICCV 2003), France, 2003.

[17] Kimme, D. H. Ballard, J. Sklansky, "Finding circles by an array of accumulators", Communications of the Association for Computing Machinery 18, pp. 120-122, 1975.

[18] Gary Bradski, Adrian Kaehler, "Learning OpenCV", O'Reilly Media, Inc., 2008.

[19] H. R. Al-Zoubi, "Efficient coin recognition using a statistical approach", 2010 IEEE Conference on Electro/Information Technology (EIT), pp. 1-5, 2010.

ISSN 1605-8607

END-OF-PIPE WASTE WATER DISPOSAL AND QUALITY COMPARISON OF OLD AND NEW SUGAR MILLS IN SINDH

Abdul Rehman Memon*, Niaz Ahmed Memon**, Muddasar Habib***

ABSTRACT

Sugar industry is 3rd largest industry of Pakistan as well as an important player in foreign exchange earnings; but it is also a major contributor to environmental pollution particularly with its wastewater discharges. This assessment study is the first part of the research project "Recurring environmental problems caused by sugar mills and their practicable solution", whereby wastewater handling in terms of its disposal by sugar mills is looked into so as to analyze their implications on the environment. This baseline study will help us assess the severity of the problem as a result of untreated wastewater disposal into the environment, which can lead us to devise a possible treatment plan that is low-cost as well as conducive to local conditions. Towards this end, three sugar mills viz. Matiari sugar mills, Matiari, Fauji sugar mills, Tando Mohammad Khan and Habib sugar mills, Nawabshah were frequently visited along with their immediate vicinities for the collection of samples and for carrying out investigation surveys as regards effluent disposal methods adopted by the concerned sugar mills. The results of this study suggested that sugar mills dispose of its wastewater untreated in every possible way, the pollutional composition of which exceed the limits, as specified in national environmental quality standards, by multiple times with respect to parameters such as Biochemical oxygen demand, Chemical oxygen demand, Oil and greases and Total suspended solids. Our research investigation into this issue concluded that sugar mills wastewater disposal in any given mode had a negative impact on land and water resources ultimately affecting the health and safety of people as well as flora and fauna.

Keywords: Pakistan sugar industry; Baseline assessment; Environmental impact; Wastewater disposal

1. INTRODUCTION

Pakistan is ranked 11th in the world in terms of sugar production and 8th largest sugar-consumer in the world [1]. Sugar industry is the 3rd largest industry of Pakistan, which is in operation for the last 65 years [2]. It is a seasonal industry operating for maximum of 5-6 months in one season from November to April. The industry uses either sugarcane or sugar beet as their raw material along with various chemicals added to increase the face value of the final product [3]. Sugar industry is a large water consumer and there is no stage in sugar production where water in some quantity is not required. However, water consumption can vary due to the technology applied and the nature and quality of raw material used. Mostly water is required in sugar mills as cooling water for barometric condensers, as boiler feed water, for lime preparation, for dilution in evaporators, etc. It has been observed that each ton of sugarcane crushed containing 70% moisture generates 0.7 m^3 of water if sugar and water are completely separated [4]. According to Indian standards, water consumption varies from 1.3 to 4.36 m^3 per ton of sugarcane crushed [5]. Thus, during sugar manufacturing process, huge amount of water is used on daily basis and as a result the industry generates wastewater or process liquid effluent with as much proportions. Sugar mills generate wastewater from each section of a sugar mill, which is mostly disposed of as a combined effluent for its outside settlement. There are three sections in a sugar factory: Mill house, process house and boiler house. The mill house wastewater is usually rich in oil and greases likely due to the spillage of oil and grease on the floor of mill house from the machinery and equipment that is washed off during floor washing. Process wastewater mainly results from floor and equipment washing and is highly contaminated with additives and other chemicals

* Assistant Professor, Department of Chemical Engineering, Mehran University of Engineering and Technology, Jamshoro
** Professor, Department of Civil Engineering, Quaid-e-Awam University of Engineering Sciences and Technology, Nawabshah
*** Assistant Professor, Department of Chemical Engineering, University of Engineering & Technology, Peshawar

used at different processing stages, while boiler house chiefly contributes to air pollution problem with its gaseous and particulate emissions and has little share in water pollution. Literature suggests that wastewater generated is about 20% of the water requirement [5]. Based on this ratio, a sugar mill with crushing capacity of 4500 tonnes of cane per day require 9000 m^3/day of water with 1: 2 ratio and hence the mills generate the wastewater in the range of 1800 m^3/day. The sugar mill wastewater is characterized by its brown color, burnt-sugar like odor, high temperature, low pH, high ash or solid residues and contains high percentage of dissolved organic and inorganic matter of which 50% may be present as reducing sugars [3]. In addition, sugar mill effluents carry the constituents such as Biochemical oxygen demand (BOD), Chemical oxygen demand (COD) and oil and greases (OG) in the range, which more often than not exceed the limits as given by national environmental quality standards (NEQS).

Sugar mills' wastewater is mostly disposed of as untreated in the outside environment subject to the mode of disposal adopted. Every sugar mill has its own way to dispense with the wastewater that it generates. In order to investigate this issue, three sugar mills comprising of one new (Matiari sugar mills, Matiari) and two old plants (Fauji sugar mills, Tando Mohammad Khan and Habib sugar mills, Nawabshah) were made the focus of this study in terms of wastewater characteristics along with its disposal mode into the environment.

2. METHODOLOGY

2.1 WASTEWATER DISPOSAL BY SUGAR MILLS

For comparative analysis of old and new plants of sugar mills in terms of wastewater disposal and their characteristics, three sugar mills viz Fauji Sugar mills (FSM), Tando Mohammad Khan, Habib sugar mills (HSM), Nawabshah (both of them old plants) and Matiari Sugar mills (MSM), Matiari (new plant) were frequented for investigation of disposal methods adopted by these mills' along with the collection of wastewater samples for their analyses. Wastewater disposal mode adopted by these sugar mills is discussed hereunder.

2.2 HABIB SUGAR MILLS, NAWABSHAH (HSM)

The factory is a private plant located in Nawabshah at about 125 km from Hyderabad. The cane crushing capacity of the plant is about 7500 tonnes per day [6]. The millers discharge their effluents through a network of long cylinders, which then transfer this wastewater into the ponds or lagoons built outside the plant premises at a distance of 1.5-2 km from the mills. The survey of these lagoons yielded that there were at least 14 lagoons, constructed in a haphazard way over an area of 2 acres, having an average depth of 8-10 ft, where HSM stagnate their effluents for a period of 7-8 months. The schematic presentation of wastewater disposal pattern indicating the relevant routes adopted by Habib sugar mills, Nawabshah is given in Figure 1.

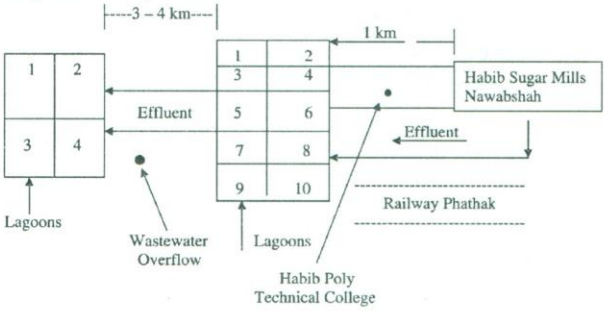


Figure 1: Schematic for wastewater disposal method adopted by Habib sugar mills, Nawabshah.

2.3 MATIARI SUGAR MILLS, MATIARI (MSM)

Figure 2 schematically depicts the wastewater disposal method practiced by MSM. The MSM is located at a distance of 6 km from Matiari town and about 35 km from Hyderabad. The MSM dispose of their plant effluent untreated through two large size concrete made drains, which carry the wastewater up to a distance of about 300 meters from the mills, before the entry of the piped wastewater into a big underground cylindrical pipe. This underground transportation of wastewater culminates into a canal, known as Chhandan Mori, which is 4 km away from the mills and serves as the final disposal point for MSM wastewaters. This canal is 25-28 feet wide and 8-10 km long having a depth of 2-3 feet.

