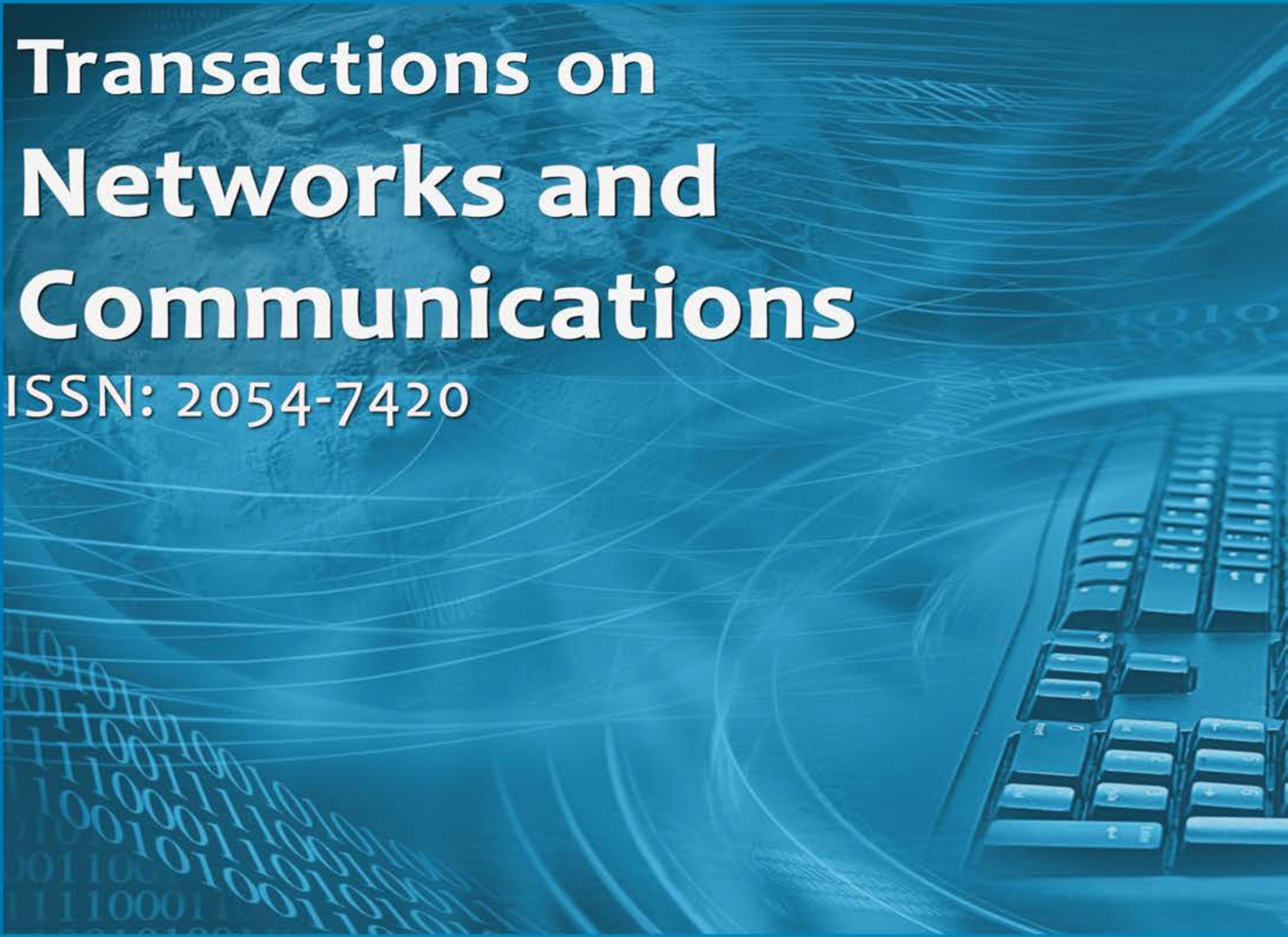
Volume 6 No. 5, October 2018



Transactions on Networks and Communications

ISSN: 2054-7420

Society for Sciences and Education - United Kingdom

TABLE OF CONTENTS

Editor In Chief

EDITORIAL ADVISORY BOARD

DISCLAIMER

All the contributions are published in good faith and intentions to promote and encourage research activities around the globe. The contributions are property of their respective authors/owners and the journal is not responsible for any content that hurts someone’s views or feelings etc.

TNC TRANSACTIONS ON NETWORKS AND COMMUNICATIONS

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM

VOLUME 6, NO. 5
ISSN: 2054 -7420

Efficient On-Line Traffic Policing in Software Defined Network

Lie Qian

Dept. of Chemistry, Computer & Physical Science, Southeastern Oklahoma State University, Durant, OK, USA
lqian@se.edu

ABSTRACT

On-line traffic such as conversational call and live video on the Internet are not pre-recorded and has no exact information about each session before the traffic happens. S-BIND (Confidence-level-based Statistical Bounding Interval-length Dependent) traffic model characterizes such traffic for QoS admission (GammaH-BIND) and policing purpose. A state-dependent token bucket based statistical regulator was proposed to police the traffic using S-BIND parameters. Recently, Software Defined Network (SDN) is proposed to decouple the control plane from the data plane, which enables low cost commodity design in switches and flexible network feature deployments through software implementation in centralized controllers. To deploy the state-dependent token bucket statistical regulator in SDN, extensions to the current SDN are needed. In this paper, design options for S-BIND traffic regulator in SDN switches are presented and analyzed. Among these options, a single meter design is chosen based on the cost and efficiency comparison between them. The needed switch implementation and OpenFlow protocol's extensions to realize the regulator in SDN is given at the end of the paper.

Keywords: Software Defined Network, OpenFlow, QoS, Policing

1 Introduction

After 30 years' fast growth, Internet is still growing in number of users, amount of data transmitted and data transmission speed. Internet is playing a more important role in economy, politics, education, entertainment and healthcare every day. Users are expecting more than simply retrieving desired information from the internet now days. Quality of Service (QoS) refers to the capability of a network to provide more than best effort service to chosen network traffic [1]. In best effort service, the network doesn't distinguish between packets from different users, or different applications. All packets are treated in the same way while the network devices try to forward every packet as fast as possible. There is no guarantee to the service received by any packets in the term of delay, packet loss rate, jitter and etc. A packet delivering Facebook page update is treated in the same way as a packet delivering video streaming on Netflix.

To fulfill QoS requirements, traffic in the network need to be controlled. A new traffic flow needs to be approved by a process called Admission Control [1, 13-20]. Algorithms are used in Admission Control to decide if the existence of the new traffic flow will affect all traffic's QoS. Traffic descriptor-based admission control [17-20] uses existing and new traffic's characterizations (parameters) to calculate the QoS that

DOI: 10.14738/tnc.65.4975
Publication Date: 14th September 2018
URL: http://dx.doi.org/10.14738/tnc.65.4975

could be received. To effectively characterize the network traffic with a set of parameters, especially for online traffic, which is not pre-recorded such as conversation call, live video, Confidence-level-based Statistical Bounding Interval-length Dependent (S-BIND) [1] model was proposed along with GammaH-BIND admission control algorithm. After being admitted, the traffic also need to be constantly monitored and regulated to ensure the real traffic going through the network be bounded by the declared S-BIND parameters. The regulator proposed in [2] is composed of a series of state-dependent token buckets. In [3], these token buckets are optimized to regulate (drop the excessive packets) the incoming traffic in a more efficient and accurate way. However, to deploy S-BIND together with the GammaH-BIND admission control algorithm and the state-dependent token buckets based regulator, there is a very common challenge in now days' network device industry: vendors of the network devices are reluctant to implement new services into their products before the services become widely accepted globally. On existing devices, making changes to deploy new services is not easy while traditionally forward functions (data plane) and control functions (control plane) are coupled together in network devices. [4-6]

Network functions are categorized into data plane and control plane. Data plane includes functions such as packet switching, dropping and modification. Control plane is responsible for routing, admission control, network monitoring, per-flow control, topology discovery, QoS support, etc. In traditional network devices, data plane and control plane are bounded together. Functions from both planes are implemented in the device by its vendor, which makes the control plane very difficult to change when new network service is needed on existing network devices. To deploy a new internet architecture, a new protocol or a new network policy, the control plane in every device needs to be reconfigured, tested, debugged and even rebuilt by its vendor.

Software Defined Network (SDN) is proposed to decouple the control plane from the data plane [7]. In SDN, data plane is still implemented by the device vendor in hardware while the control plane is realized by software in one or multiple centralized controllers. Software such as network management applications, network operating systems and OpenFlow protocols work together to control how packets are handled in each devices' data plane. Deploying new internet architecture, network management, services or protocols is simplified to software update in the controllers and no change is needed in the large number of traffic forwarding devices. Network services such as QoS, virtualization, security, traffic engineering, Information centered networking, forensic analysis, transferrable network applications, deep packet inspection, cloud data center, dynamic middle box deployment, virtual collaborative working environment, VLAN, etc. are becoming more realistic in SDN [8-9]. While many complicated control functions are moved to controllers, the majority network devices only need to deal with data plane which is relatively simple and standardized. This encourages low-cost commodity design in the data plane network devices, which lowers the cost, improves performance and enables cross vendor compatibility.

With SDN in the picture, deploying the on-line traffic S-BIND admission control [1] and traffic regulator solution [3] becomes more practical than before. The admission control could be implemented in the centralized controllers of SDN as software to make admission decisions. If admitted, new table entries for new flows are installed in the switches on the flow's route using OpenFlow protocol [10]. Controllers also need to install regulators in the ingress switches to regulate the flows. Current OpenFlow's meter design is not capable of providing needed token bucket's behavior as described in [3]. In this paper a solution to enable S-BIND characterized traffic's policing in Software Defined Network is proposed. First, couple meter design options are presented. The comparison and analysis of these design options lead to the choice of the single meter design option. To realize the chosen single meter option, needed OpenFlow protocol extensions and switches' meter implementation are presented the next.

The remainder of this paper is organized as follows. Section 2 presents the background review including S-BIND model together with its regulator and software defined network. In section 3 and 4, couple SDN switch token bucket design options are presented and compared. The new regulator design, implementation and protocol extensions are discussed in Section 5. Section 6 gives the future work and conclusion.

2 Related Works

In this section, a brief review of the S-BIND traffic model, its regulator, software defined network and OpenFlow is given. In this paper's network model, a centralized controller is used in a network domain. This controller serves as QoS controller and SDN controller. It is responsible for admission control decision and policies installation in the forward elements (switches) in the domain.

2.1 S-BIND Traffic Model

Binding traffic models use parameters to describe and bound the traffic volume. Traffic constraint function, denoted as b(t), is the essential part for binding traffic models. A network traffic flow is said to be bounded by b(t) if during any interval of length of u, the amount of traffic transmitted by this flow is less or equal to b(u),

In [1], authors developed a Confidence-level-based Statistical BIND (S-BIND) traffic model. The S-BIND model defines *p* time interval rate pairs as:

$$\{(R_k, I_k) \mid k = 1, \dots, p\} \tag{1}$$

where, $I_1 < I_2 < \dots < I_{k-1} < I_k$. I_k denotes the time interval length, and R_k is the rate, at which the flow is allowed to send in any period of I_k statistically. In S-BIND, random variable S_k is defined as:

$$S_k(a) = prob\left(\frac{A_j[t, t+I_k]}{I_k} \le a\right), \forall t \ge 0 \tag{2}$$

Here, $A_j[t_1,t_2]$ denotes the amount of arrivals in traffic flow within time interval $[t_1,t_2]$. S_k reflects the distribution of flow's transmission rate over time interval I_k and has density function $s_k(a)$. For each time interval I_k, R_k is defined in S-BIND by using random variable S_k's density function $s_k(a)$ as following:

$$\int_0^{R_k} s_k(t)dt = \varepsilon \tag{3}$$

ε is set between 0 and 1. The smaller ε is set, the smaller R_k will be, and the higher network utilization can be expected in admission control because more traffic will be admitted. For different kinds of on-line traffic such as conversation, videoconference, high motion video, low motion video, game data and etc., S-BIND parameters can be pre-defined with different confidence levels through experiments or statistical data analysis.

Along with GammaH-BIND admission algorithm developed in [1], S-BIND can achieve maximum valid network utilization for both low bursty and high bursty traffic.

2.2 Token Bucket Based Statistical Regulator

Admitted traffic flows need to be monitored and regulated during their lifetime so that they don't violate their declared S-BIND parameters. A token bucket based statistical regulator was originally proposed in [2] and later improved in [3]. In the solution, multiple state-dependent token buckets with dynamic token generation rates are cascaded together to regulate S-BIND traffic. One bucket is deployed for each linear segment in the constraint function b(t) which is derived from the S-BIND parameters. A bucket for a linear segment between (b_s,t_s) and (b_e,t_e) has the behavior given in Figure 1, where $\Delta b=b_e-b_s$, $\Delta t=t_e-t_s$, $r=\Delta b/\Delta t$, and b_{jump} is defined as:

$$b_{jump} = \begin{cases} b_s - \Delta b & if\ \Delta t < t_s \\ b_s - t_s r & if\ \Delta t \geq t_s \end{cases} \quad (4)$$

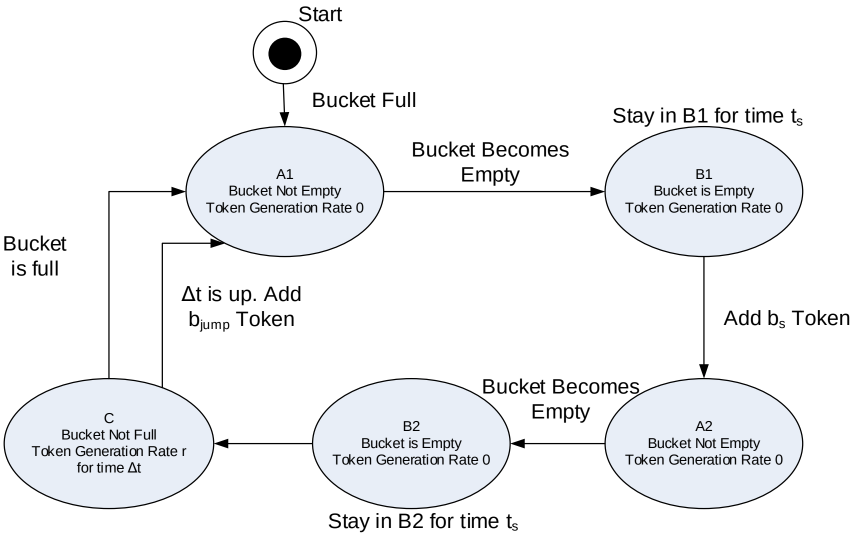


Figure 1: Token Bucket's Behavior

In the regulator designed in [3], it is possible that a bucket, say *B*, has its constraint function below another bucket *C*'s constraint function at all time interval t. In another word, bucket *B*'s output is always less than the amount of traffic allowed by bucket *C* at any time interval. In this case, there is no need for bucket *C* and an algorithm is developed in [3] to reduce the traffic policing burden in the regulator by removing *C*'s bucket.

Different from traditional deterministic token buckets which drop incoming packets when there is no enough token left in the bucket, the regulator proposed in [2] and [3] is allowed to forward packets even when the token is not enough statistically based on the confidence level *ε* specified in S-BIND traffic model. When a packet *p* with size *s* needs to be transmitted while the token left in bucket is *t<s*, the transmission probability of the packet is calculated as (5).

$$P_{transmit} = \begin{cases} \dfrac{(1-\varepsilon) - \dfrac{t_{neg}}{t_{neg}+t_{pos}}}{(1-\varepsilon)} & if \dfrac{t_{neg}}{t_{neg}+t_{pos}} < (1-\varepsilon) \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

In (5), the transmitting probability *P* mainly depends on the variables t_{neg} and t_{pos}, while *ε* is a confidence level parameter from S-BIND traffic model, which should be a constant for the life time of a bucket. t_{neg} is counted as the time when the token bucket has negative amount of token and t_{pos} is calculated as the total time minus t_{neg}.

2.3 Software Defined Network

Traditional network devices like switches bound data plane (packet forwarding, dropping and modification) and control plane (QoS, routing, network monitoring, per-flow control) in the same equipment. When any new network service, protocol or architecture need to be deployed, the control plane usually requires significant modifications. Even when the change could be accommodated by reconfiguration, still all devices need to be reconfigured by the network administrator. More often, the change needed is too significant and go beyond what is allowed in reconfiguration. In such cases new devices that support the new control plane need to be purchased and deployed if you are lucky to find them on the market while most vendors are reluctant to implement new service into their products before the service become widely accepted on the market, which is not always the case for newly proposed services, protocols or architectures.

Software Defined Network (SDN) was proposed to decouple the control plane and data plane [7]. The rationality is based on the fact that the data plane is relatively simple, stable and high performance demanding while the control plane is complicated, changes along with new services and requires flexibility. In SDN, the data plane remains in network devices like switches and the control plane is moved to one or multiple centralized controllers. The controllers install policies in the network devices to control how network traffic are processed in each switch. Network device vendors can focus on improving the data plane performance and lowering the cost of the devices and don't need worry about the constantly changing control plane [11]. Whenever the control plane needs to be changed, controllers could be reprogrammed (install a new software) to be capable of installing different policies in switches to realize the new services or protocols. With SDN, the network service that can be provided in a network is not limited by what the vendor implemented inside the devices anymore, which encourages and enables new network technology, services, protocols and architecture's design; as well as testing and deployment, services such as QoS, virtualization, security, traffic engineering, Information centered networking, forensic analysis, transferrable network applications, deep packet inspection, cloud data center, dynamic middle box deployment, virtual collaborative working environment, VLAN, etc. [12]

In a SDN, one or multiple controllers are deployed to make network management decisions. Network management applications, network operating systems and protocol interfacing the controller and switches are installed in the controllers. Network management applications is responsible for making network control decision. Network operating systems provide API to allow quick and easy network management applications programming and shield the network management application from the heterogeneous network technology. Openflow [10] protocol is installed in both switches and controllers. Openflow defines how the rules from the controller could be installed, updated, and removed in the switches and how these rules can be used in the switches for packet processing.

When the first packet of a new flow comes to a switch in the network domain. The packet could match no rule in the switch and be encapsulated and forwarded to a controller. The network operating system in the controller runs the network management application to process the packet and decides how this flow should be processed in the future. The decision is translated to a series of rules to be installed in on route switches in the network. Using OpenFlow protocol, these rules will be sent to and installed in the involved switches.

2.4 OpenFlow Switches

In an OpenFlow switch [10], there are one or multiple flow tables and one group table. Using OpenFlow protocol, the controller can add, modify and remove entries in the tables inside each switch. Each entry in any table consists of match fields, counters, and a set of instructions to apply to matching packets. Packets may need to be processed by multiple flow tables one by one when they come to a switch. When a packet is processed in a table, if a match is found, the associated instruction of that matched entry will be executed. If no match is found, a special miss entry in that table will determine the fate of the packet. The packet could be forwarded to the controller using Packet-in message, dropped or forwarded to the next flow table in the same switch depending on the policy installed in the miss entry. Controller can use Modify-State message to add flow entries in any table.

For a matched packet, instructions associated with the matched flow entry are executed. Some of the instructions edit the action set associated with the packet. All actions in the action set will be executed after the packet finishes its processing in the last flow table in the same switch. In addition to the actions in the action set, “Apply-Action” instruction could execute action of choice during the table’s processing. Instruction “Goto” specify which table is the next to process the packet. Instructions also can send metadata to the next table to assist the packet processing there. Actions are defined in OpenFlow to describe the forwarding, modification of the packet, applying meters to the packet, sending the packet to specific queue for QoS purpose, etc.

Meters are used in the switch to compare the incoming packets rate from one flow or a group of flows against certain predefined rate to decide if a packet should be forwarded or not. Action “meter” in the action set or action list (used in “Apply-Action” instruction) triggers a specific meter to be applied to a packet. One meter could be used against multiple flows. Also one packet from a flow could have multiple meters applied to it and the packet need pass all meter’s check to be forwarded. Inside a meter, there could be one or more meter bands. Each meter band has its own rate setting. When one packet needs to be processed by a meter, at most one meter band will be applied. The meter bands with the rates lower than current measured rate will be considered and among which the highest one will be chosen (if the measured rate is lower than all band rates, no meter band will be chosen). Meters together with associated meter bands are installed by SDN controllers through OpenFlow protocol messages.

3 Token Bucket Design in SDN

Enabling token bucket’s behavior as described in [2][3] is critical in developing online traffic regulator in a Software Defined Network. In this section first the SDN network management model is presented, which describes the overall architecture in SDN for new traffic admission control and traffic policing. Then three token bucket design options are discussed.

3.1 SDN Network Management Model

When the first packet of a new flow comes to the network, the ingress switch cannot find a flow entry (except the default no-match entry) matching this packet in its table and will forward this packet to the controller, which is defined in OpenFlow. Network management application in the controller extracts the S-BIND traffic parameters from the packet and performs admission control based on admission algorithms such as [1]. If the new flow is denied, a packet with cancelation flag will be sent toward the source and no new policy needs to be installed in any switch. If admitted, a packet with acceptance flag will be sent to the source of the flow. The controller also calculates how many token buckets are needed and the parameters of each bucket. Based on the calculation, table entries and meters will be installed in the ingress switch to regulate the new flow's traffic.

3.2 Regulator Design Options

As described in [2], regulators for S-BIND traffic are composed of multiple token buckets, which has state-dependent dynamic token generation rate and statistical packet forwarding. Each bucket is deployed for one linear segment on S-BIND's constraint function. Optimization algorithm proposed in [3] could reduce the number of buckets needed through removing redundant token buckets. Still multiple buckets are usually needed for each flow. Fortunately, meters are defined in OpenFlow.

Meters are used in OpenFlow to regulate traffic in SDN. Switch table entries could use instructions to apply meters to incoming packets. Multiple meters could be applied to the same packet. Vice versa, the same meter could be used to regulate traffic from a group of flows. S-BIND and its regulator are per-flow control. Therefore, one meter only applied to one flow for S-BIND traffic. Multiple meters could be needed for each flow to realize multiple token bucket's behavior.

The meter and meter bands defined in OpenFlow [10] at this moment can only control traffic using 2 simple parameters: rate and burst size. When the incoming traffic burst exceed the burst size, the rate is applied to decide if forward or flag the packets. The measuring window length (1ms, 100ms, 1s or others) are determined by the manufacturer and not modifiable in the controller on per flow base. Such simplicity in the meter band design prohibits it from realizing needed policing behavior in S-BIND regulator.

Each token bucket's behavior is depicted in Figure 1. There are 5 states in the bucket's behavior. Among the 5 states, states A1 and A2 have similar behavior, in which there is no token generated and bucket leaves these 2 states only when the remained token value hit zero. States B1 and B2 have similar behavior, in which there is no token generated and the bucket leaves these 2 states after t_s timeout elapsed. State C is different from all 4 states mentioned above. There is a token generation rate r in state C and bucket could leave state C in 2 scenarios; either timeout Δt runs out or the bucket become full of token (b_s amount of token in the bucket). Due to the different behaviors in these 5 states, there are 3 options to implement one bucket's behavior: in one, two or three OpenFlow meters.

3.2.1 Three-Meter Design Option

Based on the discussion above, all 5 states' behavior could be categorized into 3 groups as shown in Table 1. In the first candidate design option, 3 meters are used to implement one token bucket. One meter is deployed for states in each row in Table 1. One meter is used for state A1 and A2 (say meter A). One meter is used for state B1 and B2 (say meter B). One meter is used for state C (say meter C).

Table 1: Bucket's Behavior in 5 States

State	Token Generation Rate	Exit Condition
A1, A2	0	Token =0
B1, B2	0	Reach Timeout
C	r	Token=b_s OR Reach Timeout

Three meters, A, B and C, are setup in the ingress switch to realize the behavior of this bucket. Flow entries are added into four tables (say table ID 0-3) to apply meters for one bucket. The first table (table 0) doesn't apply any meter. It decides where the packet will go to (table 1-3). Table 0 is deemed as a master table for this bucket. Table 1 has the flow entry running instruction to apply action "meter A". Table 2 does that for meter B and table 3 for meter C. When a packet is matched in table 0, it will be forwarded to one of the 3 tables (ID 1-3) to be policed by one meter. OpenFlow protocol needs to be extended to include one new state field of type unsigned integer in table's flow entry to indicate where the incoming packet should be forwarded to (only table 0 will use it). One new instruction should be added into the protocol for table 0 to forward packet to the table with ID indicated by the new state field. The new instruction could be a variation of "GOTO" instruction currently defined in OpenFlow used to forward a packet to another table. The existing "GOTO" instruction requires an argument to specify the table ID [10]. The new version "GOTO" instruction could be executed without argument and use the new state field's value by default as the ID of the destination table.

The new state field should be initialized to 1 (if table 1 is used to apply meter A) by the controller during flow entry installation. One important modification to the switch is to allow the flow entry's new state field be accessed by meters. One link or address to the new state field needs to be added into OpenFlow's meter design. If the meter is used to police S-BIND traffic, the link will point to the new state field in the policed flow's master table's entry. All A, B, C meter's behavior should include the action of updating the new state field when the bucket leaves its current state, so that the next packet will be forwarded to a different table to be policed by another meter.

3.2.2 Two-Meter Design Option

In the second design option, the fact that state C is the only state with non-zero token generation rate is considered. Two meters are used to implement one token bucket. One meter is used for state A1, A2, B1 and B2 (say meter AB), where token generation is zero. One meter is used for state C, called meter C. Three tables (table ID 0-2) are used to realize one bucket's behavior. Table 0 still serves as the master table, which is responsible for forwarding the packet to either table 1 or table 2. In order to do that, the new state field is still needed in table 0 and is accessible to both AB and C meters. Table 0 can use the new argument-less "GOTO" instruction to forward packets to table 1 or 2 based on the new state field's value. Table 1 applies meter AB and table 2 applies meter C.

3.2.3 Single-Meter Design Option

The last design option in consideration is to use one meter (say meter ABC) for one token bucket. All 5 states' behavior are implemented in one meter. One table (any table is fine, not necessarily table 0) is used to apply the ABC meter. In this design, packets of the same flow will always be policed by the same ABC meter applied in the same table, therefore the new state field mentioned in previous 3.2.1 and 3.2.2's design options is not needed in the table entry. Also the associated extra access to the new state field from any meter is not needed anymore.

4 Design Option Analysis

The three design options (1-3 meter designs) described in section 3 require some extensions in current OpenFlow protocol and SDN switches. Extensions could be needed in table's entry formats and/or meter's implementation. In this section, the three design options are compared and analyzed in the terms of implementation complexity and cost. The results show that single meter implementation is the best choice.

4.1 Single Bucket Implementation Cost Analysis

In the single meter design option (as described in 3.2.3), there is no extra cost added in the switch table. The ABC meter keeps track of variables such as tokens left, timeout left, positive token time, negative token time, and state (token generation rate is a constant set by the controller, no need to be maintained). Positive token time and negative token time are used to track time duration in which the token amount is above or below zero. These two variables are used for statistical packet forwarding when there is no enough token to forward a packet [1]. Variable “state” is used to trace the state of the meter. The ABC meter could be in 5 different states and the variable uses values 0-4 to indicate the current state so that the meter could behave accordingly.

When there are 2 meters used for each bucket (Two-Meter design option in 3.2.2), 3 table entries need to be installed in 3 different tables for one bucket. One of them serves as the master table, and it needs a new field state in the entry to indicate where the packet should be forwarded to. The second and third table are meter tables, which apply meter AB or meter C respectively. In meter AB, variables for token left, timeout left, positive token time, negative token time, and state need to be maintained. State is used to distinguish states A1, A2, B1 and B2. In meter C, variables for token left, timeout left, positive token time, and negative token time need to be maintained (token generate rate is a constant set by controller, no need to be maintained).. Positive token time and negative token time should be shared between 2 meters, to achieve that, 2 more fields need to be added into master table’s entry to record these 2 values. Similarly, an extra token field needs to be added in the master table’s entry. It is used to transfer the remained token value from AB meter to C meter or vice versa. When AB meter leaves state B2 or C meter leaves state C, the meters will write the 3 values (positive token time, negative token time and token fields) to the master table’s entry. The next time when the meters start to process packets, they will copy the two values from the master table.

When there are 3 meters used for each bucket (Three-Meter design option in 3.2.1), 4 table entries needed to be installed in 4 different tables, one of which is the master table. Master table entry needs a new state field and 3 fields for positive/negative token time and token. Meter A needs to maintain variable token left, positive and negative token time. Meter B needs to maintain variables time out, positive token time, negative token time and a toggle bit. The toggle bit is used to distinguish between state B1 and B2, who has different exit behaviors (B1 switches to meter A and B2 switches to meter C). In meter C, variables for token left, timeout left, positive token time, and negative token time need to be maintained (token generate rate is a constant set by controller, no need to be maintained).

4.2 Single Bucket Cost Comparison

Table 2 summarizes cost for one bucket in all three solutions. Let’s first compare the Two-Meter design option and Three-Meters design option. Concerning the flow table cost, Three-Meter design needs to provide everything Two-Meter design needs and one extra table entry. Both design options need provide a C Meter. Three-Meter design’s B meter’s complexity is almost the same as AB meter in Two-Meter design (the only difference is that state in AB meter is an integer to distinguish between 4 states A1, A2, B1, and B2. Toggle bit in B meter is a boolean to distinguish between states B1 and B2). To gain the couple bits saving in B meter compare to AB meter, Three-Meter design has to throw in another A meter (with one more set of token left, positive time and negative time variables) which totally offsets the gain. Based on these comparisons, the Two-Meter design is obviously a better design than Three-Meter design.

Table 2: Cost Comparison Between 3 Design Options

	ABC Meter (Single Meter Design)	AB Meter +C Meter (Two-Meter Design)	A Meter + B Meter + C Meter (Three-Meter Design)
Table Cost	1 table entry	3 table entries	4 table entries
	No extra field	4 extra fields in master table	4 extra fields in master table
Meter Cost	token left positive time negative time timeout state	token left X 2 positive time X 2 negative time X 2 timeout X 2 state (only in AB Meter)	token left X 3 positive X3 negative time X 3 timeout X 2 (only B, C Meters) toggle bit (only B meter)

The comparison between Single-Meter design and Two-Meter design is not obvious in the meter part. 1-meter design has one very complicated meter which traces everything. Regarding to the variable tracing, ABC meter is the same as AB and one more than C meter (variable "state"). ABC meter needs to trace 4 states which complicates the control logic in the meter's implementation. However, Two-Meter design needs to implement all these control logics separated in 2 different meters. Therefore, regarding to control logic's complexity, Two-Meter design doesn't have clear advantage over Single-Meter design and it almost doubles the number of variables need to be traced (2 copies of token left, timeout, and positive/negative times in two meters)

Finally let's consider the flow table part between Single-Meter design and Two-Meter design. Single-Meter design doesn't need any modification on current OpenFlow's flow table design. No new field needed and only one table needed for flow traffic policing purpose. In Two-Meter design, more serious extra burden is that the new fields need to be accessed in the meter (both read and write). First this requires the meter to remember which table applies it. That adds one more address or link field in the AB and C meter and OpenFlow's meter setup message. Generally, a meter could be applied in multiple tables and link it with one specific table violate the switch design principle. Secondly, letting the meter access flow table's entry, even modify table's entry, could be denied by switch's architecture for security reason. Single-Meter design totally avoided such extra access to tables from meters and it is the obvious preferred solution for the design.

4.3 Regulator with Multiple Buckets

Now let's consider the case when multiple buckets are used to regulate one flow. When Single-Meter design is used, only 1 table is needed for all buckets. The same table entry can apply meters from all buckets. Therefore, the number of table entries and tables will not scale up with the number of buckets needed for a flow. When Two-Meter design is used, we have to use a different set of three tables for each bucket. If we try to use the same set of three tables, a packet will be either sent to one table to be policed by all AB meters from all buckets or sent to another table to be policed by all C meters from all buckets. This is wrong. Because of different token bucket sizes and different time out length between different buckets, entering and exiting states are not synchronized between buckets. It is possible that at a time, a packet needs to be policed by AB meters from some buckets and C meters from the other buckets. In this case, the master table will have to forward the packet to both tables and that will force the packet to be policed by both AB and C meters from all buckets, which first is unfairly too strict to the packet. Secondly it is incorrect because one packet's processing moves both AB and C meters' state forward (reducing token twice, countdown timeout twice, once in each meter), which could cause too many packets dropped and the flow is over policed. To solve this problem, a different set of 3 tables needed for each bucket. If there are *n* buckets needed, *3n* tables and *3n* table entries are needed. If Three-Meter design is chosen, *4n* tables and *4n* table entries will be needed for the same reason.

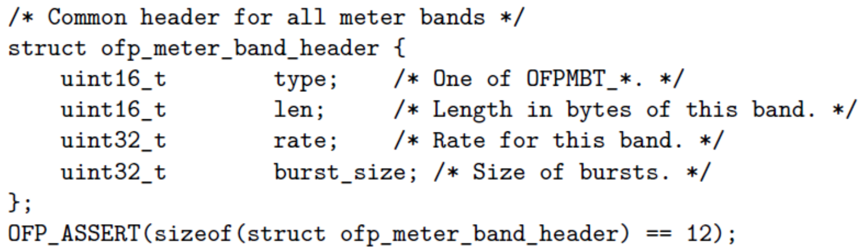
From the discussion above, the conclusion is that when the number of buckets needed in a regulator increase, Single-Meter design's table cost (number of table entries) is constant while Two and Three-Meter designs' cost are linear to the number of buckets. Again Single-Meter design is the obvious best choice.

5 Protocol Design and Meter Implementation

To implement the Single-Meter design discussed in previous sections in a Software Defined Network, the OpenFlow protocol needs to be extended to enable the installation of the new ABC meter in switches. ABC meter needs to be implemented in switches. This section presents the protocol extension and ABC meter implementation in SDN

5.1 OpenFlow Protocol Extension

To install meter bands together with a meter in a switch, the controller in the SDN sends a Meter Modification Message with the structure of opf_meter_mod to switches. The opf_meter_mod structure contains header, command, flags, meter id and a list of meter band head structures [10]. Each meter band header structure (defined as ofp_meter_band_header structure) in the list defines one meter band in the meter. To install an ABC meter, the opf_meter_mod structure will use the header, command, flags and meter id as they are defined in current OpenFlow protocol. Exactly one extended meter band header structure will be attached to the end of the meter modification message to install the newly defined meter band (say ABC meter band). No list needed, there is only one ABC meter band in an ABC meter.



```
/* Common header for all meter bands */
struct ofp_meter_band_header {
    uint16_t        type;    /* One of OFPMBT_*. */
    uint16_t        len;     /* Length in bytes of this band. */
    uint32_t        rate;    /* Rate for this band. */
    uint32_t        burst_size; /* Size of bursts. */
};
OFP_ASSERT(sizeof(struct ofp_meter_band_header) == 12);
```

Figure 2: Current Meter Band Head Data Structure

Currently ofp_meter_band_header structure is defined as Figure 2 [10]. First extension needed is to include the new ABC meter band type in the meter band type enumerates. There are 3 types defined already OFPMBT_DROP = 1, OFPMBT_DSCP_REMARK = 2, and 0xFFFF for experimenter. In the extended version, one new type OFPMBT_TOKEN = 3 is added in the ofp_meter_band_type enumerates for the new ABC meter band type. In an ABC meter band, several constants need to be setup by the controller as listed in Table 3. These constants' values are calculated by the controller based on the flow's S-BIND parameters and sent to switches in ofp_meter_band_header. Therefore, The fields listed in Table 3 need to be added into the ofp_meter_band_header structure. The new ofp_meter_band_header is shown in Figure 3.

Table 3: Constants Setup by Controller in a ABC Meter Band

Constants Field Name	Type	Description
b_start	uint32_t	b_s for the S-BIND segment
b_end	uint32_t	b_e for the S-BIND segment
t_start	uint32_t	t_s for the S-BIND segment
t_end	uint32_t	t_e for the S-BIND segment $r= \Delta b/\Delta t$
conf_lev	uint32_t	Confidence Level ε from S-BIND parameters

```
struct ofp_meter_band_header{
        uint16_t        type;
        uint16_t        len;
        uint32_t        rate;
        uint32_t        burst_size;

        /*new field for CS-BIND regulator*/
        uint32_t        b_start;
        uint32_t        b_end;
        uint32_t        t_start;
        uint32_t        t_end;
        uint32_t        conf_lev;
}
OFP_ASSERT(sizeof(struct ofp_meter_band_header) == 32
```

Figure 3: Extended Meter Band Head Data Structure

5.2 Meter Implementation

An ABC meter band is configured by the controller during meter setup with following constants: b_start, b_end, t_start, t_end, conf_lev, b_jump. Values of these constants are sent to the switch from the controller attached in oft_meter_band_header structure as described in Figure 3 except b_jump. Value of b_jump could be calculated from b_start, b_end, t_start and t_end as in formula (4) by switch.

Table 4: Variables Maintained in an ABC Meter Band

Variable Name	Type	Initial Value
token	int32_t	b_start
state	uint16_t	0
timeout	uint32_t	FFFF
p_time	struct ofp_time	0
n_time	struct ofp_time	0
Time	struct ofp_time	Current time
Rate	uint32_t	0

The ABC meter band also maintains variables during its life time to keep track of the status in the meter. These variables and their initial values are listed in Table 4. Struct ofp_time is based on the format defined in IEEE 1588-2008 with 2 subfields. A “seconds” field (64bit) and a “nanoseconds” field (32bit). “rate” is in kilobits per second. “timeout” is in millisecond.

The implementation of the meter band is given in Figure 4. The procedure described in pseudocode in Figure 4 is executed when this meter band is applied to a packet p. The meter band makes decision either to forward or drop the packet. All meter variables listed in Table 4 will be updated as described in the pseudocode.

```
Procedure PacketProcessing (packet p)
Step 1:
        if token>0, update p_time
        else update n_time
        if rate>0, update token
        if timeout<FFFF, update timeout
Step 2:
        if token>=sizeof(p)
                update token
                forward p
        else
                use formula (5) to make forward decision
                drop or forward p
                if forward, update token
Step 3:
        Update time with current system time
        Update state, token, timeout based on following
        if state=0 AND token<=0,
                state++, timeout=t_start
        if state=1 AND timeout=0,
                state++, token+=b_start, timeout=FFFF
        if state=2 AND token<=0,
                state++, timeout=t_start
        if state=3 and timeout=0,
                state++, rate = (b_end-b_start)/(t_end-t_start),
                timeout= t_end-t_start
        if state=4 AND timeout=0,
                state=0, token+=$b_{jump}$, timeout=FFFF, rate=0
        if state=4 AND token=b_start,
                state=0, timeout=FFFF, rate=0
```

Figure 4: ABC Meter Band Pseudocode

Each incoming packet is regulated by the ABC band in three steps. First step, time related variables are updated such as positive time, negative time, token amount and remaining timeout. In the second step, the forward or drop decision is made based on current token amount, packet size and positive/negative time if needed. In step 3, the band checks to see if switching state is needed. If needed, the state variable, which is used to indicate which one of the 5 states the band is in (A1, B1, A2, B2 and C), will be updated along with the variables that need to be reset during state switching.

6 Future Work and Conclusion

When there are multiple buckets' ABC meters for the same flow setup in a switch, the controller has total control over how to setup the apply instructions in the table to apply these meters. In paper [3], optimization is presented to optimize the number of buckets but no discussion about the sequence in which these buckets should be applied to regulate the traffic. In SDN, adapting the meter application sequence for different flows becomes possible. For a packet that is supposed to be dropped by the regulator, applying the meters in a proper sequence could help to drop most of such packets in first or second bucket's meter band and save the rest meter's processing cost. How the controller can determine such optimal meter application sequence for different flows will be the next step. With the optimization the real cost of the regulator such as CPU cost, memory cost, and communication cost could be evaluated through simulation and experiments.

In this paper, the design options for CS-BIND traffic regulator in SDN network are presented and analyzed. The single table, single meter design is the preferred choice due to its performance advantage and implementation efficiency. To implement the regulator in SDN, OpenFlow protocol needs to be extended to accommodate the new meter band type. The meter band's implementation is presented in pseudocode.

REFERENCES

[1] Lie Qian, Anard Krishnamurthy, Yuke Wang, Yiyan Tang, P. Dauchy, and Albert Conte, *A New Traffic Model and Statistical Admission Control Algorithm for Providing QoS Guarantees to On-Line Traffic*. Proceedings of IEEE Global Telecommunications Conference, 2004, GLOBECOM, vol. 3, pp. 1401-1405.