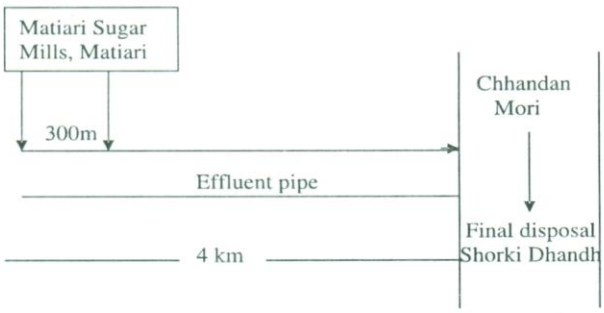


Figure 2: Wastewater disposal method for Matiari sugar mills, Matiari.

2.4 FAUJI SUGAR MILLS, TANDO MOHAMMAD KHAN (FSM)

Fauji sugar mills (FSM) is located within the city population of Tando Mohammad Khan, around 50 km away from Hyderabad. The wastewater of the mills is neither discharged to the ponds nor it is thrown away into any canal; rather, it is disposed of via a storm wastewater drain passing through the immediate vicinity of Mir Mumtaz agricultural lands. The drain water finally settles down in Tarai Nallah in Tarai village near Golarchi after bifurcating the Lakhat road. The wastewater disposal in case of FSM is schematically presented in Figure 3.

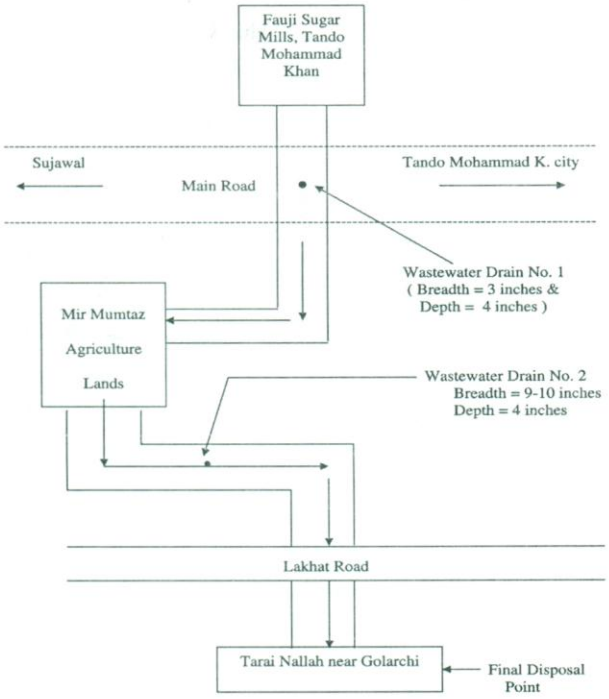


Figure 3: Wastewater disposal method for Fauji sugar mills, Tando Mohammad Khan.

2.5 WASTEWATER QUALITY COMPARISON OF THE SELECTED SUGAR MILLS

Each section of sugar mills generates wastewater with varying composition. Table 1 gives information about each section of the sugar mills along with the sources of wastewater generation and relevant parameters of concern within each section.

Table 1. ouse-wise wastewater characteristics in a sugar factory

Main Input	Unit house	Wastewater characteristics
Sugarcane	Mill House	Mostly contains suspended Solids and oil and grease contents. Also includes washing water used for floor cleaning, which contains sugar contents.
Sugar Juice	Process House	Washing of different omponents such as, Evaporator, Juice heater, Vacuum pan, clarifiers, generates aggressive effluents. Also includes manufacturing process water, which are high in BOD_5, COD concentrations.
Bagasse / Furnace oil	Boiler House	Wastewater from this unit has high pH, TDS and conductivity values

Wastewater samples were collected from two sections of the mills i.e. Mill house and Process house and one composite sample from the final drain carrying wastewater of all the sections and going outside the plant premises for the final settlement. Physical analyses was done *in-situ* (on site), while for chemical analyses, the samples had to be preserved by maintaining the sample container's inside temperature at 4^0C before being analyzed in the laboratory. Since sugar mills lie in category B of polluting industries by Pakistan Environmental Protection Act [7], the analysis of its effluents is comprised of five parameters namely pH, total suspended solids (TSS), BOD_5, COD and oils and greases. The sampling strategy for each parameters and point sources are given in Table 2:

2.6 ANALYTICAL PARAMETERS

Jenway direct reading pH meter was used to measure pH value. Oil contents were extracted from water by colorimetric extraction method using DR-2000 Spectrophotometer [8]. TSS were assayed via Photometric method using Hach's DR – 2000 Spectrophotometer. Whereas, BOD_5 test was carried out by making two diluted samples; one sample for the determination of its initial dissolved oxygen (DO_i) and the other sample was kept in incubator for five days at 20^0C. After five days the sample was tested for its final dissolved oxygen (DO_f) and with the help of following formula BOD_5 of the given sample was determined:

$$BOD_5 \text{ in mg/L} = \frac{DO_i - DO_f}{\text{Sample volume taken for dilution}} * \text{Total Volume of BOD bottle}$$

COD test was conducted by standard dichromate reflux method using the following equation [9]:

$$\text{COD in mg/L} = \frac{(A - B) \times 1000 \times M}{C}$$

Where: A = mL used in titration of blank
B = mL used in titration of prepared sample
C = mL used in titration of standard solution
M = Multiplier (depending upon the sample of volume taken)

Table 2: Wastewater sampling and analytical parameters.

Sample Point Source	EPA Prescribed Parameters	No. of samples from each sugar mill
Mill House	Oil and greases, pH, TSS	6
Process House	BOD_5, COD, pH, TSS	6
Combined or mix Effluent	TSS, pH, Oil and greases, BOD_5, COD	6

2.7 SURVEYING THE SEVERITY OF PROBLEMS IN THE VICINITY OF SUGAR MILLS

In order to know people's opinion about the problems caused by the disposal of wastewater into the surrounding environment, the people living nearby the vicinity of the mills were asked certain questions. The Specimen of the questionnaire is shown in Table 3:

Table 3: Specimen questionnaire for public opinion about sugar mills wastewater disposal effects.

Name of Industry :
Area / Locality :

Sr. #.	Name of Person and Age	Profession	Level of Impacts on				
			Human Health	Live-stock	Agricultural Lands	Odor Problem	Surface Or Ground Water
			A = 45% B = 55%	B =45% D =25%	C = 35% D = 45%	B = 55% D = 25%	A = 50% B = 50%

A = High / severe, B = Mild, C = Less, and D = No Problem

3. RESULTS AND DISCUSSIONS

3.1 WASTEWATER DISPOSAL BY SUGAR MILLS

All the three sugar mills namely Habib sugar mills (HSM), Fauji sugar mills (FSM) and Matiari sugar mills (MSM) discharge their wastewaters untreated, which can adversely affect the quality of receiving sources of water bodies or soils and agricultural lands. Major reported effects as obtained during environmental surveys of the three sugar mills were concerned with odors from settled untreated wastewaters, contamination of groundwater sources, soil erosion along with production of sludge from settled effluents with the passage of time. In case of FSM, it was noticed that the FSM effluent passing through Mir Mumtaz agricultural lands had rendered a portion of soil, parallel to the effluent drain, infertile and barren. The effects were apparent up to 100 m giving a blackish look to the soil. The area people also said that their animals were observed to fall ill especially after taking frequent dips into the sugar mills wastewater drain. While in case of HSM, the after effects of untreated wastewater disposal included contamination of groundwater resources, especially near the wastewater ponds. This information or observation was validated when a sample of groundwater was collected from the railway crossing hand-pump situated at about 600 m from the wastewater ponds for its physico-chemical analysis. In addition, the said hand-pump was reportedly not in use for the last two years due to poor quality of the water. The sample

analysis revealed that it had zero DO, pH 3.8, blackish color, undesirable smell and turbidity of > 500 NTU. It is more likely that the seepage or percolation of HSM wastewater might have caused the pollution of underground water reserves in the vicinity of unlined wastewater lagoons of HSM. Unlike surface water pollution, ground water contamination does not occur within a short time, but takes many years to contaminate the quality of ground water. This is because the mechanism of groundwater contamination is not direct and too slow as well, hence, its decontamination *in-situ* may also take long time in order to improve its quality [7, 8]. The children of the area were also seen splashing around in the stinking ponds of HSM making them susceptible to skin diseases such as acne and rash problems. When asked from a local Dermatologist about this, he said that since the wastewater of the ponds was highly organic in nature, therefore, children in this case were an easy target for skin diseases like darkening of skin and rashness, besides attracting many insects and flies towards their sugary skins. MSM wastewater was disposed of via a network of cylinders for its final settlement in the canal known as Chhandan Mori, which was a source of very pungent odor, as experienced beforehand. Moreover, the stagnant MSM wastewater was seemingly a breeding ground for mosquitoes, insects and other fliers. The data obtained from the survey questionnaire (Table 3) suggested that on average over 50% people were of the opinion that the sugar mills wastewater posed a direct or indirect effect on their health and on their livestock. They also complained of the odor nuisance caused by the stagnant wastewater in the ponds, as well as contamination of groundwater sources in the immediate vicinity of wastewater ponds.