[2] Lie Qian, Yuke Wang, and Hong Shen, *Token Bucket Based Statistical Regulator for S-BIND Modeled On-Line Traffic*. Proceedings of IEEE International Conference on Communications, 2005, ICC, vol. 1, pp. 125-129.

[3] *Lie Qian, Efficient On-Line Traffic Policing for Confidence Level based Traffic Model*. Transactions on Networks and Communications (TNC), Vol. 4, No. 5, 2015, pp. 28-41

[4] L. Sun, K. Suzuki, C. Yasunobu, Y. Hatano, and H. Shimonishi, *A net- work management solution based on OpenFlow towards new challenges of multitenant data center*, in *Proc. 9th APSITT*, 2012, pp. 1–6.

[5] Sharafat, S. Das, G. Parulkar, and N. McKeown, *MPLS-TE and MPLS VPNS with OpenFlow*, *ACM SIGCOMM* Comput. Commun. *Rev.*, vol. 41, no. 4, pp. 452–453, Aug. 2011.

[6] N. Blefari-Melazzi, A. Detti *et al.*, *An OpenFlow-based testbed for information centric networking*, in Proc. Future Netw. Mobile Summit, 2012, pp. 4–6.

[7] H. Yin *et al.*, *SDNi: A Message Exchange Protocol for Software Defined Networks (SDNS) across Multiple Domains*, Jun. 2012, Internet draft. [Online]. Available: http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360.pdf

[8] W. Xia, Y. Wen, C. Foh, D. Niyato and H. Xie, *A Survey on Software-Defined Networking*, IEEE Communication Surveys & Tutorials, Vol. 17, no. 1, pp. 27-51, 1st Quarter 2015

[9] F. Hu, Q. Hao and K. Bao, *A Survey on Software-Defined Network and OpenFlow: From Concept to Implementation*, IEEE Communication Surveys & Tutorials, Vol. 16, no. 4, pp. 2181-2206, 4th Quarter 2014

[10] *OpenFlow Switch Specification, version 1.5.1*, Open Networking Foundation, March 26, 2015, [Online]. Available: https://www.opennetworking.org/wp-content/uploads/2014/10/openflow-switch-v1.5.1.pdf

[11] S.H. Yeganeh, A. Tootoonchian, and Y. Ganjali. *On scalability of software-defined networking*. IEEE Commun. Mag., 51(2):136–141, February, 2013.

[12] E. Kissel, G. Fernandes, M. Jaffee, M. Swany, and M. Zhang, *Driving software defined networks with xsp*. In SDN12: Workshop on Software Defined Networks, 2012.

[13] J. Qiu and E. W. Knightly, *Measurement-based admission control with aggregate traffic envelopes*, IEEE/ACM Transactions on Networking, vol. 9, no. 2, pp. 199-210, April 2001.

[14] B. Statovci-Halimi, *Adaptive admission control for supporting class-based QoS*, 2010 6th EURO-NF Conference on Next Generation Internet (NGI), pp. 1-8, 2010.

[15] L. Breslau, E. W. Knightly, S. Shenker, I. Stoica, and H. Zhang, *Endpoint admission control: architectural issues and performance*, Proc. Of ACM SIGCOMM'00, pp. 57-69, September 2000.

[16] V. Elek, G. Karlsson, and R. Ronngre, *Admission control based on end-to-end measurements*, Proc. Of IEEE INFOCOM, March 2000.

[17] E. W. Knightly, *H-BIND: a new approach to providing statistical performance guarantees to VBR traffic*, Proc. Of IEEE INFOCOM '96, pp. 1091--1099, March 1996.

[18] E. W. Knightly and N. B. Shroff, *Admission control for statistical QoS: theory and practice*, IEEE Network, vol. 13, Issue 2, pp. 20--29, 1999.

[19] H. A. Harhira, and S. Pierre, *A Mathematical Model for the Admission Control Problem in MPLS Networks with End-to-End delay guarantees*, Proc. Of 16th International Conference on Computer Communications and Networks, pp. 1193-1197, 2007.

[20] S. Alwakeel, and A. Prasetijo, *Probability admission control in class-based Video-on-Demand system*, 2011 International Conference on Multimedia Computing and Systems (ICMCS), pp. 1-6, 2011.

Society for Science and Education
United Kingdom

TNC Transactions on Networks and Communications

Volume 6, No. 5
ISSN: 2054 -7420

Load Balanced Network: Design, Implementation and Legal Consideration Issues

Abuonji Paul, Rodrigues Anthony, J., George O. Raburu
School of Informatics and Innovative Systems, Jaramogi Oginga Odinga University of Science and Technology, P. O. Box 210- 40601, Bondo, Kenya.
pabuonji@jooust.ac.ke; tonyr@jooust.ac.ke; graburu@jooust.ac.ke

ABSTRACT

Computer networks have become extremely useful in the modern fast paced work and business environments. Every user of a computing device- desktops, laptops, tablets and mobile phones- wants to connect to a network and communicate with others in near real-time. So networks have become largely ubiquitous. However many challenges exist for perfect network ubiquity. To many users, a network is not useful if they cannot access it when they need it. Many factors such as congestion, disconnection, misconfiguration, network loops, host-source outage and device errors lead to unavailability of a network or network enabled services. This study focused on the design, implementation and legal issues of load balanced networks in order to increase the availability of bandwidth supply. Several network designs were developed and tested situ in a real organization and resulting data used to make decisions on what adjustments to make in the subsequent designs until the best design was achieved in an iterative manner.

Index Terms: load balancing, network, bandwidth, firewall, router.

1 Introduction

Computer networks are very critical in the modern technology driven business environments where they are the facets that tie the entire business processes and applications supporting them into one synchronized and coherent whole. Unlike in the past when computers and networks were mainly used by university researchers for sending and receiving e-mails and by corporate employees for sharing printers [1], the situation has since changed drastically, and currently millions of ordinary citizens use computers and networks for their day to day activities like shopping, banking, studying, communication, entertainment and many others. With the advent and rapid development of social media and mobile digital communication devices, the Internet has become the perpetual global meeting place where young people discover, rediscover and develop their own personalities, influenced by millions of fellow young or older people all over the world [2]. This trend has put much pressure on computer networks and raised the bar significantly high for the caliber of network that would satisfy user needs [3].

Network security now receives unprecedented attention unlike the past. From the perspective of information system security, the three core principles of IS security involve maintaining confidentiality, integrity, and availability of information resources [4]. These three concepts form what is normally referred to as the CIA triad, depicted in figure 1 below. From the standpoint of balanced security [5], some systems such as those storing trade secrets have critical confidentiality requirements while some like

DOI: 10.14738/tnc.65.5099
Publication Date: 14th September 2018
URL: http://dx.doi.org/10.14738/tnc.65.5099

financial transaction values have critical integrity requirements whereas others like e-commerce servers have critical availability requirement. This explains why some organizations opt to change the CIA triad to the AIC triad to underscore the fact that they put more emphasis on availability.

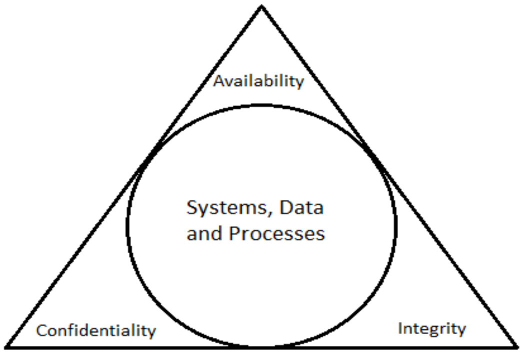


Figure 1: The Information System Security Triad

This viewpoint is further reinforced by the argument that a system must first be available for its confidentiality or integrity to be realized. That is why this diagram places it at the apex of the triangle, and for the same reason, the paper concentrates on how to enhance network security to improve its availability.

2 Related Works

Availability is the assurance that systems and data will reliably be accessed and used whenever needed by authorized users [6]. There are several threats that target system or network availability. They include denial-of-service or distributed denial-of-service attacks, worms, viruses which can clog the whole memory or CPU and render it ineffective and theft of physical computing devices [7]. Security and network administrators therefore need to implement systems with high level of availability. When considering the performance of communication lines, there are four main parameters that one needs to look at namely: bandwidth, delay, jitter and packet loss [8] since in a network, there are applications that require high bandwidth while others are more sensitive to delays or jitters. To succeed in enhancing system availability, an organization must develop appropriate technical mechanisms, security policies and well thought out contracts with external players such as suppliers, contactors and users.

Several controls can be implemented to safeguard availability of an information system and its resources such as redundant array of inexpensive disks (RAID), clustering, load balancing [9], redundant data and power lines, software and data backups, disk shadowing, co-location and off-site facilities, roll-back functions, fail-over configurations and service level agreements [5]. When dealing with outsourced services, the first aspect to consider is the quality of service level agreements (SLA) the organization negotiates and signs with its service providers [10]. Availability is normally computed in terms of mean time to failure (MTTF) and mean time to repair or restore (MTTR) service [11].

$$Node\ Availability = \frac{mttf}{mttf + mttr} \tag{1}$$

Calculating the MTTF and MTTR are fairly simple when dealing with one service like internet bandwidth; however it becomes a little complex when handling sophisticated distributed systems running multiple services. In such a case, to compute availability, those nodes need to regularly log its timestamps, enabling them to compute their uptimes and downtimes. To compute MTTF and MTTR the node reads the logged timestamps and uses that to calculate the averages for the time between the downtimes. Figure 2 below illustrates the time line of a node experiencing failure, indicating time to failure and time to repair.

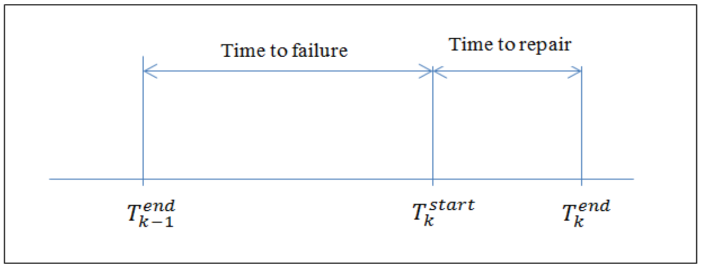


Figure 2: Time line of a node experiencing failure (Tanenbaum & Steen, 2014)

This gives node availability as expressed in the following formula:

$$Node\ Availability = \frac{\sum_{k=1}^{n}(T_k^{start} - T_{k-1}^{end})}{\sum_{k=1}^{n}(T_k^{start} - T_{k-1}^{end}) + \sum_{k=1}^{n}(T_k^{end} - T_{k-1}^{start})} \tag{2}$$

This must meet the minimum required service availability level, according to the SLA. Most high availability SLAs require at least 99.9 %, and can only compromise 0.1 %. Any service outage beyond this may require the service provider to give credit note to the client or pay for damages that occurred during the outage. For accurate data on service availability and down times to be accurately collected, there is need for constant monitoring and logging of the system [11].

The second approach for enhancing availability is by designing and robustly implementing a fault tolerant system. These systems have redundant processors, links, peripherals and software with fail-over or load balancing capacities [12]. Such systems can provide fail-safe capabilities that can enable them to operate at reasonable levels even if there is a major software or hardware failure. Redundancy and replication allows the system to have multiple components that can perform the same task so that if one component fails, the other components can take over the services. However, the limitation is that some components may remain idle while other are being used, thereby bringing about superfluous costs. An example is an organization that engages two internet service providers at the same time in order to mitigate the effects of service outages. If they opt for “redundant” or “backup link” approach, one link will be used until such a time when it fails then the second line will be plugged in. This approach has the advantage of being easy to implement, but very costly due to underutilization of the procured resources.

Another approach is the load balanced mode. Singh and Gangwar [13] defined load balancing as a methodology for distributing workload across multiple computers or other resources over the network links. When applied to management of network traffic, this approach allows multiple routes to the same place to be assigned to the traffic and will cause traffic to be distributed appropriately over those routes [14]. Suppose an organization has subscribed to bandwidth of 20 mbps form two different ISPs- where each ISP provides half of the total bandwidth, the system will be configured such that both the two ISPs supply their bandwidths to a common intelligent device like router or firewall with load balancing capabilities. The bandwidth will be distributed to all users from the common pool while the users will not know which of the two links is transmitting their data. However, when one link fails, all the users will automatically and seamlessly be redirected by the router or firewall to the existing link. All this should be transparent to the users except that they may notice a drop in transmission speed due to the fact that the available bandwidth is half the original bandwidth. This is illustrated in figure 3 below.

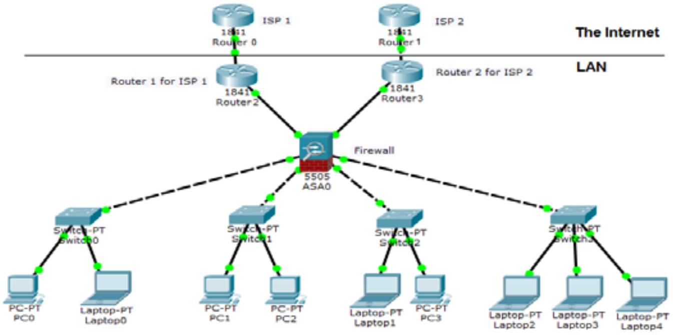


Figure 3: An Illustration of Bandwidth Supply in Load Balanced Mode

The system will need to be configured with two gateway IP addresses and domain IP addresses from both the two ISPs. Supposing the gateway IP for ISP 1 is 64.25.201.140 and that of ISP 2 is 41.206.114. 23, then the pseudo code for the gateway failover will be as follows:

```
If…
Not able to ping on IP Address "64.25.201.140"
Then…
Shift to another Available gateway
```

In this case the alternative gateway is 41.206.114. 23. Here none of the gateways should be set as backup because they are expected to be active all the time, and should be given the same weight if the bandwidth procured from each ISP is the same. The failover timeout must also be configured at reasonable time – that is neither too short nor too long for maximum efficiency. Too short a time will create high overhead due to frequent switching between gateway and a longer time may also create some window of service outage when one link fails but the other has not picked up.

The desired failover time can be expressed as follows:

$$Optimum\ failover\ time = \ t_{fmin} < t_f < t_{fmax} \tag{3}$$

Where:

t_{fmin} is the minimum time set for failover.
t_{fmax} is the maximum time set for failover.
t_f is the optimal time which must neither be too short nor too long.

Note that this design is meant to automate link stability vigilance and can greatly improve performance since it is self-organizing or self-adjusting [15]. The system should also log the events including up-times and down-times thereby making it easy to monitor and enforce the SLA. Much as this system looks better and more effective in improving availability, it requires more sophisticated network tools and advanced technical skills on the part of security or network administrators.

Many algorithms exist for load balancing implementation. They are normally referred to as packet scheduling algorithms. Patel and Dalal [16] broadly classifies them into time stamp based and round robin based algorithms. They explain that the latter is more efficient than the former because it eliminates time stamping and sorting overheads. Vashistha and Jayswal [17] broadly classified load balancing algorithms as static and dynamic. Under static algorithms, they listed round robin, randomized, central manager and threshold algorithms. While examples of dynamic algorithms included sender initiative, receiver initiative, symmetrically and periodically exchanged, central queue and local queue algorithms. Whereas Elngomi and Khanfar [18] also classified the algorithms the same way, they gave central queue and local queue algorithms as the only examples of dynamic algorithms.

However Kaur and Kaur [19] took a different approach. They outlined many load balancing algorithms without any classification. Their list included round robin, weighted round robin, throttled load balancer, active monitoring load balancer, adaptive resource allocation and skewness algorithms. Ray and Sarkar [20] also took a similar approach and listed load balancing algorithms as token ring, round robin, weighted round robin, randomized, central queue algorithm and connection based mechanism.

Several studies have been conducted to assess the performance of various load balancing algorithms. For example, Singh and Gangwar [13] compared the performance of round robin, active monitoring and throttled load balancer in a virtual environment in terms of response time. Their experimental results showed that if they increased the number of datacenter computers, this led to increase in overall average response time in all the algorithms under study. However, the increase in response time was least in throttled load balancing algorithm. This showed that it had better performance. However the algorithms were not tested for their effectiveness in managing network traffic.

Another comparative study was done by Sharma, Singh, and Sharma [21], and a similar one by Vashistha and Jayswal [22]. They identified performance parameters against which to test the algorithms. The parameters included: overload rejection, fault tolerance, forecasting accuracy, stability, centralized or decentralized decision making, nature of load balancing algorithms, cooperativeness of processors or hosts, process migration and resource utilization. The performance comparison is a shown in table 1 below. In summary, static load balancing algorithms were more stable, more accurate at forecasting and use fewer resources.

Table 1: Comparative Study of Performance of Various Load Balancing Algorithms: (Vashistha & Jayswal, 2013)

Parameters	Round Robin	Random	Local Queue	Central Queue	Central Manager	Threshold
Overload Rejection	No	No	Yes	Yes	No	No
Fault Tolerance	No	No	Yes	Yes	Yes	No
Forecasting Accuracy	More	More	Less	Less	More	More
Stability	Large	Large	Small	Small	Large	Large
Centralized/ Decentralized	D	D	D	C	C	D
Dynamic/ Static	S	S	Dy	Dy	S	S
Cooperative	No	No	Yes	Yes	Yes	Yes
Process Migration	No	No	Yes	No	No	No
Resource Utilization	Less	Less	More	Less	Less	Less

A review conducted by Elngomi and Khanfar [23] concluded that there is no particular load balancing technique that fits all information systems. Therefore system designers and developers must be able to carefully choose the technique that is suitable to a given system architecture and it requirements. For instance if the uncompromisable requirement of the system is stability, forecasting accuracy and optimum utilization of resources, then we must choose a static load balancing algorithm regardless of other performance indicators.

It is for this reason that most modern UTM systems with inbuilt load balancing capabilities use weighted round robin algorithm (WRR). This is a Round Robin based scheduling algorithm used in packet-switched networks with static weight assigned to queues of various connections. It was designed to mitigate the failures of the original round robin algorithm [24]. In this algorithm each connection is assigned a weight and those with higher weight receive more traffic than those with lower weights. But in a situation where all the weights are equal, all connections will receive uniformly balanced traffic. Figure 4 below illustrates how WRR works.

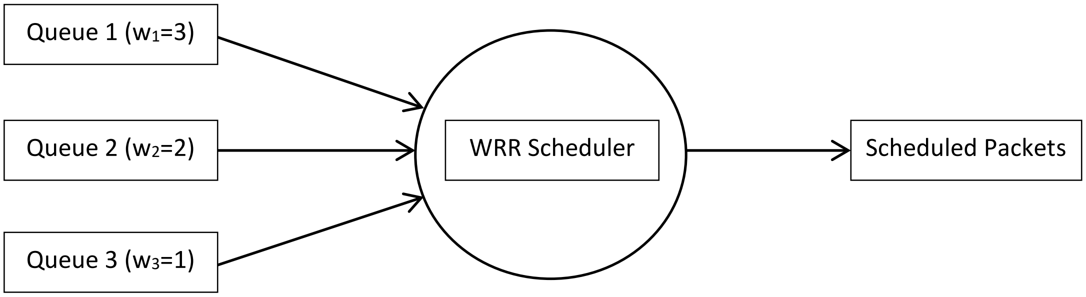


Figure 4: Weighted Round Robin (WRR) Algorithm

The algorithm allocates time for each queue and allows each to transmit data packets based on its weight. By so doing, it ensure that high priority queues do not suffer equal completion from lower priority queues but at the same time prevents lower priority queues from being starved of bandwidth for a long time. It has processing or computation complexity of O(1), thus making it feasible for high speed interfaces in both core and at the edge of computer networks [16].

WRR scheduling is based on assigning a fraction weight $\emptyset_i$ to each service queue such that the sum of weights of all service queues is equal to one.

$$\sum_{i=1}^{N} \emptyset_i = 1 \tag{4}$$

Considering that the weight is a fraction of the total number of packets to the transmitted, and we need to determine the number of integer packets to be transmitted from each queue, the fraction weight is then multiplied by a constant integer M. and the product is rounded off to nearest larger integer to obtain integer weight w_i. This integer weight value of each queue specifies number of packets to be transmitted from that queue. The total sum of these counter values is referred to as round robin length. Therefore the integer weight of i_{th} queue is:

$$w_i = \lceil \emptyset_i * M \rceil \tag{5}$$

The sum of existing *N* active connections in the network is defined as round robin length *W* and is given by the following formula

$$W = \sum_{i=1}^{N} w_i = M \tag{6}$$

As earlier stated, this algorithm provides stability, forecasting accuracy and good resource utilization. It has a drawback which is common to most scheduling algorithms that latency is affected by transmission rate of output link and the number of connections [16].

3 Methodology

The study adopted descriptive and diagnostic research design. Descriptive study was used to collect and record primary data depicting the problems, issues or concerns within the system under study [25], [26] while diagnostic study was used to facilitate an in-depth analysis of the research variables by first investigating the root cause of the problem and dealing with it in terms of emergence of the problem, diagnosis of the problem, solution for the problem and where no concrete solution has been found, a suggestion for the problem solution or escalation [27]. A network infrastructure was designed and deployed in a university. The process was guided by the Prepare Plan Design Implement Operate Optimize (PPDIOO) Network Life Cycle [28]. This design methodology was adopted because it was developed to support evolving networks like the one used in the research. Three Cisco routers, several switches and a cyberoam firewall were configured and used at various levels of the study. Two ISPs supplied internet bandwidth. The supply and use of the bandwidth was monitored to test availability. In-situ data was captured by real-time bandwidth monitoring tools deployed at the gateway to monitor bandwidth availability in terms of stability and amount supplied. This data was presented graphically and compared for different designs and results used to make decisions on what adjustments were supposed to be made on the current design.

4 Load Balanced Network Topologies and Architectures

Several network topologies and architectures were developed and tested for security and bandwidth load balancing using two ISPs.

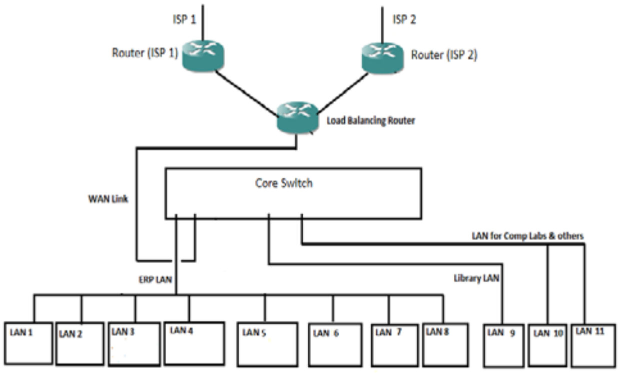


Figure 5 (a): LAN Topology with two ISPs and Load Balancing on Router

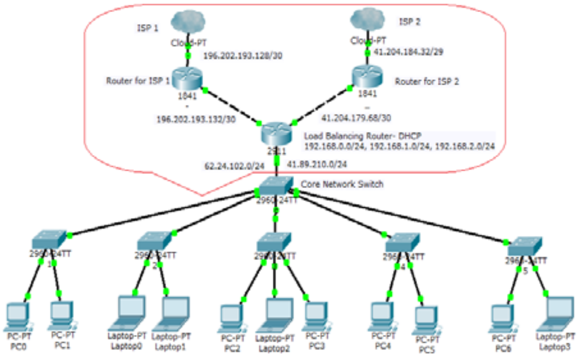


Figure 5 (b): LAN Architecture with two ISPs and Load Balancing on Router

Figures 5 (a) and (b) above illustrate a topology and architecture respectively of a network with two ISPs, three routers and a core network PnP switch. In this arrangement, the link from each ISP terminates on a different router. Then the third router is dedicated to perform load balancing functions such as aggregation of traffic from or to the two ISPs. The load balancing router performs DHCP, NATing and routing functions. This arrangement though was effective in improving link availability still left the LAN in desperate need for security- both at the gateway and within.

As illustrated in figures 6 and 7, this configuration worked well except that the network was vulnerable to cybercrime activities since there was no firewall. The load balancing router performed NATing, DHCP as well as managing the traffic load on both links. This setup was a major achievement as it improved availability by providing two simultaneous links- whereby if one link fails, all the users' traffic is seamlessly and transparently routed in and out through the remaining link. However the main concern was that the absence of firewall still exposed the network to numerous external threats. Besides, the network still had one broadcast domain containing three pools of /24 IP addresses provided by three DHCP servers with a single default gateway- 192.168.2.0. There was dire need to address these issues so a new design was proposed.

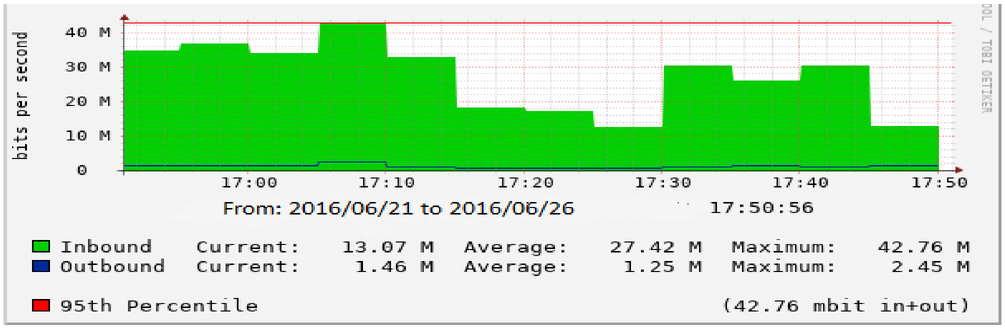


Figure 6: Traffic on ISP 1

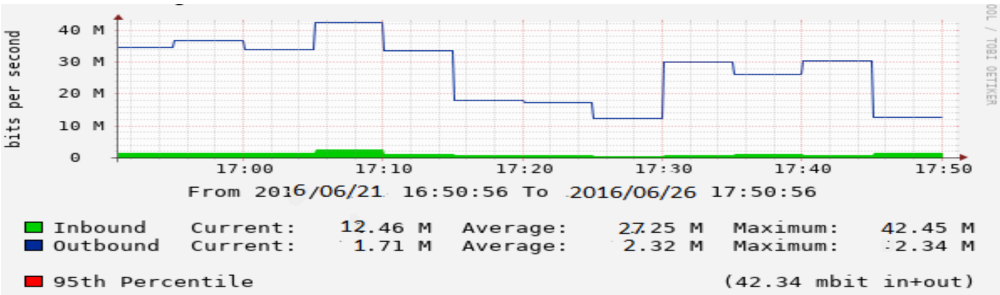


Figure 7: Traffic on ISP 2

The need for security triggered a desire to develop a system that would provide a high availability link as well as border and internal LAN security. This led to the topology and architecture in figures 8 (a) and (b) below respectively. In this set up, the load balancing was done by the third router that aggregates traffic from/ to both ISPs. However, instead of this router feeding into the LAN directly, it was connected to a UTM configured in gateway mode.

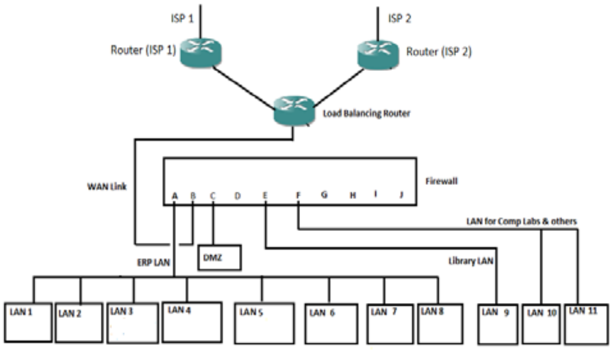


Figure 8 (a): LAN Topology with two ISPs, three routers and Firewall

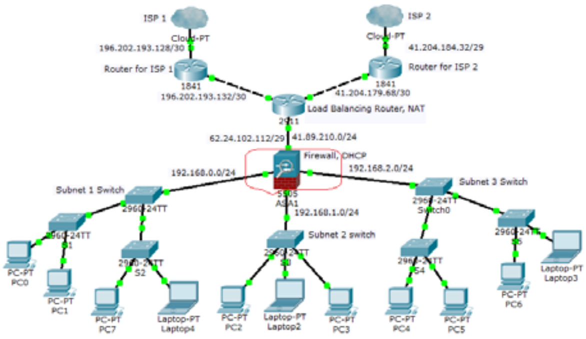


Figure 8 (b): LAN Architecture with two ISPs, three routers and Firewall

Even though this design improved the security of the network, it had considerable operational challenges due to configurational intricacies. The UTM was implemented in gateway mode in order to utilize all its required features (as opposed to bridge mode which is limited in many ways). In this gateway mode, the UTM could take only one IP address as gateway on its WAN interface. This meant that either (not both) of the ISP gateways could be used at a time. The UTM was configured to provide security as well as perform NATing, subnetting and DHCP functions. However load balancing activities were supposed to be performed by the "load balancing router" as was the case in design 2 (a). The justification of this design was to distribute network tasks so that no single device is overworked since this could reduce network efficiency, after all the intention of the design was distribution and balancing of network load. Nevertheless, this design did not work as expected. Figure 9 below shows graphs illustrating how traffic was transmitted through the two network links ISP 1 and ISP 2 if the gateway on UTM is from ISP 2.

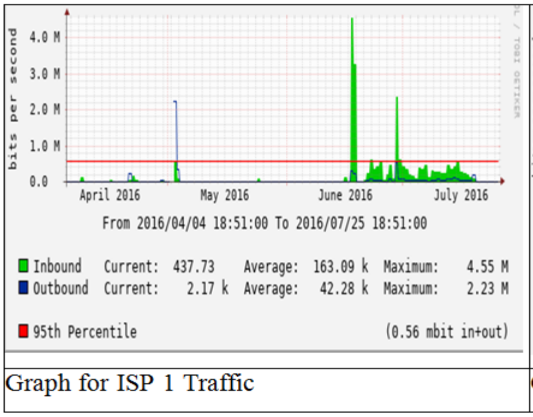
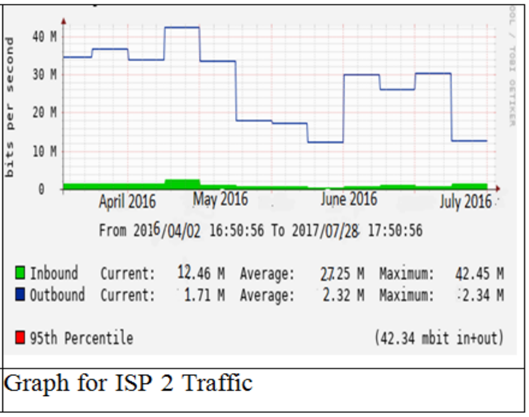
	
Graph for ISP 1 Traffic	Graph for ISP 2 Traffic

Figure 9: Data Traffic in Design flow

It was observed that if the gateway on the WAN interface of the UTM was from ISP 2, all the date exiting the LAN would be transmitted through ISP 2's network and vice versa. Even though the load balancing configuration on the router remained the same, the added layer of security and associated configurations had inadvertently invalidated some aspects of the routers configurations. With this configuration intact, the only way to remedy the situation was to configure other upstream routers of the two ISPs to allow Border Gateway Protocol (BGP) to enable the exchange of routing and reachability information between the two autonomous ISP systems involved. However this would have been a complex process which required even more sophisticated configurations, prolonged boardroom negotiations and possible legal implications for one of the two ISPs since it was not strictly a commercial ISP but one that supplied bandwidth to educational institutions and recognized as such by the government.

This challenge motivated the researcher to design another topology and architecture that would take advantage of the salient feature of the previous set ups but mitigate their weaknesses. Figures 10 (a) and (b) below depicts the new arrangement. A public switch was introduced between the load balancing router and the UTM. Date emanating from the two ISPs was aggregated at the load balancing router and then passed on to the public switch which was connected to two different WAN interfaces of the UTM.

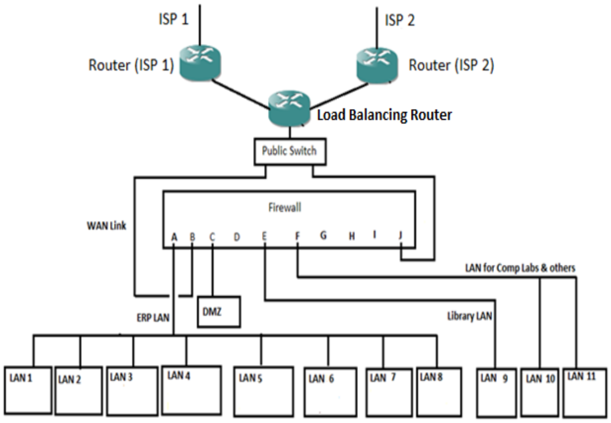


Figure 10 (a): LAN Topology with two ISPs, three routers, firewall and public switch

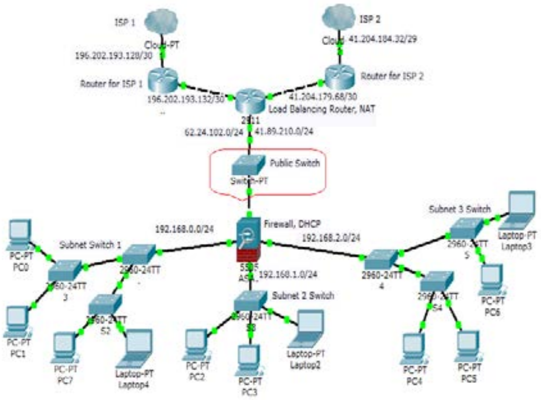


Figure 10 (b): LAN architecture with two ISPs, three routers, firewall and public switch

This topology initially alleviated some of the problems of the previous topologies. At first the internet access speed from the LAN increased tremendously as was expected since the two links were being used simultaneously and the LAN had also been subnetted thereby creating many smaller collision domains and decreasing the broadcast domain. However this was as far as the success went. This setup added more confusion to the system. Byzantine and transient faults characterized the design. For example in one office with six users, four users would get internet access whilst two could not. When one removed the Ethernet cable from the NIC port for a short while and reconnected it then that would solve the problem in one case but did not in another case. Technically speaking, byzantine faults are difficult to troubleshoot and solve since it is difficult to identify the lucid cause. Because of these problems, a new design was developed and implemented.

5 The Ideal Load Balanced Network for Security and High Availability

Due to the problems described above, a fundamentally different design was developed to solve them. This structure used two routers and the UTM. The load balancing function was configured in the UTM instead of the router. Other functions such as NATing, DHCP, subnetting and security were also configured on the UTM. The downside was that this configuration overloaded one devise. However the upside is that this was the configuration that could effectively solve the problem. The topology and architecture are illustrated in figures 11 (a) and (b) below respectively.

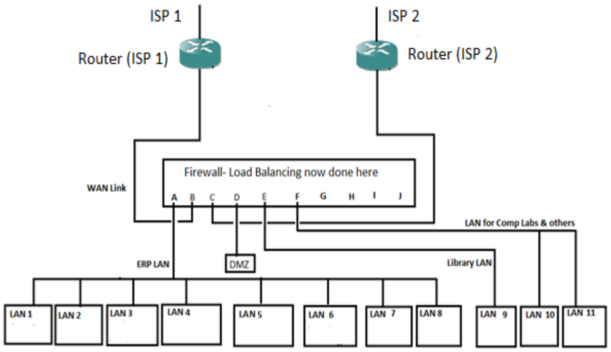


Figure 5 (a): LAN topology with two ISPs, two routers and firewall

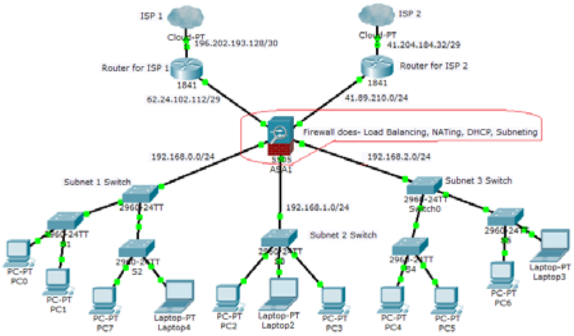
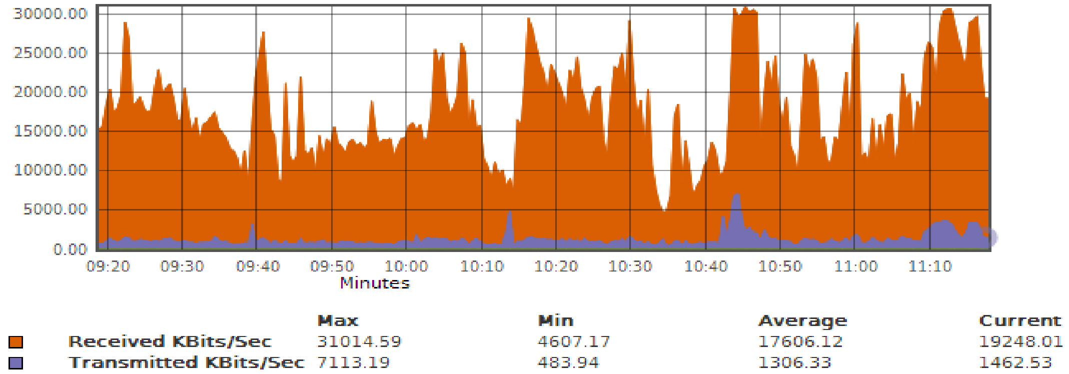


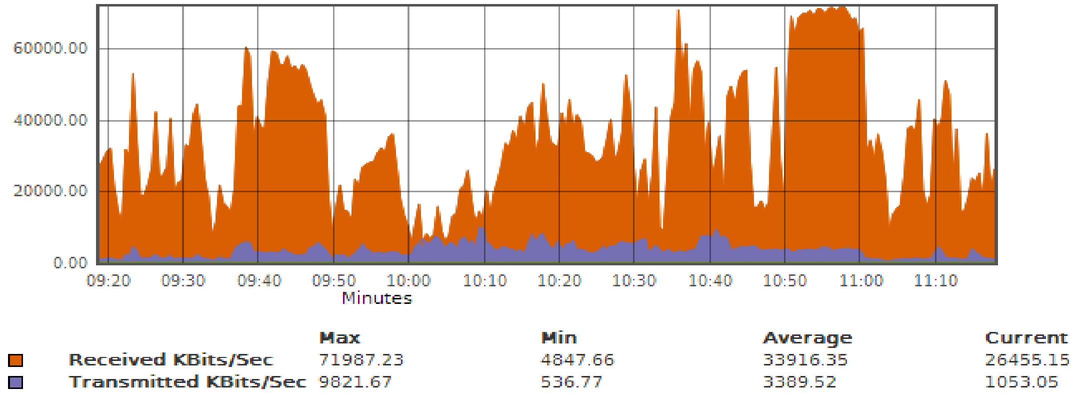
Figure 5 (b): LAN architecture with two ISPs, two routers and firewall

Figures 12 and 13 below are graphs viewed from the UTM WAN interfaces illustrating the transmission of traffic through the two network links ISP 1 and ISP 2 respectively.



		Max	Min	Average	Current
🟧	Received KBits/Sec	31014.59	4607.17	17606.12	19248.01
🟪	Transmitted KBits/Sec	7113.19	483.94	1306.33	1462.53

Figure 12: Data Transmission through ISP 1



		Max	Min	Average	Current
🟧	Received KBits/Sec	71987.23	4847.66	33916.35	26455.15
🟪	Transmitted KBits/Sec	9821.67	536.77	3389.52	1053.05

Figure 13: Data Transmission through ISP 2

The graphs demonstrated that the load was evenly distributed between the two internet links as was expected. This design provided both security and load balancing services to the LAN. However it still had a few limitations which were ingrained in design weakness of the UTM used. One major weakness was that whereas the UTM provided for many simultaneous WAN connections, its design only had provision for three DNS IP addresses as opposed to four required for a good UTM that can take two ISP connections. The second weakness emanated from the algorithm used, as illustrated in the following pseudo code.

```
If…
Not able to ping on IP Address "64.25.201.140"
Then…
Shift to another Available gateway
```

Note that there are instances when the internet bandwidth may not be available but the ISP's gateway is reachable. In such cases this pseudo code states that the UTM users connected through that particular gateway should not be switched over to another gateway. This explains why there were instances in the network when one connection failed and the failover to the next gateway was not seamless. Figure 14 below illustrate three different scenarios.