3.2 WASTEWATER QUALITY COMPARISON OF THE SELECTED SUGAR MILLS

(a) Physical Analysis

The color of all the wastewater samples was juicy or brownish with a sugarcane juice-like smell. This implies towards the presence of sugar residues in the wastewaters. The wastewater temperature ranged from 42 to 45^{0}C, which is above the standard value of 40^{0}C (NEQS limit). It is an established fact that the solubility of oxygen decreases as the temperature and pressure increases. Hence, warm wastewater discharges tend to lower the value of DO. Since metabolism rates may go up at higher temperatures, warm wastewater discharges may disturb the ecosystem of the receiving water bodies.

(b)Chemical Analysis

The wastewater pH for the three sugar mills is graphically shown in Figure 4. The pH results for process house and final effluent samples showed that HSM wastewater was more acidic in nature as it contained the lowest pH valued of 5. 15 and 3. 93 respectively, as compared to the other two samples. The lower pH values for HSM may likely be due to the difference in the nature of manufacturing process of each sugar mills, with HSM using Defecation Remelt Carbonation Sulphitation process in which derivative acidic compounds of both sulphur and carbon are formed, while FSM use Defecation Remelt Carbonation process without sulphitation factor. Secondly, HSM wastewater temperature was 45^{0}C as compared to 42^{0}C for FSM effluent, thus increasing the metabolic activity rate in the HSM wastewater, which would have resulted in the decrease of pH value. While pH of mill house samples of all the three sugar mills was found in alkaline range and within the specified limits of 6- 8.

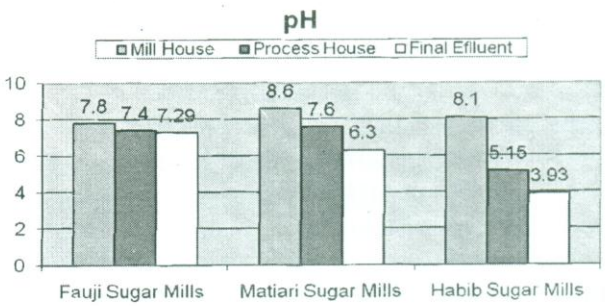


Figure 4: Wastewater pH for the three sugar mills.

Total suspended solids (TSS) were determined for the effluents of process house and combined mixture. TSS values were also found to be much higher than the standard value of 150 mg/L. TSS results suggested that MSM wastewater had the highest TSS of 300 mg/L from process house section as compared to other mills samples. This may largely be due to the lack of observation of in-plant control measures. Whereas, final effluent of HSM contained higher TSS value of 653 mg/L than those of the other two sugar mills, indicating higher cane crushing capacity of the industry and as a result producing large amounts of both fly ash and bottom ash and their subsequent possible presence in the effluent streams. In addition, the industry is also running a fiber plant within its premises to process bagasse for further uses, hence the

coarse bagasse particles are always found there suspended in the air and getting mixed with the final effluent. TSS results are graphically shown in Figure 5.

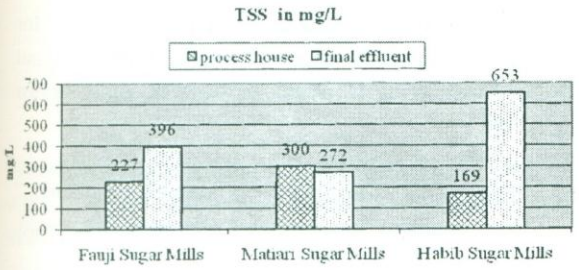


Figure 5: Total suspended solids of both process house and final combined effluents for the three sugar mills.

5-day biochemical oxygen demand (BOD_5) and chemical oxygen demand (COD) were analyzed for process house wastewater and final combined effluents and the results are shown in Figures 6 and 7 respectively. Like TSS, BOD_5 and COD values were also on the higher side as compared to their NEQS values. MSM final effluent had the highest BOD_5 and COD values of 673 mg/L 3673 mg/l in comparison to those of other two samples. While process house BOD_5 value of 365 mg/L for FSM was the highest among the process house BOD_5 values for the three mills. As for COD, MSM effluent registered the highest value among all the samples i.e. 1995 mg/L. However, HSM effluent had the lowest BOD_5 and COD values measured both for process and combined effluents i.e. 93 mg/L and 181 mg/L for BOD_5 and 390 mg/L and 634 mg/L respectively. Increase in BOD and COD values may by attributed to the spillage of molasses and lost or leaked sugar contents on the floors of the mills, which are swept or washed away to be mixed with the effluents causing higher levels of BOD and COD [10].

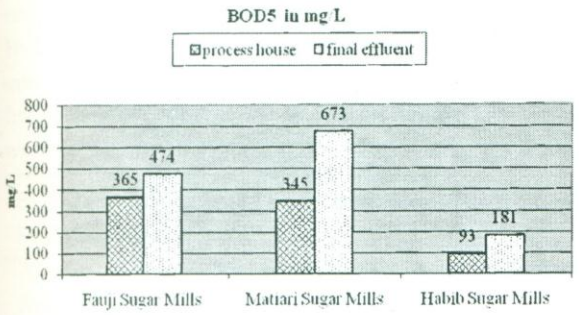


Figure 6: BOD_5 for process house and final combined effluents for the three sugar mills.

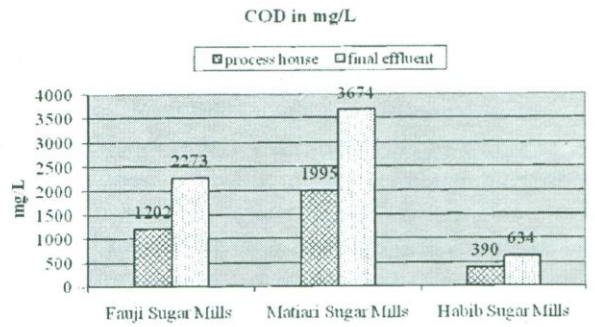


Figure 7: COD for process house and final combined effluents for the three sugar mills.

Oils and greases (OG) were determined for mill house samples, where oils and greases are particularly used, as well as for final effluent samples. Figure 8 show the results for oils and greases present in the wastewater samples from three sugar mills. These results show that OG were also present in wastewaters in higher quantities as against the set standard value of 10 mg/L. HSM effluent from mill house recorded the lowest value of OG i.e. 20 mg/L, while similar sample of FSM gave the value of 49 mg/L, which was the highest among the three samples. Similarly, final effluent sample of HSM contained lowest OG contents i.e. 27 mg/L and FSM effluent had the highest OG contents i.e. 185 mg/L. The final effluent OG contents are on the higher side as it carries the effluents of all the lubricated houses of the mills.

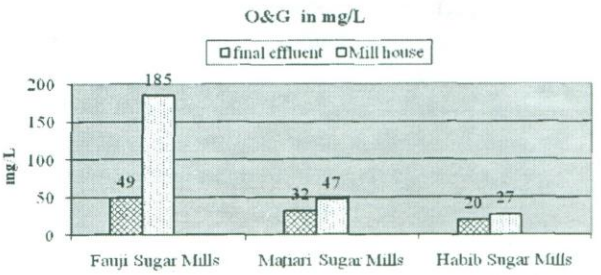


Figure 8: Oils and grease contents for mill house and final combined effluents for the three sugar mills.

4. CONCLUSIONS

From this research study, it was concluded via the conduct of environmental survey that people living in the vicinity of the selected sugar mills were being affected by the discharge of untreated sugar mills wastewater. In case of Habib sugar mills, Nawabshah, wastewater ponds built

outside the plant premises were the source of pungent and unpleasant odor that is bound to increase with time. In addition, this study has transpired that the groundwater quality in the immediate vicinity of HSM ponds has been affected severely to the extent that it is no longer in use for drinking purposes, which was also verified via laboratory analysis. The disposal of sugar mills wastewater in open pits is a source of ailments for humans, flora and fauna. While the wastewater seepage into soils may result in the loss of soil fertility. The effluent parameters such as pH, temperature, BOD_5, COD, TSS, Oils and greases were all found higher than the standard limits, which will be a source of further pollution with each passing day. Hence further work in this regard should focus on the treatment of sugar mills wastewater before their discharge into the environment.