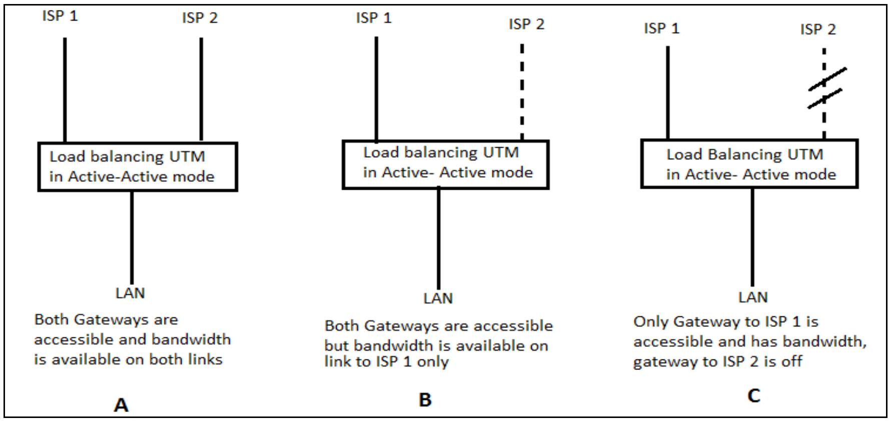

ISP 1
ISP 2
Load balancing UTM in Active-Active mode
LAN
Both Gateways are accessible and bandwidth is available on both links
A
ISP 1
ISP 2
Load balancing UTM in Active- Active mode
LAN
Both Gateways are accessible but bandwidth is available on link to ISP 1 only
B
ISP 1
ISP 2
Load Balancing UTM in Active- Active mode
LAN
Only Gateway to ISP 1 is accessible and has bandwidth, gateway to ISP 2 is off
C

Figure 14: Load Balancing Failover Scenarios

In scenario A, internet users in the LAN are all expected access the internet as they required. The load will be distributed to the two ISP links by the UTM in the ratio of 1:1. In scenario B, the internet users who were connected to ISP 1 will access the internet as they required, but the users connected to ISP 2 will not access the internet until either of the following two actions are taken: (i) they disconnect their Ethernet cables from their computers NICs for a short while and reconnect them so that the UTM connects them to the internet through ISP 1 or (ii) the network administrator disconnects the cable connecting the UTM to ISP 2 so that the UTM will not be able to ping the gateway of ISP 2, in which case all the internet users in the LAN will be connected to ISP 1. Finally in scenario C, as soon as the gateway to ISP 2 goes off, the UTM will seamlessly and automatically connect all internet users in the LAN to ISP 1.

6 Discussion

There are many factors that affect the stability of bandwidth supply in a network. Some of the factors are beyond the control of the organization receiving the service and thus the guaranteed can only be got by negotiating water-tight contracts and SLAs with the service providers. However for those organizations that operate in remote places, far away from the internet backbone cable, it is prudent to adopt link replication as a reinforcement to the administrative controls to enhance link availability. In this process, link redundancy acquired for backup in not advisable due to superfluous costs that come with it- for example the primary link will be used exclusively or largely while the backup or secondary link is either left idle or underutilized until the primary link fails. The ideal approach therefore is load balanced connection which, even though comes with advanced and sophisticated technological demands, will provide a better solution at economically justifiable cost. Each of the four designs described above has its strengths and weaknesses as earlier explained. The final design depicts a network in which bandwidth from the two links are aggregated and redistributed in a LAN network that is also protected by a gateway firewall.

7 Conclusion

We have demonstrated that if an organization requires high internet bandwidth availability in its LAN which is protected by a gateway or border firewall and therefore decides to be connected to two IPSs simultaneously whereby one of the ISPs is a commercial ISP (ISP 1) but the other is not an ordinary commercial ISP (ISP 2) which is restricted from carrying any traffic of commercial ISPs by its contractual or legal obligation to a third party such as the government or any other regulatory authority, so that ISP 2 cannot configure its BGP policies to allow traffic from ISP 1 to transit through ISP 2's network, then the best configuration for load balancing is to use a firewall with load balancing capabilities.

REFERENCES

[1] Tanenbaum, A. S. (2011). Computer Networks; 4th ed. Prentice-Hall, Inc: New Jersey.

[2] Greenwald, G. (2014), No Place to Hide: Edward Snowden, the NSA & the Surveillance State; Penguin Random House, UK.

[3] Membrey, P., Plugge, E. & Hows, D. (2012); Practical Load Balancing: Ride the Performance Tiger: The ACM Digital Library; Apress Berkely, CA, USA ©2012 ISBN:1430236809 9781430236801

[4] Cocca, P. (2004). SANS Institute InfoSec Reading Room: Email Security Threats. Retrieved on 17th November, 2012 from: http://www.sans.org/reading_room/whitepapers/email/emailsecurity_threats_1540

[5] Harris, S. (2013), All in One CISSP. McGrow-Hill: New York

[6] Stallings, W. (2011). Network Security Essentials: Applications and Standards, 4th Ed; Pearson Education, Inc: Prentice Hall

[7] Laudon, K. C. & Laudon, J. P. (2012). Management Information Systems: Managing the Digital Firm, 12th ed. Pearson Education Limited: Edinburgh Gate, Harlow.

[8] Orzach, Y. (2013), Network Analysis using Wireshark Cookbook; PackT Publishing: Birmingham- Mumbai.

[9] Zhang, J. *et al*. (2018), Load Balancing in Data Center Networks: A Survey; *IEEE Communications Surveys & Tutorials (Early Access); Electronic ISSN: 1553-877X; CD-ROM ISSN: 2373-745X. Retrieved on: 15th August, 2018; from: https://ieeexplore.ieee.org/document/8316818/*

[10] Stewart, J. M., Tittel, E. & Chapple, M. (2005), CISSP: Certified Information Systems Security Professional Study Guide; 3rd ed. Sybex Inc.: London

[11] Tanenbaum, A. S. & Steen, M. V. (2014), Distributed Systems: Principles and Paradigms, 2nd ed. Edinburg Gate: Pearson Education Limited.

[12] O'Brien, J. A. & Marakas, G. M. (2011). Management Information Systems, 10th ed. McGrow- Hill/ Irwin: New York.

[13] Singh, H. & Gangwar, R. C. (2014), Comparative Study of Load Balancing Algorithms in Cloud Environment: *International Journal on Recent and Innovation Trends in Computing and Communication; ISSN: 2321-8169; Volume: 2 Issue: 10; PP: 3195 – 3199*

[14] Peterson, L. L. & Davie, B. S. (2007). Computer Networks: A systems Approach, 4th ed. Elsevier, Inc.: San Francisco.

[15] Ma, Y., Chen, J. & Lin, C. (2018); Automated Network Load Balancing and Capacity Enhancing Mechanism in Future Network: *IEEE Access, Volume 6.*

[16] Patel, Z. & Dalal, U. (2014), Design and Implementation of Low Latency Weighted Round Robin (LLWRR) Scheduling for High Speed Networks: *International Journal of Wireless &Mobile Networks (IJWMN):* Vol. 6, No. 4.

[17] Vashistha, J. & Jayswal, A. K. (2013), Comparative Study of Load Balancing Algorithms: *IOSR Journal of Engineering (IOSRJEN) e-ISSN: 2250-3021, p-ISSN: 2278-8719 Vol. 3, Issue 3; PP 45-50*

[18] Elngomi, Z. M. & Khanfar, K. (2016), A Comparative Study of Load Balancing Algorithms: A Review Paper: International Journal of Computer Science and Mobile Computing; *Vol. 5, Issue. 6, June 2016, pg.448 – 458*

[19] Kaur, P. & Kaur, D. (2015), Efficient and Enhanced Load Balancing Algorithms in Cloud Computing: *International Journal of Grid Distribution Computing:* Vol.8, No.2 (2015), pp. 9-14.

[20] Ray, S. & Sarkar, A. D. (2012), Execution Analysis of Load Balancing Algorithms in Cloud Computing Environment: International Journal on Cloud Computing: Services and Architecture (IJCCSA), Vol.2, No.5, October 2012.

[21] Sharma, S., Singh, S. & Sharma, M. (2008), Performance Analysis of Load Balancing Algorithms: *International Journal of Computer, Electrical, Automation, Control and Information Engineering Vol. 2, No: 2, 2008.*

[22] Vashistha, J. & Jayswal, A. K. (2013), Comparative Study of Load Balancing Algorithms: *IOSR Journal of Engineering (IOSRJEN) e-ISSN: 2250-3021, p-ISSN: 2278-8719 Vol. 3, Issue 3; PP 45-50*

[23] Elngomi, Z. M. & Khanfar, K. (2016), A Comparative Study of Load Balancing Algorithms: A Review Paper: International Journal of Computer Science and Mobile Computing; *Vol. 5, Issue. 6, June 2016, pg.448 – 458.*

[24] Daryapurkar, A. & Deshmukh, V. M. (2013), Efficient Load Balancing Algorithm in Cloud Environment: *International Journal Of Computer Science And Applications:* Vol. 6, No.2

[25] Kothari, C. R. & Garg, G. (2014): Research methodology: Methods and techniques. New Delhi: New Age International (P) Ltd, Publishers.

[26] Nzioki, P.M., Kimeli, S. K., Abudho, M. R., Nthiwa, J. M. (2013): Management of working capital and its effects on profitability of manufacturing companies listed at NSE, Kenya: *International Journal of Business and Financial Management Research*, *1:35 - 42.*

[27] Farooq, U. (2013), Types of Research Design. *Referred Academic Journal, 08:21*

[28] Cisco Study Guide (2007), Designing Cisco Network Service Architecture. Retrieved on 24th June 2017, from: www.itsolutions.pro%2Fimages%2Fstories%2Fdocs%2Fdesigningcisconetworkservicearchitecturesarchv2.0volume1.pdf&usg=AFQjCNEjONV tN0daSRWcsOfDYIDc0FOpbw

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM

TNC TRANSACTIONS ON NETWORKS AND COMMUNICATIONS

VOLUME 6, NO. 5
ISSN: 2054 -7420

Secured Communication through Wireless Sensor Network

[1]Talal Alkharoubi, [2]Abdullatif Albaseer, [3]Gamil Ahmed
[1,2,3]King Fahd University of Petroleum & Minerals
Computer engineering department
Dhahran, 31261, Saudi Arabia
talalkh@kfupm.edu.sa; Abdullatif2009@gmail.com; g201302310@kfupm.edu.sa

ABSTRACT

Nowadays, life seems to be deficient without Internet. The Internet of Things (IoT) is heavily affecting our daily lives in many domains, ranging from tiny wearable devices to large industrial systems. In face of this rapid improvements, security threats and privacy issues also have brought critical challenges in designing and implementing such applications. In this work, we aim to find a way to protect the data from the WSNs devices to the cloud server or control unit. Also, the task of data management in WSNs is a vital issue that can be performed with limited resources such as processing, memory and energy. So, we have proposed a light implementation for AES 128 key in order to be used to encrypt the data sent by these sensors. We have adopted three different platforms which are Sky and Z1 motes to test this algorithm. Applying such algorithm leads to consume more power but guarantees a secure communication against malicious nodes.

Keywords: IoT, Threat, Attack, WSNs, AES, TelosB, Z1, Contiki, Cooja.

1 Introduction

Internet of Things (IoT) is a concept that enables various physical objects and methods of communication to achieve a certain task by exchanging information. IoT exploits underlying technologies to make these objects much smarter such as wireless sensor networks (WSNs), applications, Internet protocols and ubiquitous and, embedded devices. (Al-Fuqaha et al. 2015).

Nowadays, the IoT has brought the opportunities to have a smart home and business applications which contribute to increase the quality of our life and grow the world's economy. From a functionality viewpoint, IoT technology promises to improve our life style and make our lives easy and comfortable. IoT depends on the internet infrastructure in order to reach its goals. Consequently, it opens itself into all the known conventional cyber-security threats as well as it opens doors for new threats. Cybersecurity attacks towards IoT not only endanger the IoT functionality but also it may expose human lives to risks in its environments due to security breaches. Even though a plenty of security defense and countermeasures have been developed since the early stage of the internet era, these solutions cannot be applied directly into IoT infrastructure due to a major difference in computation capability between conventional computing devices and IoT devices. Not only that, but there is no a comprehensive study that presents the panoramic picture of the IoT security threats(Abdur et al. 2017).

According to Sharma and(Sharma and Bhadana 2010), sensor nodes are low power, have limited functionality, and are not individually capable of multi-hop routing. These nodes tend to be application

DOI: 10.14738/tnc.65.5101
Publication Date: 14th September 2018
URL: http://dx.doi.org/10.14738/tnc.65.5101

specific to monitor temperature, video, or pressure. Most often, sensor nodes are grouped in clusters and sited at strategic locations. Sensor nodes monitor applications or provide surveillance to send back to the local forwarding nodes (FN). For each sensor node cluster, there is an individual forwarding node (FN). Forwarding nodes receive the sensor node cluster information and then process the information to obtain aggregate results. These nodes also verify the information received from the SN cluster. This “middleman” node consists of two wireless interfaces between the lower level sensor nodes and the next higher level of the node, the access points (AP).

Possessing both wired and wireless interfaces, access points (AP) utilize its multi-hop routing capabilities to send SN and FN packets to wired networks within a designated radio range as well as to forward control information between SNs and FNs and wired networks. APs also can re-verify the information previously verified at the FN node level. At each of these node points, protected and authenticated communication between the various sensor nodes are key security concerns. WSN sensor node vulnerabilities arise from four separate areas: the open nature of wireless channels, the absence of infrastructure, its rapid speed of deployment, and hostile deployment environments (Maw. 2008).

Due to these four vulnerabilities, security protocols centered solely around physical security cannot be successfully used. Security only becomes more critical against security attacks because sensor nodes are heavily constrained and limited in terms of its internal energy, memory, computational and communication abilities. Because it’s routing paths and relative neighborhood are subject to constant change, networks frequently cannot provide adequate security measures against posed threats such as breaches in confidentiality, integrity, authentication, and authorization(Banković et al. 2012).

There is little or no capability to identify new threats or impending attacks and to react proactively to prevent damage. Security, then, becomes a paramount concern because roaming nodes must constantly be authenticated within neighboring nodes through secure communication keys. Attacks on WSNs can be categorized as passive or active and internally-sourced versus externally-sourced attacks. More specifically, there are two view levels of attacks: security mechanism attacks and basic mechanism attacks. Major attacks can consist of wormhole attacks, spoofing, selective forwarding, black-holes or sinkholes, Sybil attacks, HELLO flooding, and denial of service of these attack sources, wormhole attacks, the focus of this paper, constitute one of the highest continuing threats to WSNs (Ronghui et al. 2009). Wormhole attacks are malicious, passive, external laptop-class threats.

In a wormhole attack, at least two colluding nodes maliciously “create a higher-level virtual tunnel (wormhole) in the network and transport message packets between the tunnel endpoints” (K. E. N. Kumar, Waheed, and Basappa 2010) by offering shorter network links. Unsuspecting nodes are deceived into selecting the shorter routes and replaying the message in a separate part of the network and corrupting data or disabling networks through faulty information. Wormhole tunnels can be established through wired infrastructure links or hidden within out-of-band channels, through high powered transmission lines, or through packet encapsulation above network layers.

1.1 WSNs security challenges

WSNs design model has been divided into three layers: Application level, Communication level, and Perception level. Each layer is a potential target for several designated attacks. Figure 1.1 describes these three layers.

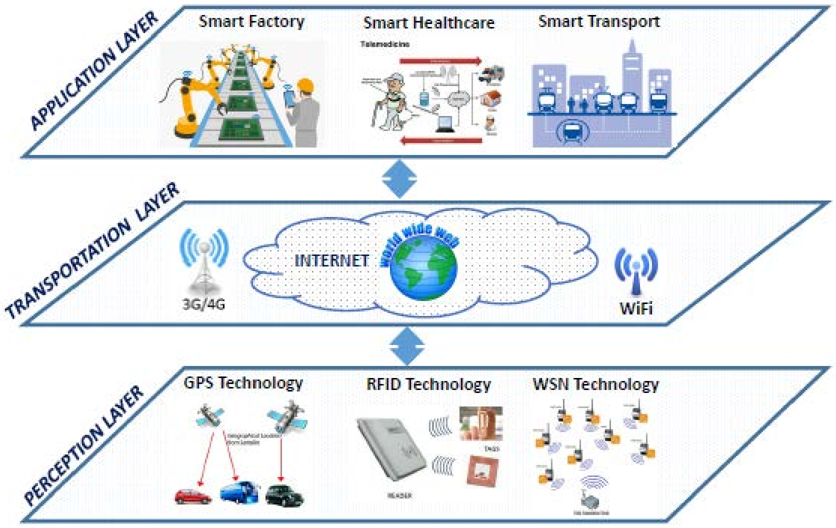


Figure 1: IoT as a Layered Approach [2].

1.2 Application layer attacks

The application layer is a place where specific application business situated such as Smart City, Smart Factories, Health Care etc(Abdur et al. 2017). Each business environment might have its own security and privacy challenge since there is no security standard for IoT for the application layer. Furthermore, at, this layer, some security threats at this layer cannot be avoided at the other layer such as privacy and protection. Even though this layer may have regularly a complex logical business at the different environment with a different purpose, it suffers from various common security threats:

- Data Leak: An adversary could steel client’s confidential information such passwords due to an existence of vulnerability on for example in either application authentication implementation or session management.
- DoS attack: Due to this attack, application service availability could be targeted with thousands of fake requests in order to block or shout down the service from other legitimate customers.
- Malicious code injection: an attacker could utilize a vulnerable application for example with maliciously crafted input to push it to action that it’s not intended to do.

1.3 Perception Layer

Perception Layer of IoT basically involves gathering and processing data over RFID (Radio-Frequency Identification), WSN (Wireless Sensor Network), RSN (RFID Sensor Network) and GPS(Mendez, Papapanagiotou, and Yang 2017). It consists of sensors and actuators that aim to either query the location or measure for example temperature, acceleration, humidity, etc.)

This layer is a potential target to each of the following attacks:

- Physical attacks: The aim of the attackers her is either to cause the damage to the sensor node or physically insert a malicious code into the targeted node.
- Impersonation: attackers utilize the vulnerabilities in the authentication to insert a fake node for malicious or collusion attacks.

- Data Transit Attacks: unsecured communication at this layer opens doors for several potential attacks such as Eavesdropping, Man-in-the-Middle etc.
- Routing Attacks: an intermediate fake node may modify the routing paths.

1.4 Transport Layer

Communication layer is responsible for delivering all kind of communication traffics between the applications and their related outsourced peripherals such as sensors, actuators, networking devices and so on connected through either wireless or wired mediums. Communication Layer is a potential target to one of the following attack(Zhou, Zhang, and Liu 2018):

- Routing Attacks: an intermediate fake node may modify the routing paths 4
- DoS Attacks: due to the heterogeneous nature of IoT, makes the Communication layer a potential target to DoS Attacks.
- Data Transit Attacks: unsecured communication at this layer opens doors for several potential attacks such as Eavesdropping, Man-in-the-Middle etc.

2 Related Work

Due to the advancement in IoT and their applications, the security threats in these fields have received a good attention from some researchers in the last few years.

In (Yang et al. 2017), Yang et al have published an article survey regarding the privacy and security issues in IoT. Their work studied the security issues from four different perspectives. They initially highlighted on the limitations of applying security in IoT devices (e.g., computation power, the battery lifetime etc.) and the suggested solutions for these problems (e.g. lightweight encryption scheme designed for embedded systems). Then, they introduced a classifications summary of different IoT attacks (e.g. local, remote, physical etc). After that, they paid much attention to the mechanisms and architectures of designing and implementing for authorization and authentication purposes. Finally, they did analyze the security issues at different layers (e.g. Application, Transport, etc.) composed of Application, Network, and Perception layers. The Perception layer belongs to the physical devices that identify and sense analog data and then digitize it for transportation purposes. Infrastructure protocols such as ZigBee, Z-Wave, Bluetooth Low Energy (BLE)(Gomez, Oller, and Paradells 2012), Wi-Fi, and LTE-A run in the Network layer(Ghosh et al. 2010). The Application layer is the interface for end-users to access data and talk to their IoT devices. It supports standard protocols such as HyperText Transfer Protocol (HTTP), Constrained Application Protocol (Yang et al. 2017). (CoAP) (Shelby, Hartke, and Bormann 2014).

Also, the security and privacy issues in IoT have been addressed by (J. S. Kumar and Patel 2014) & (Vikas 2015). They studied the security issues at each layer characterized by the three-layer architecture(Vikas 2015) surveyed most of the security threats in IoT, resulted from the various communication technologies used in wireless sensor networks.

The authors in (Vikas 2015) have proposed an authorization access model. They recommended using this model as a security framework for the IoT to assure access control and legitimate authority for users only.

Authors in (Fremantle and Scott 2017) reviewed the challenges and approach proposed to overcome the security issues of the IoT middleware, where a large number of existing systems inherit security properties from the middleware frameworks. Depending on the well-known security and privacy threats, the authors analyze and evaluate the available middleware approaches and show how security is handled by each approach. The work concludes by illustrating a set of requirements to have a secure IoT middleware.

3 System Model and problem statement

Let us consider the proposed wireless networks (WSNs) is comprised of multiple sensor nodes. The nodes are deployed to monitor a certain phenomenon, collect an information and report to the base station through multi-hope forwarding scheme. The forwarded data can be read by any intermediate node consequently this node could corrupt this packet and retransmit it again until reaching the base station. In some application, the collected data is crucial and very sensitive such as poison gas detection, fire detection, and health care. The data resulting from these applications should be protected in order to provide confidentiality to make the right decision. Corrupted data may lead to a disaster. As a result, designing and implementing a secure path for this data is essential to prevent such actions. However, most existing encryption solutions cannot be applied in WSNs due to limited capabilities in processing, transmission, and energy.

As pointed out, we propose to implement a lightweight algorithm such as AES light version which is appropriate for such limited resources devices.

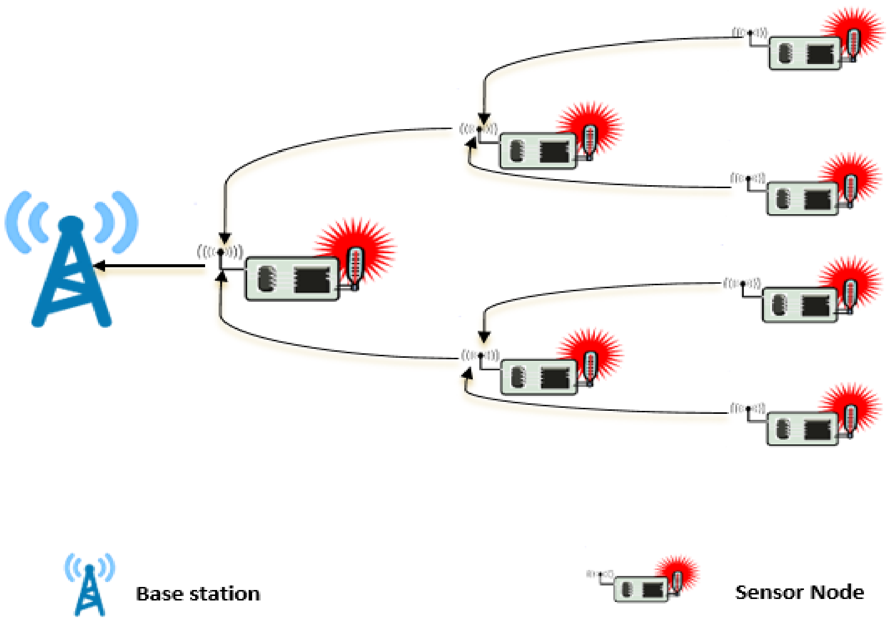


Figure 2: System Model of the proposed application

4 AES light Algorithm

AES Encryption is a method for scrambling data. A key is utilized to mix up data such that it can be securely stored or transferred over the network and the only person with the key can unscramble the data. Algorithm implementation is highly depending on microcontroller technology.

Due to the limited resources of IoT devices, a lightweight algorithm version of AES is developed to be used with these limited resources devices. The lightweight version of AES has small block size, small key size, simple round, and simple key scheduling.

5 Sensors Specification

5.1 TelosB (Sky mote)

TelosB is a platform designed for low power sensor network applications. It uses CC2420 radio chip with a transmission speed of 250 kbps and works on 2.4 GHz radio frequency. Figure illustrates TelosB mote specifications(Memsic 2003).

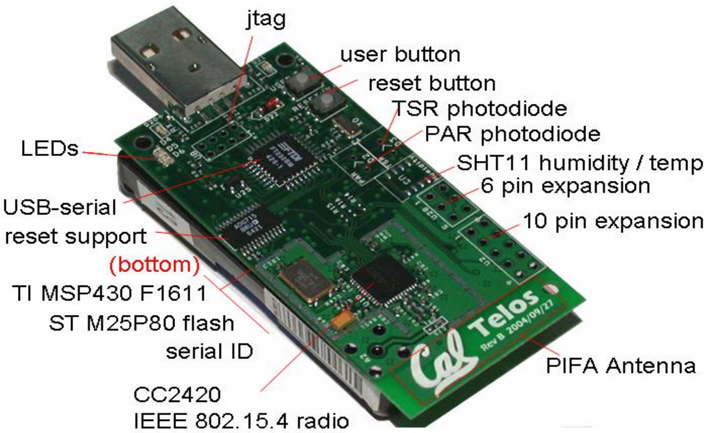


Figure 3: TelosB mote Specification

5.2 Z1 mote

Z1 mote is a low power sensor equipped with a second generation MSP430F2617 microcontroller, which features a powerful 16-bit RISC CPU @16MHz clock speed, built-in clock factory calibration, 8KB RAM and a 92KB Flash memory. Also includes the well-known CC2420 transceiver, IEEE 802.15.4 compliant, which operates at 2.4GHz with an effective data rate of 250Kbps(Zolertia 2010).

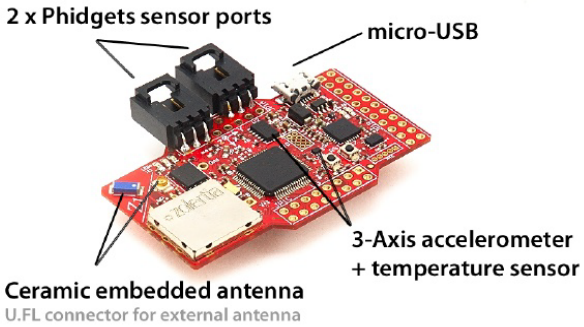


Figure 4: Z1 mote specification

5.3 Experimental Setup

In order to conduct the proposed solution, we have implemented the AES light algorithm on different platforms (mentioned in section) each experiment has been conducted using the following components:

5.3.1 Simulation Scenarios

Initially, TelosB motes have been adopted with different solution scenarios to trace the behavior of this algorithm in different network size. Table 1 summarizes the components of this experiment

Table 1: simulation parameters

Parameter	Value
Tool	Cooja
Network size	Variable (5, 10, 15)
Battery capacity	2600 mAh
Voltage	3 V
Time of sending a packet	10 seconds
Transmission power	0.052 watts
Receiving power	0.069 watts
Active mode	0.0018 watts
Sleep mode	0.0000153 watts

Secondly, the same simulation parameters described in Table also are implemented with Zolertia mote.

5.3.2 Results and Analysis

The analyzed performance metrics that have been used in this study are:

1. Total power consumption: this metric measures the total energy of each sensor nodes as in the following equation. This metric also shows how the effects of the proposed solution are in term of how much of the energy is consumed compared to the unsecured model.

$$Energest_{value}\ per\ cycle = current\ Energest_{value} - previous\ Energest_{value}$$

Where $Energest_{value}$ is the times that the mote spends in this state.

$$Energy\ consumption(mW) = \frac{Energest_{value} * current * Voltage}{RTIMER_{SECOND} * Runtime}$$

Where the $RTIMER_{SECOND}$ is the number of ticks per second.

2. Network lifetime: this metric measures the estimated lifetime of each sensor nodes based on the equation. This metric also determines the lifetime of the whole network. In addition, this metric shows how the ability of the proposed AES encryption in term of how much of the network lifetime is reduced due to security measures.

$$P_{total} = P_{Tx} + P_{RX} + P_{LPM} + P_{CPU}$$

The performance of the two approaches has been investigated using different setups to explore the effect of using different size of sensor nodes in simulation environments. First, we show the effects of applying such algorithm to total energy consumption.

To start with, Figure illustrates the cumulative power consumption for both scenarios (by applying an encryption algorithm and without applying it) when the network size is 10 sensor nodes. It can be noticed that, applying the security measurements require more power due to the complex processing operation used by the AES algorithm for encryption when the packet is sent. Also, as depicted in Figure , the power consumption increases by 39% as a maximum.

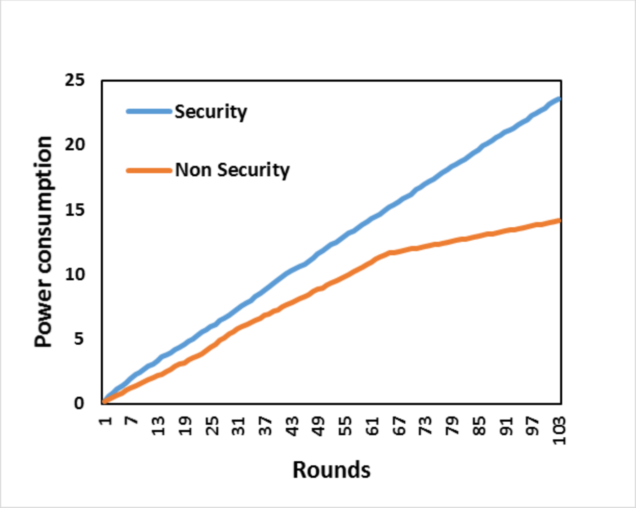


Figure 5: Power consumption (TelosB mote) when the network size is 10 nodes for both security and none security scenarios

For the second scenario when the network size is 15 sensor nodes, as shown in Figure , applying security measurements requires more power compared to the first scenario due to increasing the traffic loads in this scenario. The power consumption increases approximately by 65% compared to none security scenario.

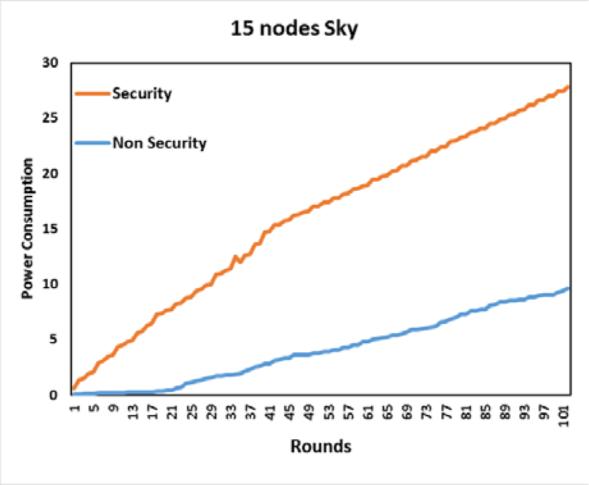


Figure 6: Power consumption (TelosB mote) when the network size is 15 nodes for both security and none security scenarios

In addition, Figure shows the lifetime of both scenarios for with security measurements and without security measurements. It could be seen that the lifetime decreased when the encryption algorithm is applied either when the network size is 10 nodes or when the network size is 15 nodes. This results from the heavier traffic as the network size increases.

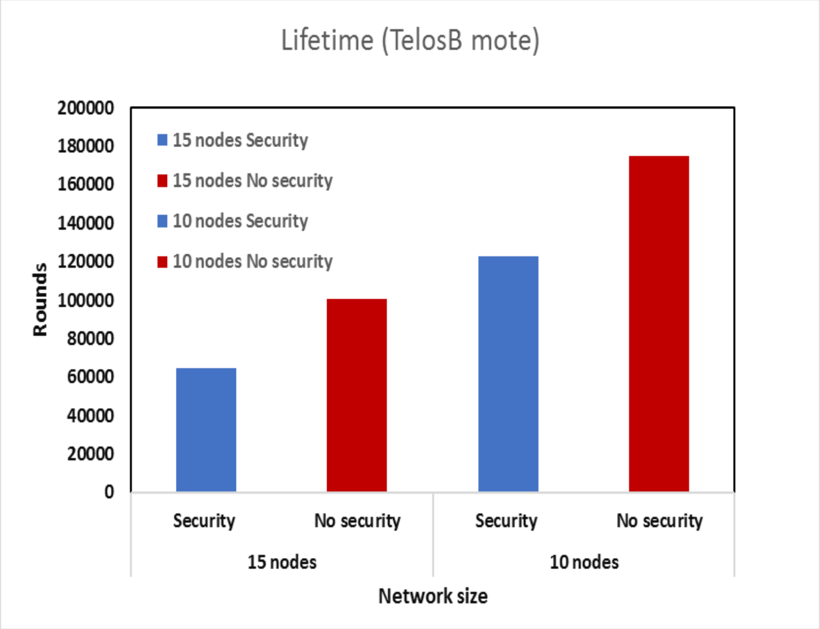


Figure 7: The average network Lifetime when the network size is 10 and 15 nodes for both security and none security scenarios

To see the behavior of this algorithm among different sensor platforms, we have adopted Z1 mote to evaluate the performance of this algorithm. The same simulation setup and performance metrics mentioned in the previous section have been considered in these experiments.

Initially, the first experiment has been conducted when the network size is 10 and 15 sensor nodes and similar to the same scenario in the TelosB motes, applying security algorithm requires more power compared to traditional transmission operation as shown in Figure and Figure .

In contrast, Z1 mote is outperforming TelosB motes in both scenarios either for security or without security measurements.

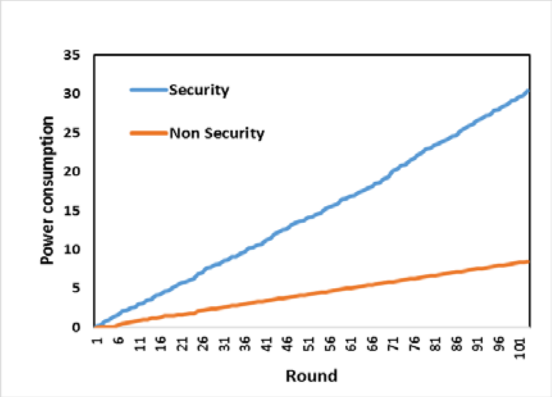


Figure 8: Power consumption (Z1 mote) when the network size is 10 nodes for both security and none security scenarios

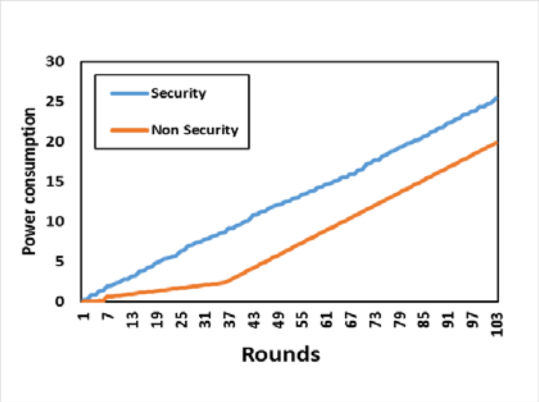


Figure 9: Power consumption (Z1 mote) when the network size is 15 nodes for both security and none security scenarios

6 Conclusion

With its spatially distributed nodes used to monitor physical and environmental conditions in hostile or unattended sites, Wireless Sensor Networks represent a major means of sensing, processing, and communicating data results for military and civilian purposes and applications. Because data is being transmitted and shared, basic security issues such as authentication, integrity, confidentiality, and availability arise. While a variety of threats can be mounted against WSNs, different types of attacks represent one of the major threats to a wireless sensor network's security. Wormhole attacks result from the compromising of two or more sensor nodes. Applying a light version of advanced encryption security algorithm to tackle the gap between the complex computation of encryption algorithm and resource limitation in WSNs. Different experiments with different network sizes have been conducted and the results show that applying such algorithm requires more power but provides a secure communication through multi-hops forwarding schemes. Also, different platforms have been adopted to evaluate the behavior of this algorithm among different sensor manufacturing architecture.

REFERENCES

[1] Abdur, Mirza, Sajid Habib, Muhammad Ali, and Saleem Ullah. 2017. "Security Issues in the Internet of Things (IoT): A Comprehensive Study." *International Journal of Advanced Computer Science and Applications* 8(6). http://thesai.org/Publications/ViewPaper?Volume=8&Issue=6&Code=ijacsa&SerialNo=50.

[2] Al-Fuqaha, Ala et al. 2015. "Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications." *IEEE Communications Surveys & Tutorials* 17(4): 2347–76.

[3] Banković, Zorana, David Fraga, José M. Moya, and Juan Carlos Vallejo. 2012. "Detecting Unknown Attacks in Wireless Sensor Networks That Contain Mobile Nodes." *Sensors (Switzerland)* 12(8): 10834–50.

[4] Fremantle, Paul, and Philip Scott. 2017. "A Survey of Secure Middleware for the Internet of Things." *PeerJ Computer Science* 3: e114. https://doi.org/10.7717/peerj-cs.114.

[5] Ghosh, Amitava et al. 2010. "LTE-Advanced: Next-Generation Wireless Broadband Technology [Invited Paper." *IEEE Wireless Communications* 17(3): 10–22. http://ieeexplore.ieee.org/document/5490974/ (December 24, 2017).

[6] Gomez, Carles, Joaquim Oller, and Josep Paradells. 2012. "Overview and Evaluation of Bluetooth Low Energy: An Emerging Low-Power Wireless Technology." *Sensors* 12(9): 11734–53. http://www.mdpi.com/1424-8220/12/9/11734.

[7] Kumar, J Sathish, and Dhiren R Patel. 2014. "A Survey on Internet of Things: Security and Privacy Issues." *International Journal of Computer Applications* 90(11).

[8] Kumar, K E Naresh, Mohd Abdul Waheed, and K Kari Basappa. 2010. "TCPL: A Defense against Wormhole Attacks in Wireless Sensor Networks." In *AIP Conference Proceedings*, , 633–38.

[9] Maw., Z. Tun and A.H. 2008. "Wormhole Attack Detection in Wireless Sensor Networks." *Proceedings of World Academy of Science Engineering and TechnologyEngineering and Technology* 46(3): 545–50.

[10] Memsic. 2003. “MICAz Datasheet: 6020-0065-05 Rev.” *San Jose, CA, California* Revision 6: 1–2.

[11] Mendez, Diego M., Ioannis Papapanagiotou, and Baijian Yang. 2017. “Internet of Things: Survey on Security and Privacy.” : 1–16. http://arxiv.org/abs/1707.01879%0Ahttp://dx.doi.org/10.1080/19393555.2018.1458258.

[12] Ronghui, He, Ma Guoqing, Wang Chunlei, and Fang Lan. 2009. “Detecting and Locating Wormhole Attacks in Wireless Sensor Networks Using Beacon Nodes.” *Engineering and Technology* 3(7): 286–90.

[13] Sharma, Pooja, and Pawan Bhadana. 2010. “An Effective Approach for Providing Anonymity in Wireless Sensor Network: Detecting Attacks and Security Measures.” *International Journal on Computer Science and Engineering* 02(05): 1830–35.

[14] Shelby, Zach, Klaus Hartke, and Carsten Bormann. 2014. “The Constrained Application Protocol (CoAP).”

[15] Vikas, B O. 2015. “Internet of Things (IoT): A Survey on Privacy Issues and Security.”

[16] Yang, Yuchen et al. 2017. “A Survey on Security and Privacy Issues in Internet-of-Things.” *IEEE Internet of Things Journal*.

[17] Zhou, Wei, Yuqing Zhang, and Peng Liu. 2018. “The Effect of IoT New Features on Security and Privacy: New Threats, Existing Solutions, and Challenges Yet to Be Solved.” : 1–11. http://arxiv.org/abs/1802.03110.