ACKNOWLEDGEMENT

Authors are highly thankful and appreciative of the cooperation extended in this regard by the management of the sugar mills in question.

REFERENCES

[1] Sugar – World supply and demand summary, Report by Spectrum Commodities, USDA Economics and Statistics System, USA, referred to from the website: www.spectrumcommodities.com/education/commodity/statistics/sugar.html

[2] Aftab Ahmed,"Pollution Prevention and in-plant control measures in sugar industry", A draft Report by Technology Transfer for sustainable Industrial Development, Islamabad, pp: 16 – 23. 1997.

[3] Asadi, M., 2007, "*Beet-Sugar Handbook*", John Wiley & Sons, Inc., USA.

[4] ESCAP/ FAO/ UNIDO, "Agro-chemicals – News in brief," Fertilizer Advisory, Development and Information Network Asia & the Pacific, Thailand, 1994.

[5] Trivedy, R.K, "Advances in Wastewater Treatment Technologies," 1st Edition, Global Science, Aligarh, pp. 33, 1998.

[6] Abdul Qayoom Jakhrani, "Impact of oxidation ponds on groundwater quality at Hyderabad," MS thesis, Faculty of Engg. MUET, Jamshoro, 2002.

[7] Azhar Uddin Khan, "Opportunities for Environmental solutions in Sugar sector," Journal of Engineering Horizons, Issue. 133, pp.7 August 1999.

[8] Hach, "DR/2000 Spectrophotometer Handbook," Hach company for systems for analysis, Loveland, Colorado, USA, 1993.

[9] APHA, "Standard Methods for the Examination of water and wastewater," 19th Edition, APHA, WPCF, AWWA, New York, USA, 1995.

[10] Musa Kaleem et al, "Environmental impact of industrial effluents with reference to the quality of groundwater", Pakistan Journal of Analytical Chemistry, Centre of Excellence in Analytical Chemistry, University of Sindh, Jamshoro, Vol. 2, pp: 48-52, 2001.

ISSN 1605-8607

ASSESSMENT OF RAW WATER QUALITY SUPPLIED TO HYDERABAD FOR DRINKING PURPOSES

Niaz Ahmed Memon*, Abdul Rehman Memon**, Noorullah Soomro***

ABSTRACT

Drinking water is an indispensable life supporting ingredient for daily life. The quality of drinking water at all stages while supplied through distribution systems requires some specific monitoring. Present study is based on the assessment of quality of raw water supplied to treatment plants for onward distribution to the city of Hyderabad (Pakistan) serving about 60% of the population and the 40% population is served with untreated water. The major intake for supply of water is from River Indus at Kotri barrage Jamshoro where unexpected and unannounced release of highly contaminated water from Manchhar Lake is received every year, thus creating major threats to public health and aquatic life. The present study was conducted on a regular sampling basis for a period of half of the year. The aim was to assess the quality of raw water at all three stages and observe the changes in the parametric values while collected, stored in the lagoons and distributed to the end user. Spectrophotometer and Multiparameter Meter were used for the analysis of these parameters. The parameters include pH, turbidity, TDS, DO, Iron, Sulphate and Hardness. A critical evaluation of data presented indicated that pH generally ranged from 6.8–7.8; turbidity varied from 60.87 -125 NTU at location#1, 42.2-111 NTU at location #2; , TDS values were found to be ranging between 346– 487 mg/l at location 1 and were reduced in lagoons by 301-404 mg/l at location 2, the DO ranges were found 8.16-9.98 at location 1, 8.01-9.43 at location2. The observed variations in iron were found from 0.59 -1.34 at location1 and were increased in settled water at lagoons with a variation of 0.92-1.65 mg/l. It was further revealed that the turbidity and iron contents of the raw water at two locations were elevated though 16.76% of turbidity was reduced during the storage in lagoons but the iron content remained elevated while comparing with WHO standards.

Keywords: Raw water quality, River Indus, Drinking water, Hyderabad

1. INTRODUCTION

The current investigations on the quality of groundwater of Pakistan and application of the strategies for monitoring it, has given an idea that there is lack of sufficient data availability for raw water supplied for drinking purpose to various cities including Hyderabad [1]. From the very limited Physico- chemical analysis on raw water supplied for drinking purpose to the city of Hyderabad it is indicated that the most critical water quality problem relates to the deterioration of raw water from rivers [2]. The major source for supply of drinking water to the city of Hyderabad is River Indus. The canals from this river are supplying the domestic water to the city having an approximate population of 1.8 million. Due to sudden and spontaneous discharge of highly polluted water from the Manchhar Lake, river water becomes highly contaminated specially at the major intake (Location 1) as shown in figure 1 [3]. From this river at Jamshoro, water is supplied to the lagoons named North and South lagoons of 400 MG for pre settlement and then brought to the treatment plants for further distribution through network. Physico chemical analysis of raw water indicates that the turbidity and iron were elevated at both locations during the study period started from February 2012 to July 2012. Pakistan Council of research in Water

*Professor, Department of Civil Engineering, Quaid-e-Awam University of Engineering sciences & Technology, Nawabshah, Sindh, Pakistan

**Assistant Professor, Department of Chemical Engineering, Mehran University of Engineering & Technology, Jamshoro, Sindh, Pakistan

***Assistant Professor, Department of Textile Engineering, Mehran University of Engineering & Technology, Jamshoro, Sindh, Pakistan

Resources (PCRWR) concluded that "the only problem" in untreated water supplied to many parts of the city of Hyderabad by Water and Sanitation Authority (WASA) Hyderabad is the presence of excessive turbidity and consequent bacteriological contamination [4]. In 2004, 11 people were known to die, and thousands of children were exposed to diarrhoea in Hyderabad due to contaminated drinking water and that the source was identified as a lake discharging polluted saline water to the river Indus near Jamshoro, location 1 in this study [5]. Based on the reports from World Health Organization (WHO) Pakistan is declared at second amongst 31 Asian countries in annual diarrheal episode among young children [6]. Physico chemical analysis of this study indicates the changes of the various critical parameters of raw water at three important different locations.

2. MATERIALS AND METHODS

A descriptive research study is carried out in accordance with [7, 8] based on the Physico- chemical analysis of the raw water supplied to the city under observation.

To make the samples representative of the entire assessment of raw water quality and the selection of parameters is based on three categories. First category is the raw water taken from the River Indus (Major intake) to the treatment plants and direct supply (untreated water) to some areas (location 1). Second category is the Lagoons water quality (location 2) where the raw water is pumped to store the water for a certain period for settlement and onward supply of treated water through treatment plants; and that the rest of the water is supplied untreated to 40% of the population of the city. Third category is the raw water entering into the inlet of the treatment plants (location 3).

Figure 1: showing the satellite view of the locations Sampling Frequency & Equipment In this study, 120 samples are analysed for 03 selected locations started from the month of February to July 2009. The 06 months comprehensive descriptive study was planned in a way to cover both seasonal changes of summer and winter. The sampling frequency is shown in Table 1. Spectrophotometer DR 2700 (HACH) and SensIon 59 Multiparameter Meter (HACH) were used in this research work.

Table 1: Sampling Frequency

No.	Month	Week	Observations
1	February 2012	1	02
2	March 2012	4	09
3	April 2012	4	08
4	May 2012	4	09
5	June 2012	4	09
6	July 2012	2	03
Total samples on each location			40
No. of observations = 03 x 40			Total 120

3. RESULTS AND DISCUSSIONS

The present study revealed that the values of the parameters at location 2 and location 3 are almost same as the mean values were found similar for these locations. Therefore the observations at location 1 and 2 are presented and discussed. The pH values of the samples at L1 and L2 are presented in fig 2and 3 respectively. In this study, the mean values at location 1 and location 2 are 7.37 and 7.40 lying with the normal ranges available from WHO guidelines. Kandhar [9] reported the pH of river water as 7.98. According to PCRWR "National Water Quality Monitoring Program" (2011) pH variation of river water at this location is 7.30 [10]. Therefore our results are consistent with both of these results for this parameter. The statistical comparative values for pH are shown in Table 2.