[18] Zolertia. 2010. “Z1 Datasheet.” s: 1–20.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM

TNC TRANSACTIONS ON NETWORKS AND COMMUNICATIONS

VOLUME 6, NO. 5
ISSN: 2054 -7420

A Stratified Cyber Security Vigilance Model: An Augmentation of Risk-Based Information System Security

Abuonji Paul, Rodrigues Anthony, J., George O. Raburu
School of Informatics and Innovative Systems, Jaramogi Oginga Odinga University of Science and Technology, P. O. Box 210- 40601, Bondo, Kenya.
pabuonji@jooust.ac.ke; tonyr@jooust.ac.ke; graburu@jooust.ac.ke

ABSTRACT

Information system security in the current interconnected environment called the cyber-space is continually getting more sophisticated. All the players involved- governments, corporates, IS security experts and users, both naïve and sophisticated- all grapple with one big problem: how to decide on what level of security is enough for their information system since the amount of security controls applied must be commensurate with the IS assets being protected. In that regard, many organizations adopt risk-based security, in the hope that it would answer the elusive IS security question, but to no avail. Unfortunately, many such organizations still experience numerous breaches to their Information systems and some even realize they have fallen victims to cyber criminals, long after the actual compromise. It is for this reason that this paper presents a novel security model called Stratified Cyber Security Vigilance (SCSV) model that augments the standard risk-based security approach and demonstrates its ability to improve IS security.

Key Words: stratified, cyber, security, vigilance, model, risk-based

1 Introduction

As our lives continue to depend more on computers, our own security lies in their security [1]. Governments, business enterprises, political players, researchers, and many others, all depend on computer based information systems to effectively accomplish their undertakings. Therefore security of data and the systems that store and manipulate these data is of paramount importance to all technologically progressive organizations. [2] defined information system security as the set of measures and mechanisms that are put in place in order to safeguard computer or information system's confidentiality, integrity and availability with the aim of preventing unauthorized disclosure, unauthorized modification and unauthorized withholding. [3] on the other hand defines it as the protection afforded to an automated information system in order to attain applicable objectives of preserving integrity, availability and confidentiality of information system resources. In the current digital age, it is difficult to confine information, data and network software to geographical localities and so can't we confine their security [4]. It therefore means that when conceptualizing, designing and implementing IS security must take global perspective and the mechanisms put in place must consider both stationary and transit data as well as being cognizant of the complexities and challenges emanating from diverse legal jurisdictions involved. This whole complex process must be done in a risk-based, cost-effective manner [5].

DOI: 10.14738/tnc.65.5166
Publication Date: 14th September 2018
URL: http://dx.doi.org/10.14738/tnc.65.5166

2 Related Works

Risk-based information system security begins by identifying the assets to be protected, calculating their values so that they are neither over protected nor under protected, assessing both internal and external threats to the system, identifying system vulnerabilities and then conducting threat-vulnerability (T-V) paring to determine which threats are likely to exploit which vulnerabilities to harm the system. In this process, the security expert deals with both real and perceived threats [6]. Information system threats and vulnerabilities cover a wide array of events, virtually none of which can be totally eliminated while still operating the system [7].

Since no system can be rendered totally secure, once threats and vulnerabilities are identified, their impact on the total system must be assessed to determine whether to accept the risk of a particular danger, and the extent to which corrective measures can eliminate or reduce its severity. The primary goal for an IS Security professional involves putting efforts to reduce the threat window, which is the time between detection of an incident and response [8]. This is one of the fundamental aspects of creating a robust security program. It therefore means that detection must be prompt and appropriate action taken immediately. For prompt detection to take place there is need to use a security model that delivers high level of vigilance. [9] said that one best practice towards network security is to build the network model in a way that can foster security. And [10] indicated that, lack of attention on the part of management has exposed many information systems to security breaches. These facts elevate the concept of vigilance to the centre stage of an effective information system security program.

Many information system security models have been proposed in attempt to improve security. These models range from specific to general purpose and do sometimes overlap, depending on designers or implementers. Whatever the case, the models should help in achieving the main information security goals namely confidentiality, integrity, availability, authentication and accountability [11]. The system should be secure, both from the user and administrator perspective. [12] explained that secure systems must have been built using secure components such as strong cryptographic algorithms, secure key redistribution mechanisms and robust authentication protocols. Some of these systems include Pretty Good Privacy (PGP) which provides email security and operates at the application layer, Secure Shell (SSH) which provides security during remote system login. At the transport layer are the Secure Socket Layer (SSL) and the newer version called Transport Layer Security (TLS). Finally IPsec protocol operates at the IP or network layer.

Some common security models include access control list, Bell-La Padula model, Biba model, Brewer and Nash model, Capability-based security, Clark-Wilson model, Graham-Denning model, Harrison-Ruzzo-Ullman model and lattice-based access control model. Others are mandatory access control, object-capability model, role-based access control, take-grant protection model and protection ring. Some implement specific aspects of security- for example Bell-La Padula for confidentiality while Biba and Clerk-Wilson models are geared toward safeguarding data integrity. While Bell-La Padula and Biba models approach access control from a static standpoint, the Brewer and Nash model, also called Chinese wall model provides information security access controls that have the capacity to change dynamically. Its intention was to provide controls that can mitigate conflict of interest in commercial organizations. Most of these models need to be integrated into a more comprehensive and holistic model in order to be able to protect an enterprise end-to-end in line with the organization’s security policy [13].

An example of such a model was proposed by [3]. He described a more general and holistic model for network security as shown in figure 1 below. He demonstrated a scenario where the sender intends to transmit a message to the recipient over the internet. Since the internet is inherently insecure by virtue of its pervasive and ubiquitous use, it is expected that some secure mechanism must be put in place to ensure safe communication.

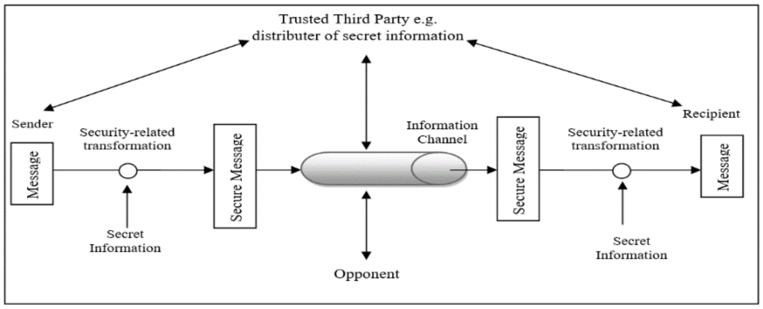


Figure 1: Model for Network Security (Stallings, 2011)

This model borrows heavily from the ideas of [12] explained above. Security related transformation is done using secret information provided by the trusted third parties- in this case encryption keys. This model if adopted and properly implemented can safeguard confidentiality of data through encryption. However it is silent on integrity and availability of data and information systems. Moreover, this paper contends that security in whatever form- whether administrative, physical or logical is miserably impractical without vigilance on the part of security agents, other stakeholders including those being protected and the information system itself.

Generally speaking, vigilance is the process of paying more careful attention, especially in order to notice possible danger [14]. From psychology perspective, vigilance refers to the ability of organisms to maintain their focus of attention and to remain alert to stimuli over prolonged periods of time [15]. In information systems, security of computers or programs largely depends on the efficiency of its vigilance mechanisms that prompts its users and administrators when an abnormal activity occurs [16]. With brisk and widespread adoption of computer systems coupled with quick-fix application development and deployment tendencies, everything in the information system infrastructure appears to be vulnerable. To that end, security of information systems has become a cause of great concern.

In its cyber security report, Deloitte indicated that a good financial system must have three main features- secure, vigilant and resilient. It explained that a secure system must have enhanced risk prioritized controls to protect against known and emerging threats and should comply with industry cyber security standards and regulations. A vigilant system must detect violations and anomalies through better situational awareness across the environment. Finally, a resilient system must establish the ability to quickly return to normal operations and repair any damages to the business [17]. It further illustrates that organizations need multipronged approach to cyber security management involving automated systems and people. This is illustrated in figure 2 below.

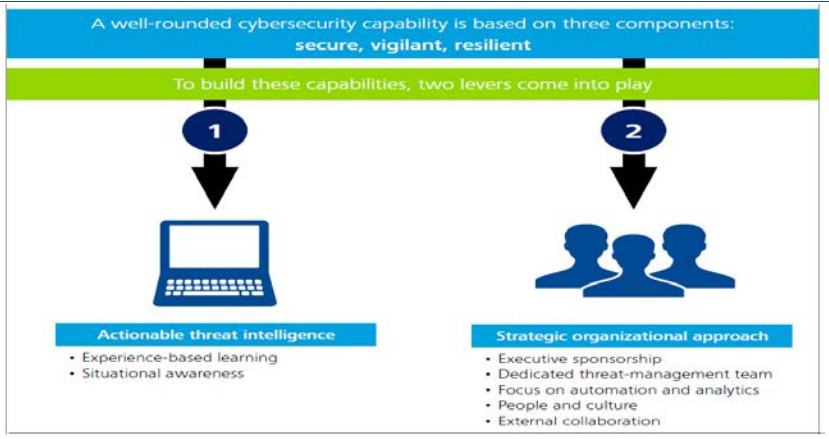

A well-rounded cybersecurity capability is based on three components:
secure, vigilant, resilient
To build these capabilities, two levers come into play
1
2
Actionable threat intelligence
• Experience-based learning
• Situational awareness
Strategic organizational approach
• Executive sponsorship
• Dedicated threat-management team
• Focus on automation and analytics
• People and culture
• External collaboration

Figure 2: Multipronged Approach to Cyber Security Management
Source: Deloitte Center for Financial Services Analysis, Whitepaper (2014)

Several studies indicate that many cyber-attacks take place against individuals, organizations or states and either go unnoticed or are detected too late thereafter. These attacks are perpetrated either by lone cyber criminals, organized criminals like Anonymous hackers and hacktivists or government agencies. According to [18], United States, British and Chinese governments participate in large scale cyber-crimes in the name of gathering intelligence to curb terrorism or for economic and political espionage. To prevent such acts of law breaking by the supposed law enforcers, citizens ought to be vigilant and defend themselves through civil rights activism or the technology community need to over-engineer their systems to prevent unlawful government surveillance that impinge on citizen’s fundamental rights to privacy.

[17] also reported that the response time to cyber-attacks by global financial services firms indicates significant gaps in their preparedness. This is due to inability of these firms to quickly detect and respond to cyber threats. As illustrated in figure 3 below, attack success is the time to compromise the system. It measures time from the first malicious action taken against the victim until the point at which an information asset is negatively affected. Discovery success is the time from compromise of the system to discovery. It measures time from initial compromise to when the victim first learns of the incident. Finally, restoration success is the time from discovery to containment. It measures the time between discovery of a breach to when it is successfully contained. The study showed a big margin between attack success and discovery success, and another big margin between discovery success and restoration success. This clearly shows a lapse in vigilance.

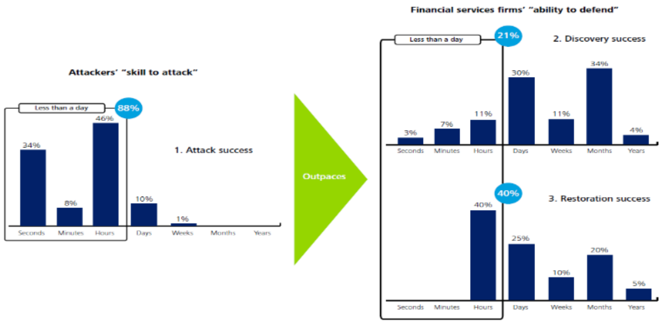


Figure 3: Response time to attacks indicates significant gaps in preparedness
Source: Deloitte Center for Financial Services Analysis, Whitepaper (2014)

The Deloitte study discovered that 34% of successful attacks to information systems happened within seconds; 8% succeed within minutes and 46% succeed within hours. These statistics showed that a whopping total of 88% of information systems that were attacked and got compromised succumbed to their victims within less than a day. Only 10% resisted attacks for days and another 1% for weeks. The remaining 1% was not allocated. This grim state of IS security in financial systems was worsened by the fact that only 21% of targeted organizations had security systems that could detect attacks within a day while majority took longer to detect- 30% took days to detect, 11% took weeks, 34% took months and 4% took years to detect that they were compromised. When it came to restoring the system, no organization was able to restore its system or services within seconds or minutes. The most efficient was discovered to be able to restore within hours, constituting 40%. The other, 25% took days to restore their systems, 10% took weeks, 20% took months and 5% took years to restore their systems. The disturbing results of this study revealed a huge lapse in vigilance and security of information systems in financial services firms.

Kenya's case is not any batter as [19] reported that various sectors of the economy were just coming to terms with the alarming fact that hacking was taking a menacing proportion and causing untold havoc. It stated that in late 2014, companies lost a mindboggling KSh. 15 billion to hackers. The statistics which they got from Kenya Cyber Security Report 2015 showed that an average of 30 companies suffered cyber-attacks in Kenya daily. The bigger problem was that the devastation caused was unlikely to be detected until up to 120 days later. This information showed that there was dire need for enhanced vigilance in order to successfully combat cybercrime.

The necessity for vigilance is usually augmented by the complexity that characterizes information systems security. This is because IS security is pervasive and is a continuous process not an event. A well designed and implemented security program, if forgotten is as good as no security [20]. Besides, vigilance should be stratified to ensure that the degree of vigilance required for any system or resource is commensurate with the level of risk faced by that system or resource. This requires organizations to diligently assess their risk levels and develop a stratified vigilance scheme based on the risk ratings [21].

There is no doubt that risk rating is a principal factor in determining security level of an information system. NIST (2003) defined risk in information systems as the net negative impact of the exercise of a threat on a vulnerability, considering both the probability and the impact of occurrence. Risk management on the other hand is the process of identifying vulnerabilities and threats to the information resources used by an organization in achieving business objectives, and deciding what countermeasures, if any, to take in reducing risk to an acceptable level, based on the value of the information resource to the organization [23].

Risk management is a continuous process that starts by identifying organization's assets or resources that need to be protected and estimating their values; then assessing the threats to those assets or resources, followed by an assessment of all the vulnerabilities that exist in the system. Risk is thereafter calculated based on the probability that a given threat would exploit a vulnerability to cause harm to the systems, and the impact this action would have on the system and business processes. After this determination has been made, the organization needs to identify, select and implement appropriate controls that are proportionate with the level of risk. The final stage is to evaluate effectiveness of the controls and start the process all over again. Note that organizations can take any of the following five possible options when dealing with risks [24]. They can- mitigate, transfer, accept, avoid or deny the risks. The decision made will depend on the rating of these particular risks, where each risk will be treated individually.

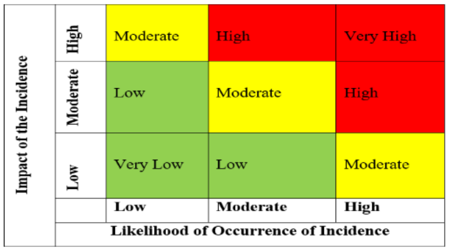
Risks can be calculated or estimated using quantitative or qualitative approaches. In information systems, the most tenable way is using qualitative approach [24]. In this approach, the probability or likelihood of a threat exploiting a vulnerability to cause harm over a period of one year will be rated as low, moderate or high, based on percentages. Table 1 below shows a sample likelihood definition that can be sued.

Table 1: Likelihood Definitions (NIST, 2003)

Ratings	Definition
Low	0-33% chance of successful exercise of threat during a one-year period
Moderate	34-66% chance of successful exercise of threat during a one-year period
High	67-100% chance of successful exercise of threat during a one-year period

On the other hand the level of damage or impact that the attack is likely to exert on the system or business is also rated as low, moderate or high. In many cases this will be evaluated based on the effect on confidentiality, availability and integrity of the information resources of the organization including business processes. Once this is satisfactorily done, the risk determination matrix can be constructed to show how various risks are rated. This is shown in table 2 below.

Table 2: Risk Rating/ Determination Matrix



Impact of the Incidence				
	High	Moderate	High	Very High
	Moderate	Low	Moderate	High
	Low	Very Low	Low	Moderate
		Low	Moderate	High
		Likelihood of Occurrence of Incidence		

As discussed in the previous sections of this research, risk and vigilance are key factors in security of information systems. Even the process of risk management itself requires vigilance. This is because risks and vulnerabilities are dynamic and a vigilant risk management team or system is required to keep pace with the ever changing cyber security landscape (NIST, 2003). A stratified vigilance scheme that matches risk ratings is therefore proposed as follows: very low vigilance, low vigilance, moderate vigilance, high vigilance and hyper vigilance. And using this rating, a risk - vigilance paring for an information system can then be done

This paring helps to provide vigilance which is commensurate with the risk to an information asset. This is supported by Resource Theory of vigilance proposed by [25], [26],[27], [28]. According to this theory, human vigilance depends on the mental capacities or resources that can be allocated to the task. The concept of resources draws on economics of vigilance which supports the assertion that the resources used to secure an asset must be commensurate with the value of that asset. In this case the organizations need to employ both human and automated vigilance which have financial implications.

The concept of vigilance is heavily applied in security [29], [30] and the whole of military spectrum [31]. This study broadly classifies vigilance into two main components namely human vigilance and automated vigilance. In the modern technology-driven enterprise environments, vigilance has ceased being an exclusive human activity as was the case in the olden days. In addition to involvement of human actors, which is very critical for making semi-structured and unstructured decisions, there are myriad automated systems designed to alleviate the physiological and psychological human resources required in vigilance as an aspect of security.

There are many automated IS devices or components for implementing vigilance. The most commonly used such devices or systems in cyber security vigilance are firewalls, intrusion detection systems (IDS), intrusion prevention systems (IPS), bandwidth management and monitoring tools, traffic analysis or discovery tools, system alerts, host based scanners, network and port scanners, among others [32], [33].

The other component is human vigilance as earlier stated. One needs to underscore the importance of involvement of human actors in cyber security because automated systems may not have the capacity to make critical unstructured and semi-structured decisions based on environmental and other factors necessary in decision making. For example humans need to view system logs, alerts, alarms and decide how to respond appropriately. There are instances when, one may even decide to shut down the system temporarily in order to solve a more catastrophic attack that would happen if the system remained running throughout. Depending on the organogram of the organization, the human actors may include operational staff, operational level managers or supervisors, middle level managers and top level or strategic level managers [34], [35].

After considering components of vigilance in an IS system, the next thing is to look at vigilance activities in an information system. These are those factors that determine whether a system is vigilant or not but may not necessary be used to measure or rate the quality of vigilance. In psychology, vigilance is described as the ability to maintain concentrated attention over prolonged periods of time with the intention to detect the appearance of a particular target stimulus [36]. The individual watches for a signal stimulus that may occur at an unknown time. This means that vigilant human actors need to constantly observe their environment. To do this, they must be able to detect activities with random occurrences, the stimuli must be transmitted into the brain for interpretation and then an appropriate action or response is made.

[37] recommended the need for a vigilance classification based on the domain of application. When this concept is applied in information system security, vigilance must exhibit similar features to those in psychology, but the components used will be different. The system- both automated and human components- must monitor, detect, interpret, report and respond appropriately to activities and events [38]. The incident or stimulus comes from the environment of the system. This incident is detected by the system, interpreted, reported and a response is made- which in this case should be a corrective action [39].

There are also parameters that can be used to measure the quality of vigilance in an information system. The process of detection, interpretation, reporting and response must be done in a manner that can help in making correct decisions in a timely manner. Consider for example, the British Royal Air Force while using RADAR during the Second World War were not able to detect their targets accurately and hit the German U-boats that were sinking allied ships. Therefore the quality of their vigilance was considered to be poor [40]. For effective and quality vigilance, the indicators of vigilance –detection, interpretation, reporting and response must be done in an accurate and timely manner. Automated and human vigilance components must never issue false alarms. However time for detection, interpretation, reporting and response can be tampered with the level of risk to the organization. The study therefore finds the need to develop a mechanism to relate risk, vigilance and time constraints allowed for each level of risk to be tackled in compliance with resource theory of vigilance articulated by [41], [42]. In that regard, as the level of risk increases so does vigilance increase and time constraints decrease proportionately. This depicts an inverse relation between risk and time constraints for action. These time constraints can further be matched to the kind of reasonable response that may be required of the system or system administration based on the organization's security policy.

It is worth remembering that vigilance classification is done based on calculated or estimated level of risk to the organization's information system assets. These assets help the organization to pursue its business strategy. From the perspective of strategic management, information technology strategy and its adoption must fit the business strategy since IT should be a driver of business strategy and processes [43]. This means that cyber security risks are not just IT problems but are strategic business problems [17], [10]. Therefore, once information about system risks has been generated, it must be reported for decision making and action. Irrespective of the organogram of the organization, risk must be responded to appropriately and proportionately.