Table 2: Comparative Values of pH

S.No.	Author		Kandhar		PCRWR	
	Min	Max	Min	Max	Min	Max
L1	6.8	7.9	---	7.98	---	7.30
L2	6.9	7.8	---	---	---	---

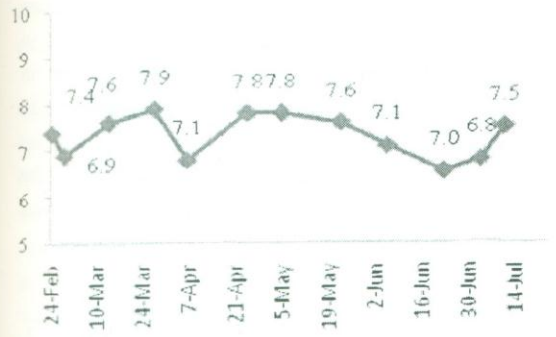


Figure 2: pH values at Location 1

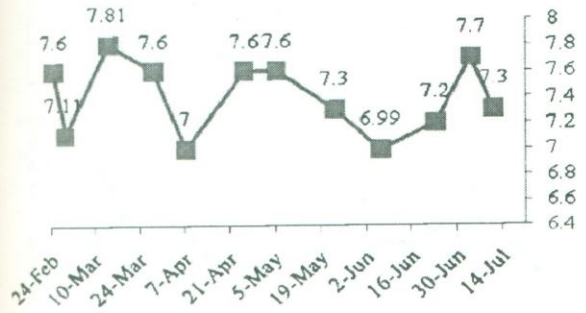


Figure 3: pH values at Location 2

One of the important parameter of raw water is turbidity which was assessed on location1 and location 2 against the dates of the experimental study as shown in figure 4 and 5. The minimum value at location 1 is observed as 60.87 NTU in the month of February (winter season) and the maximum value is 125 NTU in the month of March. However the mean value at L1 is observed as 91.2 with standard deviation of 19.54. The results of PCRWR study shows the mean value of River water as 86 NTU which confirms the raw water variation at this location. The values of turbidity on Location 2 show a minimum value of 42.4 NTU on 28th February while the maximum value is observed as 111NTU on 28th March indicating mean value of 74.49. It shows a trend of increase in turbidity from February to March. Generally the turbidity level in winter and summer seasons is different. The comparative values of turbidity are given in Table 3

Table 3: Comparative Values of Turbidity

S.No.	Author		Kandhar		PCRWR	
	Min	Max	Min	Max	Min	Max
L1	60.87	125	---	---	---	7.30
L2	42.2	111	---	---	---	86

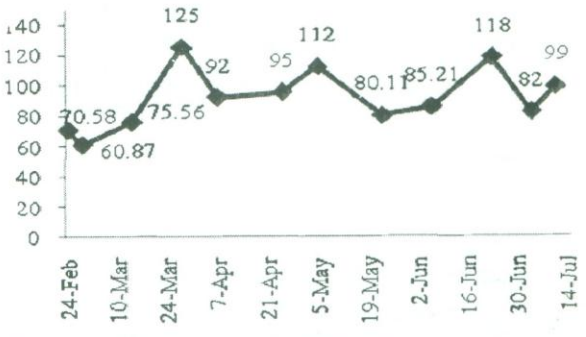


Figure 4: Turbidity values (NTU) at Location 1

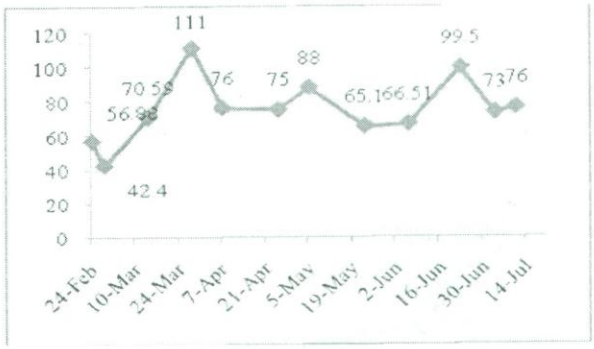


Figure 5: Turbidity values (NTU) at Location 2

TDS is a measure of the combined content of all inorganic and organic substances contained in a liquid in molecular, ionized or micro-granular (Colloidal solution) suspended form. Measured values of TDS at location 1 and 2 are shown in graphical form in the figure 6 and 7. At location 1 the mean value of TDS is 423.416 mg/l varying 346-487 mg/l. (PCRWR, 2012) reported TDS at 277 mg/l.

The mean value of TDS at Location 2 is 352.5 mg/l (301-404 mg/l). The TDS values dropped in the lagoons at this location. A decrease of 16.76% in TDS is found in the Lagoon water as the mean value of TDS is found as 352.5 mg/l. Kandhar (1998) has reported TDS value of 188 mg/l for this location. Since the mean value of TDS is below 500 mg/l so, as far as TDS concentration is concerned it can be recommended as safe for drinking according to WHO guide lines. The presence of dissolved solids in water may affect its taste. The palatability of drinking water has been rated by panels of tasters in relation to its TDS level, expressing that TDS less than 300 mg/l is "Excellent". TDS between 300 mg/l and 600 mg/l is "Good" and between 600 and 900 mg/l is "FAIR" [11]. Comparative TDS Values are shown in Table 4.

Table 4: Comparative Values of Total Dissolved Solids

S.No.	Author		Kandhar		PCRWR	
	Min	Max	Min	Max	Min	Max
L1	346	487	-	152	277	7.30
L2	301	404	-	188	---	86

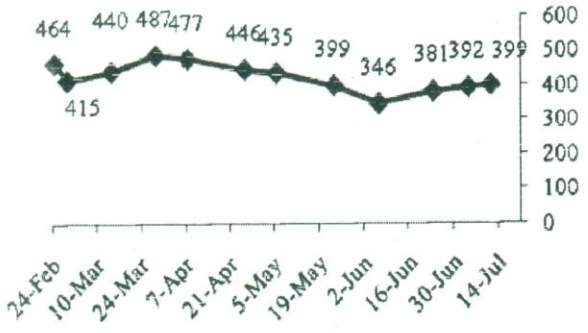


Figure 6: TDS (mg/l) values at Location 1

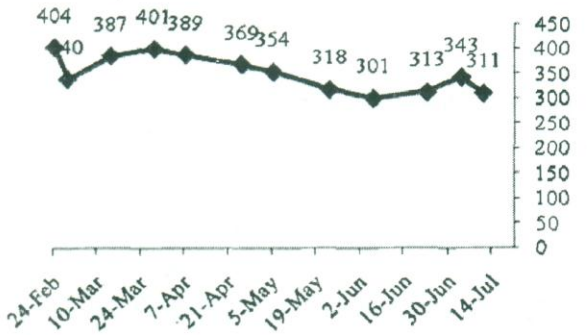


Figure 7: TDS (mg/l) values at Location 2

In drinking water supplies, iron salts are unstable and are precipitated as insoluble iron hydroxide, which settles out as a rust colored silt. If the content of iron exceeds 0.3 mg/l, staining of laundry and plumbing may occur [12].

The experimental values of Iron (Fe) on both Locations1 and 2 are shown in Figures 8 and 9 respectively. Statistics of Iron on location 1 includes mean value of 0.86 mg/l, minimum value is 0.59 mg/l, and maximum value is 1.34 mg/l. The PCRWR (2004) reported 3.43 mg/l of iron in raw water, which is very high, while Kandhar (1998) reported 0.23 mg/l for raw water. The values of this parameter on Location 2 show the mean value of 1.237 mg/l, minimum value is 0.92 mg/l, and maximum value is 1.65 mg/l. Comparing both values of location1 and Location 2, it is evident that the mean content of iron at 30.47 % in raw water is increased in lagoon. The same trend of increase of iron content in lagoon is confirmed by Kandhar (1998) reporting 0.34 mg/l compared to 0.23 mg/l of raw water. No significant change is found within the quality of raw water and the inlet of the treatment plant. The comparative statistical values of Iron are shown in Table 5.

Table 5. Comparative values of Iron (Fe) mg/l

S.No.	Author		Kandhar		PCRWR	
	Min	Max	Min	Max	Min	Max
L1	0.59	1.34	-	0.23	--	--
L2	0.92	1.65	-	0.34	---	--

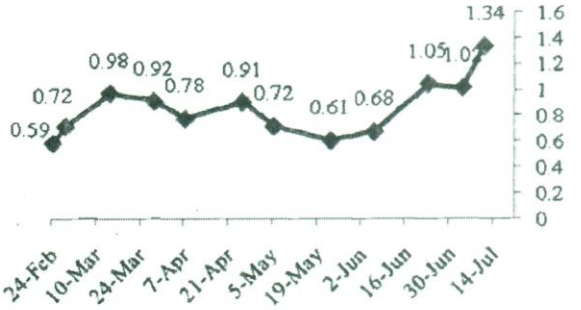


Figure 8: Iron (mg/l) values at Location 1

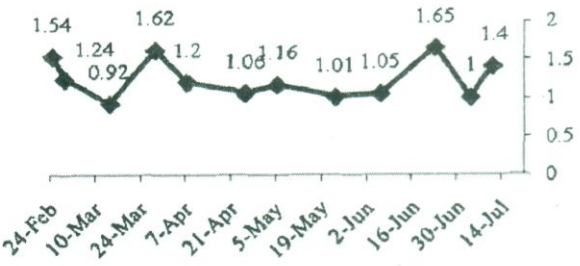


Figure 9: Iron (mg/l) values at Location 2

4. CONCLUSIONS

From the present study, summarizing the results it is concluded that during period of six months from February to July, 2012, the pH of water remains within the WHO drinking water guidelines value (6.5-8.5). Its Turbidity was found very high ranging from 60-125 NTU. (The PCRWR study is in agreement. From mean value it is

clear that that there is drop in Turbidity due to settling of river water.

The TDS value has increased (487 mg/l) in comparison to previous study-188 mg/l (Kandhar) and152 mg/l, (PCRWR, 2004) showing that there is influent of saline water in the river system from different sources, especially from Manchar lake. There is a precaution note as this value is reaching the WHO guideline value 500 mg/l. This trend must be checked and alternate measures be taken to control the TDS & other health related parameters. The raw water DO has not been determined frequently by researchers. The Iron content in raw water at the intake from river was higher (0.59-1.34 mg/l) than (0.23 mg/l). It was also noticed that the Fe content increased in lagoon water (0.92-1.65 mg/l). This increase was also reported by Kandhar (1998) who additionally witnessed the high rise of Manganese in lagoon waters. This suggests that iron and manganese dissolved solids are possibly coming from the rocky structure of lagoons, the total hardness of river water was found as 300-395 mg/l. Hardness and TDS of Raw water are found increasing continuously from 1998 to present study. The source of increase in iron also must be taken into consideration for further study of the raw water.

REFERENCES

[1] Chilton, P.J., Jamieson, D., "Pakistan Water Quality Mapping and Management Project". Scoping Study, LSHTM/WEDC Report DFID UK. 2010.

[2] Khan , K., Khan, M., "Monitoring of Pollution of Water of North West Frontier". Part I, J. of Physical Chemistry, 12, pp. 77-90, 2011.

[3] Tahir. M., "Drinking Water Quality Monitoring in the Rural Areas of Rawalpindi". In the proceedings of the National Workshop on Quality of Drinking Water, Islamabad, pp.35-39, March 2009.

[4] PCRWR, DRIP, "Technical Report on Water Quality Monitoring and Assessment of Hyderabad", Tandojam, pp. 1-99, 2012.

[5] Hassan A. 2004. "Pensioned Water Deaths in Pakistan", Report by BBC available at http://news.bbc.co.uk/hi/south_asia/3745165 last accessed 20th January 2013.

[6] World Bank, Ministry of Environment Government of Pakistan. "Pakistan Strategic Country Environmental Assessment, Final Report", Islamabad , pp.15, 2005.

[7] WHO 2006 Guidelines for Drinking Water- Water Quality 3rd Edition, Wprld Health Organization availa http://www.who.int/water_sanitation_health/dwq/gdwq3rev/en/ last accessed 8 Feb 2013

[8] APHA "Standard Examination for the examination of Water and Waste Water, 20th Edition, Washington DC, American Public Health Association, 1989.

[9] Kandhar, I.A., Ansari, A.K. "Drinking Water Quality of Hyderabad City". In the: "Proceedings of 24th WEDC Conference, Sanitation, and Water For ALL", Islamabad, Pakistan, pp. 254-257, 1998.

[10] PCRWR,"National Water Quality Monitoring Program" NWQMP Report, Annexure III-12: pp.33-40, 2011.

[11] WHO." Total Dissolved Solids in Drinking Water". Backgrpound document for development of WHO guide lines for Drinking Water Quality, 2003 (WHO/SDE/WSH/03.o4/16) available at www.who.int

[12] WHO, " Iron in Drinking Water." Background document for developmet of WHO Guide lines for Drinking Water Quality , 2003.(WHO/SDE/WSH/03.04/08) available at www.who.int

ISSN 1605-8607

ASSESSMENT OF DRINKING WATER QUALITY SUPPLIED TO QUAID-E-AWAM UNIVERSITY OF ENGINEERING, SCIENCE & TECHNOLOGY (QUEST) NAWABSHAH, A CASE STUDY

Roshan Zaman Sehto*, Abdul Rehman Memon**, Niaz Ahmed Memon***

ABSTRACT

A comprehensive descriptive study was carried out in vicinity of Quaid-e-Awam University of Engineering Science & Technology (QUEST) Nawabshah, Sindh, Pakistan with the aim to ascertain the physicochemical quality and the presence or absence of bacterial contamination of drinking water supplied to this University. The areas under study included Boys Hostel, Mosque, department of Civil Engineering, Mechanical Engineering (B-Sector), departments of electrical engineering and Computer Sciences (A-sector), Administration block and civil laboratories (C-sector). The physicochemical data were collected from 08 different locations. Two different forms of chlorine (Total and residual) were measured at each sampling point. The significant analysis from the data available indicated that the pH generally varied from 8.31-8.41, TDS ranged between 1295-1313 mg/l, electrical conductivity was found fluctuated from 2160 – 2190 µS/cm, turbidity was found within 6.85-9.72 NTU, total hardness as Ca CO3 was recorded Over Range (>500 mg/l) at each station and all the samples were having no chlorine in both Total and residual free forms. All the samples (100%) were found positive for bacterial contamination.

Key words: Drinking water quality, Chlorination, Disinfection, Distributions System, Nawabshah

1. INTRODUCTION

Drinking water is considered as an integral aspect of life sustainability. Its availability and quality has remained one of the critical issues in the water sector of Pakistan. It is considered that 60% of the population is served with safe drinking water in urban areas but the same is not true for rural areas, as the figure is not more that 45% [1]. For analyzing the quality of the water there is no any regular monitoring program in this country even the Drinking Water Treatment Plants (DWTP) are not included in the monitoring system which allows severe bacteriological contamination to go unobserved at all the levels of the distribution of the drinking water supplied to the end users [2]. It is believed that the distribution system of the drinking water plays an important and vital role for preserving and supplying the safe drinking water to the public [3]. The availability of disinfectant like chlorine in the distribution has remained as an exceptional sign for studying water quality in the distribution network [4]. Many researchers have indicated that the drinking water supplied without having the traces of disinfectant as residual chlorine bacterial growth of many microorganisms can be traced with elevated levels while the availability of residual chlorine reduces the microbial contamination level and its frequency of the occurrences at the tap of the end user [5]. Various viruses such as Hepatitis A Virus, Hepatitis E Virus, Rota Virus and parasites like Entamaeba histolytica, Giardia lamblia are also to be found in the absence of disinfectant [6-9]. The quality of the drinking water supplied through distribution network in Nawabshah city has always remained questionable and no major research projects are carried out so far strengthening the fact that the water microbiology and physicochemical analysis has remained a neglected subject [10-15]. However the present study is based on both physicochemical analysis and presence or absence of bacterial growth in the drinking water supplied to QUEST Nawabshah which indicates that all the samples (100%) are free from chlorination and

* Postgraduate Student, Department of Civil Engineering, Quaid-e-Awam University of Engineering, Science and Technology, Nawabshah, Sindh

** Assistant Professor, Department of Chemical Engineering, Mehran University of Engineering and Technology, Jamshoro

*** Professor, Department of Civil Engineering, Quaid-e-Awam University of Engineering, Science and Technology, Nawabashah, Sindh

consequently are positive for bacterial contamination.

2. MATERIALS AND METHODS

Spectrophotometers, Multiparameter Meter were used to analyze the physicochemical parameters. For presence and absence (P/A) of E. Coli in the samples, Broth Ampules method (HACH) kit was used. A total of 80 samples were collected from 08 different locations in such a way to make it representative study of the area taken into consideration. The samples were collected with intensive care in highly sterilized plastic bottles of 200 ml. All the samples were collected, transported and analyzed according to APHA [16]. Sampling locations and sampling frequency are shown in Table 1. In this study pH, TDS, Electrical Conductivity (EC), Temperature, were analyzed by SensIon 59 Multiparameter Meter (HACH), Hardness, Turbidity and Dissolved Oxygen (DO) were analyzed with Digital Spectrophotometer DR 2700 (HACH).

Table 1: Sampling locations and frequency

S. No.	Area	Location	Samples
1	Boys Hostel	Hostel Masjid	10
2	Boys Hostel	B Block	10
3	B – Sector	Civil Eng. Dept	10
4	B- Sector	Mech. Eng. Dept	10
5	A-Sector	EE Dept.	10
6	A-Sector	C.S. Department	10
7	C-Sector	Civil Labs	10
8	C- Sector	Admin Block	10

3. RESULTS AND DISCUSSIONS

From the present study it is revealed that the critical parameters of drinking water at various locations of QUEST are not found satisfactory. The pH at all locations is ranging from 8.31-8.41, with a clear indication that the pH of location 8 (Administration block) is very close to 8.5 the maximum contamination level (MCL) for pH by WHO (2007) [17]. The situation of the water supply in other cities of the country is not satisfying but for pH parameter of Rawalpindi drinking water, it has been reported as 7.3. These results are also in accordance with the earlier study conducted by PCRWR- DRIP in 2007, showing that the pH values varied from 7.00–8.30 [18]. The mean values of pH against each location are shown in figure 1.

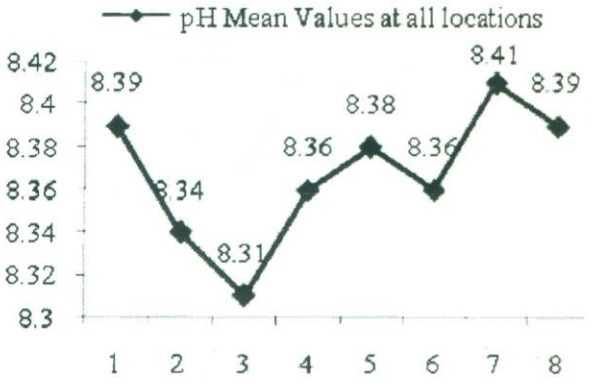


Figure 1: pH Mean values at all locations

Temperature is considered as one of the significant limitations due to the fact that the microbial growth is directly related to it. Temperature observed at the time of collection of water samples ranged from 19.6 to 20.64°C. The minimum value of 18 °C is observed at location #5 (Electrical Engineering Department). How ever the data has indicated that this parameter exceeds the WHO limit of 12°C. These observed results are in line with the study carried out by [19]. The mean values of this parameter at all locations are presented in figure 2.

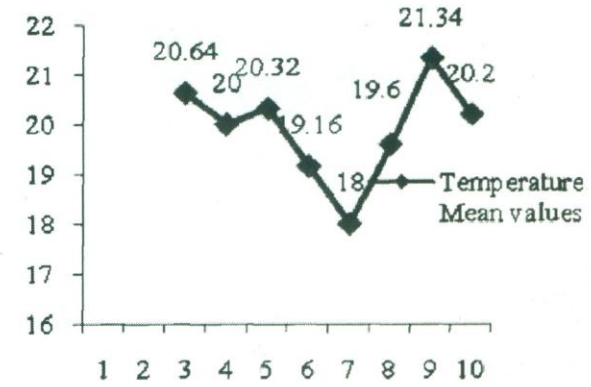


Figure 2: Temperature Mean values at all locations

The total dissolved solids (TDS) values of the samples were found fluctuated from 1295 to 1313 mg/l, these limits exceed the WHO the highest limits for TDS in drinking water of <500 mg/l consequently based on TDS decisive factor all the water samples were found unfit for drinking purposes. These results are not in accordance with [19 & 20]. The observed mean values of TDS are presented in figure 3.

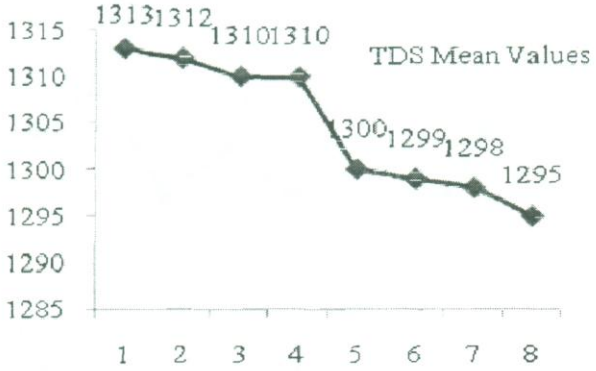


Figure 3: TDS (mg/l) Mean values at all locations

Electrical Conductivity (EC) is one of the important parameters of drinking water. In our study the values of conductivity varied from 2160 to 2190 µS/cm. In a study it is revealed that statistically a close proportionality of TDS and EC remains evident and that the increased temperature increases the conductivity due to increase of termination of ionic variety [21]. The observed mean values of EC of this study are presented in figure 4.

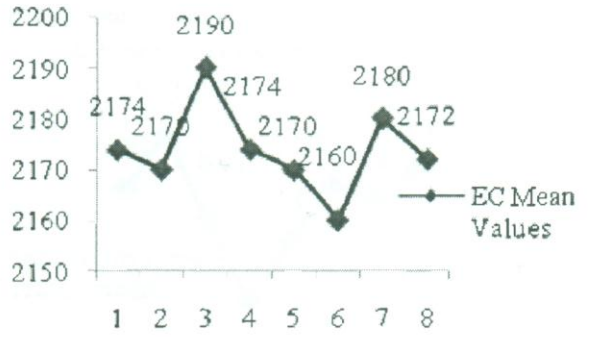


Figure 4: EC Mean values at all locations

Turbidity has a vital role to maintain the quality of drinking water especially when the microbial contamination is taken on priority. The microbiological quality of drinking water can be significantly affected by turbidity since it is dependent variable of source water quality and the basic determinant for various elements of the treatment plants as well [22]. The maximum contamination level (MCL) allowed by WHO for this parameter is 5 NTU, whereas in this study we are not within allowed MCL. The values varied from 6.85 to 9.72 NTU. Location # 8(Civil Labs) is the critical location where the maximum turbidity 9.72 NTU is observed. Based on turbidity, the drinking water quality at all locations cannot be considered as safe for drinking. The mean values of turbidity are shown in figure 5.

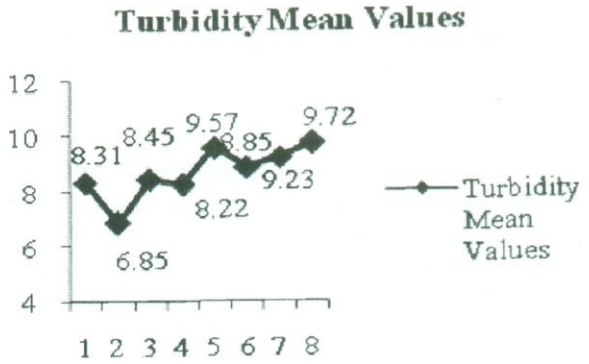


Figure 5: Mean Values of Turbidity at all locations

Chlorine in both forms as Total chlorine and free residual chlorine was not found in all the samples at all locations. Non-availability of chlorine indicates that there is potential microbiological presence of many species which may cause serious health problems for the end users. The same situation can be realized in other cities of the country as confirmed in a study [23], endorsing that the free residual chlorine was not found in drinking water samples collected from the distribution line of sukkur.

Hardness as $CaCO_3$ at all locations were found over range of (>500 mg/l) when measured on spectrophotometer. The Hardness MCL allowed by WHO is not more than 500 mg/l, therefore exceeded limits of hardness in all the samples reveals that the quality of drinking water is deteriorated. No previous study is available yet for these locations except PCRWR (2004) reporting 370 mg/l for the water supplied in Hyderabad city.

An effort was made to analyze the bacteriological contamination by performing the Broth Ampules Kit (HACH) method to confirm the presence and absence of these E.Coli. According to WHO there should be 00 / 100 ml coliform in the drinking water. The observations exposed that 100% of the samples were found positive (presence) of E.Coli.

4. CONCLUSIONS

After comprehensive critical analysis of the samples of the drinking water supplied through WASA Nawabshah

from the original source of Rohri Canal and Gajra Wah, supplied by distribution lines to the different locations of QUEST, taken in this study it is concluded that:

- pH values are within normal ranges (8.31-8.41) with a note that these values are very close to the standard values (6.5-8.5) therefore proper monitoring of this parameter is required essentially.
- TDS are also higher than the specified limits of 500 mg/l in each sample indicating the deteriorated quality of the water supplied.
- Conductivity is out of range of the standards available from WHO.
- Turbidity of all the samples is more than 5 NTU which is an indicator to preserve the bacterial growth.
- Hardness as $CaCO_3$ is found "over range" in all samples which may have some potential health problems for the end suer.
- Temperature is more than $12^{O}C$ the WHO standard.
- Bacterial presence was found in all the samples at all locations.

Based upon the research carried out in this study it is finally concluded that the drinking water supplied to the users of the vicinity of QUEST cannot be considered "SAFE" for drinking. Therefore it is recommended that a critical monitoring of disinfectant like chlorine must be carried out on regular basis to prevent the potential health risks to the end consumers. The recommended dose of free residual chlorine 0.2-0.5 mg/l must be ensured to be found in the distribution line.

ACKNOWLEDGEMENT

The authors are thankful to Mr. Asif Ali Gadani and his group who helped dedicating their valuable time for collecting and testing of the samples in the laboratory of Department of Energy and Environment (EE), QUEST, Nawabshah, to complete this research effort. The authors are thankful to the staff of laboratory of EE department (QUEST) for their cooperation and technical support.

REFERENCES

[1] Raza, H., Husaain, M. & Alibai, K. Seasonal investigation of drinking water quality. In 24th WEDC international conference, Pakistan 1998. Retrieved and available on http://wedc.lboro.ac.uk/conferences/pdfs/24/Raza.pdf. Last viewed on 22.1.2013

[2] WB-CWRAS.,The World Bank 'Pakistan country water resources assistance strategy water economy: Running dry'. Washington DC: WB-CWRAS., November, 2005.

[3] Ndambuki, J. M. Water quality variation within a distribution system: a case study of Eldoret Municipality, Kenya. In: IASTED conference on environmentally sound technology in water resources management, 2006.

[4] Lienyao, L., Chungsying, L., & Shyang-Lai, K. "Spatial diversity of chlorine residual in a drinking water distribu tion system". Journal of Environmental Engineering, ASCE, 130, pp.1263-1268, 2004.

[5] Olivieri, V. P., Snead, M. C., Kruse, C. W., & Kawata, K. "Stability and effectiveness of chlorine disinfectants in water distribution systems". Environmental Health Perspectives, 69, pp.15–29, 1986.

[6] Anwar, M.S., Lateef, S., Siddiqi, G.M. Bacteriological quality of Drinking Water in Lahore". Biomedica, 26, pp.66-69, 2010.

[7] WHO Guidelines for drinking water quality. Geneva: World Health Organization, Vol. 1, pp. 1-29, 1993.

[8] Geldreich EE. Water borne pathogens invasions: A case of water quality protection in distribution. Proceedings of American Water Works Association Water Quality Technology Conference, pp.1-18, 1992.

[9] Pommervilli,J.C., "Alcamo'sFundamentals of Microbiology. 8th Edition Massachusetts: Jones and Bartlett Publishing 2007.

[10] Park, J.E., Park, K., "Text Book of Preventive and Social Medicines 13th Edition, Jabalpur, Banarasidas, Bhanot, pp.377-382, 1991.

[11] Akhtar T, Zai S, Zahoorullah Z. Quality of drinking water in NWFP. Pak J Med Res, 25, pp.74-84, 1986.

[12] Zaheer, S., Ghani., R., " Detection and enumeration of faecal coliforms and other microorganisms in drinking water: A comparison of two techniques. Journal of Pak Med Assoc, 35, pp. 329-34, 1985.

[13] Sheikh, M.R., Shar, A. Sheikh, S."Potability of water obtained through boring in Karachi." Journal of Pak Med Assoc , 44, pp. 286-87, 1994.

[14] Waheed, T., Kausar, T.,"Quality of drinking water in Lahore". Journal of Pak Med Res 26, pp.162-65, 1987.

[15] Anwar, M.S., Chaudrhy, N.A, Tayyib, M. "Qualitative assessment of bacteriological quality and chlorination status f drinking water in Lahore. JCPSP, 14, pp.157-160, 2004.

[16] APHA, Standard Methods for the Examination of Water and Wastewater. Washington: American Public Health Association.(20th edition), 1998.

[17] WHO, "pH in Drinking Water. Revised background document for development of WHO Guidelines for Drinking water Quality, WHO, 2007, No. WHO/SDE/WSH/07.01/1, 2007.

[18] PCRWR- DRIP , " Executive Summary: Drinking Water Assessment in Karachi, Hyderabad and Kotri", March 2007, pp. 2-7, March 2007.

[19] Farooque, S., Hashmi, I., Qazi, I.A., Qaisar, S., Rasheed, S., " Monitoring of coliform and chlorine residual in water distribution network of Rawalpindi, Springier, Journal of Environmental Monitoring Assessment, 140, pp.339-347.

[20] PCRWR, "National Water Quality Monitoring Program (NWQMP) Report, Annexure III-12(a) 2004. Retrieved from http://www.pcrwr.gov.pk

[21] Kelin, H., Yuang, H. F., Hong, L., & Robert, W. E., "Spatial variability of shallow ground water level, electrical conductivity and nitrate concentration and risk assessment of nitrate contamination in North China plain. Environment International, 31, pp. 896–903, 2005.

[22] Montgomery, J. M., Water treatment principles and design. New York: John Wiley & Son, 1985.

[23] Shar, A.H., Qazi, Y.F., Zardari, M., Soomro, I. H, (2009) Bacteriological Quality of Drinking Water of Sukkur City, Pakistan Journal of Medical research, Vol. 48, Issue 4, pp. 1-3, 2009.

ISSN 1605-8607

CONTENTS

Volume 11 No. 2 JULY - DEC, 2012

1. A Study Related to Protocols and Ports usage for the ISA Server
Pardeep Kumar, Iftikhar A. Koondhar, Akhtar H. Jalbani, Agha Shiraz A. Khan, M. Aamir Bhutto 1

2. Design, Challenges and Future Research Aspects of MAC Protocols for Wireless Sensor Networks
Pardeep Kumar, Akhtar H. Jalbani, M. Ibrahim Channa, Anees A. Soomro, Iftikhar A. Koondhar 10

3. Coin Recognition Based on Heuristic Segmentation and Image Fusion Via Bayesian Approach
Akhtar H. Jalbani, Pardeep Kumar, M. Ibrahim Channa, Chee Wai Mak, Aness A.Soomro 19

4. End-of-Pipe Waste Water Disposal and Quality Comparison of Old and New Sugar Mills in Sindh
Abdul Rehman Memon, Niaz Ahmed Memon, Muddasar Habib 30

5. Assessment of Raw Water Quality Supplied to Hyderabad for Drinking Purposes
Niaz Ahmed Memon, Abdul Rehman Memon, Noorullah Soomro 37

6. Assessment of Drinking Water Quality Supplied to QUEST Nawabshah, A Case Study
Roshan Zaman Sehto, Abdul Rehman Memon, Niaz Ahmed Memon 42

Published by:
Directorate of Research & Publication,
Quaid-e-Awam University of Engineering,
Science & Technology, Nawabshah,
District Shaheed Benazir Abad, Sindh- Pakistan.
Phone # 92-244-9370380 Fax # 92-244-9370362
email: jamali_sattar@yahoo.com, editor_rj@quest.edu.pk

Composed by:
Rano Khan Meghwar
Cell # 03002307151
email: megh.warrano@yahoo.com

Printed by:
Souraj Printers
7-Rabi Chamber, Haider Chowk,
Gari Khata, Hyderabad, Sindh.