According to [34] in the process of managing different subsystems of an organization, executives at various levels of the organization need to make management decisions. He classifies these decisions as strategic, tactical and operational decisions based on three management levels namely strategic, tactical and operational levels. [44] also classify management levels into three, but with different names- senior management, middle level management and supervisory management. [35] on the other hand describes four levels of management as top management, intermediate management, middle management and supervisory or operational management.

Whichever names given to these levels of management, the single point of agreement is that the top most level of management in any organization deals with the overall strategy of the organization. Middle level deals with interpretation and implementation of the strategy and low level management deals with supervision of the day-to-day operations of his or her section or division in the organization [35], [34]. Figure 4 below illustrates this information.

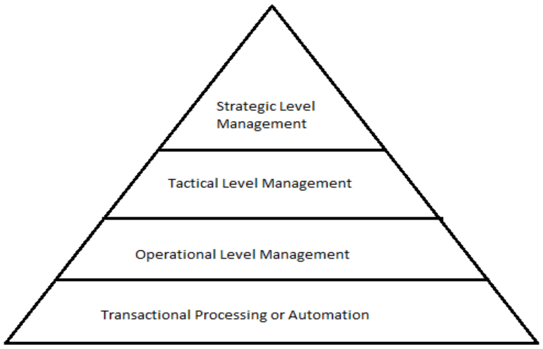

Strategic Level Management
Tactical Level Management
Operational Level Management
Transactional Processing or Automation

Figure 4: Management Levels (Lucey, 2005)

With reference to the above explanation, highly rated risks have direct impact on the strategic objectives of the organization and therefore should get direct and immediate involvement of the strategic level management of the organization. Moderate risks need the immediate attention of middle level management and low level risks may only need direct and immediate attention of operational level management. This stratification of response to risk based on management levels is in tandem with economics of risk management and is what partially informs the stratified cyber security vigilance (SCSV) model, developed in the subsequent section.

3 Model Formulation

Proper risk management is the hallmark and foundation for any effective information system security program since organizations need to assess their risks and implement appropriate security controls that can mitigate against those risks [45]. It is for this reason that the study aimed at developing a cyber-security model founded on the widely accepted risk-based information system security model, but with an enhancement called stratified vigilance. As earlier stated, risk management is the process of identifying vulnerabilities in and threats to information resources used by and organization in achieving business objectives and deciding what countermeasures, if any, to take in reducing risk to an acceptable level, based on the value of information resource to the organization [23]. It is as an iterative process comprising of the following steps:

1. Identification of all relevant assets and estimation of their values
2. Conducting threat assessment within the organization and its environment
3. Conducting vulnerability assessment within the organization
4. Performing threat- vulnerability (T-V) pairing
5. Calculating risk in terms of likelihood of occurrence and impact on the information assets
6. Identification, selection and implementation of appropriate security controls
7. Evaluation of the effectiveness of those controls that have been implemented

This process has been illustrated in figure 5 below. It is vital to note that at the center of risk management is business strategy and processes. This is because threats have the potential of disrupting business processes and the overall business strategy of the organization. Therefore risk management process must be owned and driven by the strategic level management. After risk assessment has been conducted successfully, the management of the organization can make any of the following five decisions risk by risk: mitigate the risk internally by implementing appropriate controls or countermeasures; transfer the risk to a third party for example by outsourcing the service or engaging the services of an insurance company; avoid the risk by removing the risky parts of the system being used or replacing it all together with a lower risk system; accept the risk if its likelihood and impact is very low; or deny the risk if it's not clearly articulated.

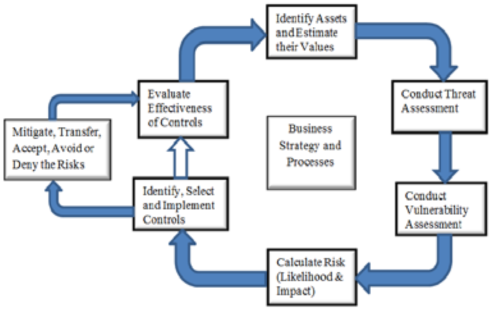

Identify Assets and Estimate their Values
Conduct Threat Assessment
Conduct Vulnerability Assessment
Calculate Risk (Likelihood & Impact)
Identify, Select and Implement Controls
Evaluate Effectiveness of Controls
Mitigate, Transfer, Accept, Avoid or Deny the Risks
Business Strategy and Processes

Figure 5: Risk Management Process

Even though risk can be calculated using quantitative or qualitative approach, in information systems risk management, the most tenable approach is using qualitative one [24]. Table 1 and table 2 above respectively show likelihood definitions and risk rating or risk determination matrix used in qualitative risk management. The five level risk rating scale developed for this study is: very low, low, moderate, high, very high. And as was discussed in previous section risk and vigilance are principal factors in information system security [22], [20], [45]. In the light of this, a stratified vigilance scheme that matches risk ratings would therefore be proposed as follows: very low vigilance, low vigilance, moderate vigilance, high vigilance and hyper vigilance. A risk-vigilance paring for an information system can then be developed as shown in figure 6 below. This rating will help in providing vigilance that is commensurate with risk to an information asset. This is supported by Resource theory of vigilance proposed by [25], [26], [27], [28].

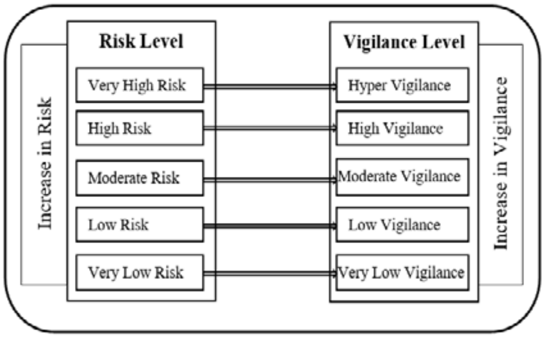


Figure 6: Risk- Vigilance Paring

Once threat - vigilance paring has been done, vigilance needs to be clearly defined in a manner that it can be clearly classified in terms of security roles by human actors [29], [30], [31] and automated technical controls [32], [33]. Additionally, [37] recommended that vigilance should be classified based on the domain of application. Therefore when vigilance is applied in the discipline of information system security, the entire system comprising of automated and human components must be able to monitor, detect, interpret, report and respond appropriately to events and actions [38].

Figure 7 below illustrates this process. The incident or stimulus comes from the environment of the system. This incident is detected by the system, interpreted, reported and a response is made- which in this case is referred to as corrective action. Note that interpretation appears twice. If the information system tools used to implement security vigilance are intelligent then interpretation of security reports are automated. In that case, response to the undesired incident can also be automated. However if the system is not intelligent enough to interpret the incidences correctly then the role is transferred to human actors who will interpret the incident reports and respond to them appropriately. In such as case, then the role of automated security of the information system will simply be to detect and report incidents in raw form for human actors to interpret and respond.

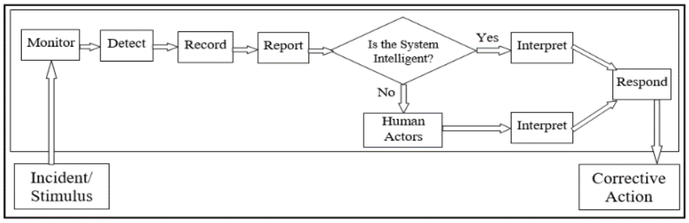


Figure 7: Ideal Information System Security Activities

The quality of vigilance in an information system can be measured in terms of accuracy and timeliness. This means that the process of detection, interpretation, reporting and response must be done in a manner that can help in making correct decisions within an acceptable time constraint. This gives an impression of good quality of vigilance [40]. Figure 8 below illustrates how the stratified vigilance model correlates risk to vigilance in terms of time constraints allowed for each level of risk. Note that as the level of risk increases so does the vigilance increase and time constraints decrease. This depicts an inverse relation between risk and time constraint for action.

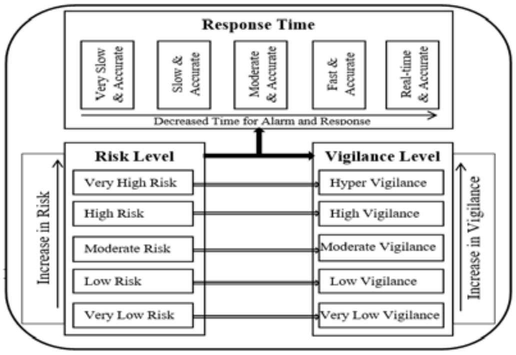


Figure 8: Relating Risk, Vigilance and Time Constraints for Actions

The time constraints described in figure 8 can further be matched to the kind of reasonable response that may be required of system administration based on the organization's security policy. Table 3 below illustrates an example of the possible activities expected of system administration at different levels of vigilance.

Table 3: Time Constrain Metrics for Vigilance

Vigilance Level	System Administration Activities
Very Low Vigilance	System is setup but unmonitored. Owners are content with its existence.
Low Vigilance	System is monitored to ascertain system health.
Moderate Vigilance	If there is a problem, system diagnosis is done to identify cause.
High Vigilance	Ensures that a solution is found and test its effectiveness.
Hyper Vigilance	If no concrete solution is found, propose a solution and escalate.

Since vigilance classification is done based on calculated level of risk, and cyber security risks are strategic business risks [10], [17] therefore risk management is a strategic management role and all other levels of management must perform their respective delegated roles in tandem with that fact [43]. Meaning the entire process must be steered by strategic level management as illustrated in figure 4 above. All these components are then put together to form the SCSV model shown in figure 9 below.

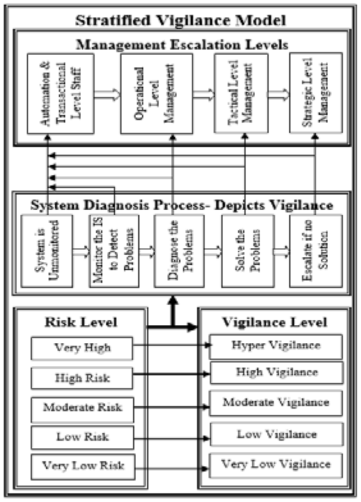

Stratified Vigilance Model
Management Escalation Levels
Automation & Transactional Level Staff
Operational Level Management
Tactical Level Management
Strategic Level Management
System Diagnosis Process- Depicts Vigilance
System is Unmonitored
Monitor the IS to Detect Problems
Diagnose the Problems
Solve the Problems
Escalate if no Solution
Risk Level
Very High
High Risk
Moderate Risk
Low Risk
Very Low Risk
Vigilance Level
Hyper Vigilance
High Vigilance
Moderate Vigilance
Low Vigilance
Very Low Vigilance

Figure 9: SCSV Model

Five goals of securing information systems and services running in a network have been identifies in this study as the need to identify and control logical threats; the need to identify and control logical vulnerabilities; the need to enhance availability of the system and outsourced services; the need to optimize the utilization of IS resources; and finally the need to produce a proper design and appropriately reengineer the information systems. These goals were also identified as the parameters with which the effectiveness of the SCSV model shall be tested. Appropriate IS design and reengineering is placed at the centre of all the other four parameters since it directly affects other parameters. Figure 10 below illustrates these goals and relates them to ideal information system security activities illustrated in figure 7. These are the activities that are expected to take place in a secure vigilant system. And consequently, figure 10 is a simplification of what a vigilant system is, combined with the parameters identified for testing the SCSV model.

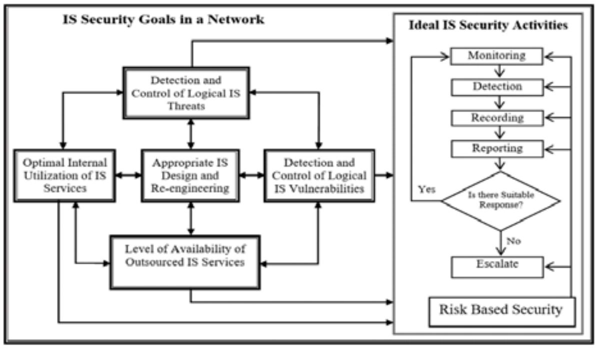

IS Security Goals in a Network
Detection and Control of Logical IS Threats
Optimal Internal Utilization of IS Services
Appropriate IS Design and Re-engineering
Detection and Control of Logical IS Vulnerabilities
Level of Availability of Outsourced IS Services
Ideal IS Security Activities
Monitoring
Detection
Recording
Reporting
Is there Suitable Response?
Yes
No
Escalate
Risk Based Security

Figure 10: IS Security Goals in a Network and Ideal IS Security Activities

The formulated stratified cyber security vigilance (SCSV) model was then tested using these parameters to validate its effectiveness in mitigating risks in a corporate information system environment. The complete model together with the testing parameters is illustrated in figure 11 below.

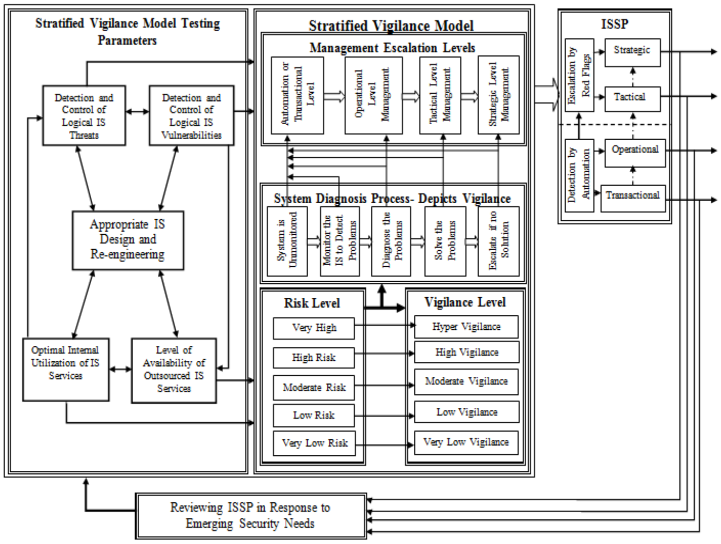

Stratified Vigilance Model Testing Parameters
Detection and Control of Logical IS Threats
Detection and Control of Logical IS Vulnerabilities
Appropriate IS Design and Re-engineering
Optimal Internal Utilization of IS Services
Level of Availability of Outsourced IS Services
Stratified Vigilance Model
Management Escalation Levels
Automation or Transactional Level
Operational Level Management
Tactical Level Management
Strategic Level Management
System Diagnosis Process- Depicts Vigilance
System is Unmonitored
Monitor the IS to Detect Problems
Diagnose the Problems
Solve the Problems
Escalate if no Solution
Risk Level
Very High
High Risk
Moderate Risk
Low Risk
Very Low Risk
Vigilance Level
Hyper Vigilance
High Vigilance
Moderate Vigilance
Low Vigilance
Very Low Vigilance
ISSP
Escalation by Red Flags
Detection by Automation
Strategic
Tactical
Operational
Transactional
Reviewing ISSP in Response to Emerging Security Needs

Figure 11: SCSV Model with Testing Parameters

In this model, risks are rated and assigned appropriate levels of vigilance. These levels of vigilance determine who first acts on the risk and how they need to respond. For instance, when a very low level risk is detected in the information system, automated vigilance systems and operational level employees are expected to deal with it conclusively. In case of a medium level risk, the automated systems and operational level managers are expected to report this to tactical level management and they handle it together. If the risk is high, this must be escalated quickly from operational managers, to tactical level managers for appropriate response. However in case of a very high risk, a real-time reporting and escalation is required all the way to strategic level managers. They will then make a decision and recommend actions commensurate with the risk, which may even include reviewing the ISSP. The reporting and escalation can be done using secure emails, SMS alerts, red flags and other forms of alarms depending on response time constraints allowed for each risk level.

4 Model Testing

Preceded by fastidious preparations and meticulous planning, the study entailed designing a sophisticated network infrastructure and deploying various security tools on this network to collect real-time and residual data. The systems were then adjusted appropriately from time to time and at each stage data was collected. Five phases of the system were deployed and tested. Each phase had different configuration and components were set up in a manner that the initial system was simpler in design, configuration and had fewer components while each subsequent phase got more complex in all the above mentioned aspects. Each added level of sophistication represented an advanced level of vigilance. Therefore each of the five levels of system design corresponded to each stratum of vigilance in the model, and this in turn corresponded to each stage of the system diagnostics ability. As illustrated in figure 12 below, the SCSV model was applied in a network to test whether it can improved the system's ability to detect and control logical threats such as computer viruses.

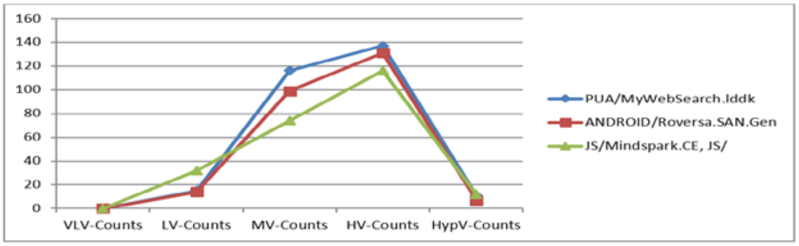


Figure 12: Effect of Vigilance of Virus Detection as Control

The data in figure 12 shows that at very low vigilance level, no viruses were detected because there were no monitoring tools deployed on the network. As the level of vigilance increased, the number of detections steadily increased from low, medium to high vigilance where the highest number of viruses were detected. However, at hyper vigilance level, we observed a sharp decline in the numbers of viruses detected in the system since most viruses have were detected and locked at high vigilance level. However, the remaining viruses needed human intervention to control them such as deployment of other additional controls that could specifically control these viruses.

The SCSV model was also applied on the network to detect and control vulnerabilities. One of the most common logical vulnerability in information systems is unprotected network ports on hosts. A scan was performed on the network and the results shown in figure 13 below were obtained.

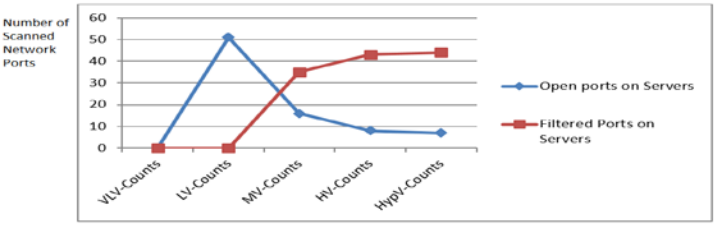


Figure 13: Effect of Vigilance on Status of Communication Ports

As depicted in figure 13 above, low levels of network vigilance was characterized by high numbers of unused open ports on servers while as vigilance increased, most of these ports were filtered using firewall rules thereby increasing the security of servers.

The SCSV model was similarly applied on the network to test stability of two ISP links and results illustrated in figure 14 below.

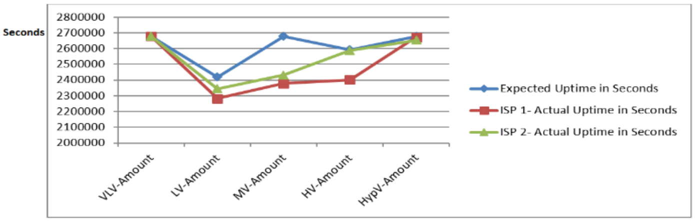


Figure 14: Effect of Vigilance on Link Stability

As depicted in figure 14, from the onset during very low vigilance, no monitoring tools were used and therefore the assumption was that the uptime was the maximum available number of second in a month. However as the level of vigilance increased, it was detected that the uptime for both links wasn't as expected since there were brief, and sometimes prolonged downtimes from time to time. This discovery was followed by timely communication between the organization receiving the service and the ISPs to ensure prompt restoration of service. It therefore led to better levels of stability for both links as indicated in the graph above.

Finally, we also applied SCSV model in managing access to YouTube which was one of the resources in the internet whose access consumed the highest amount of bandwidth of the organization where the research was done.

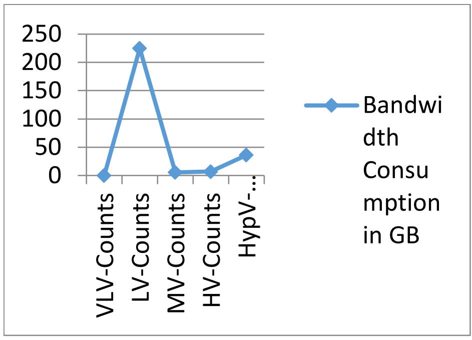


Figure 15 (a): YouTube Bandwidth Usage

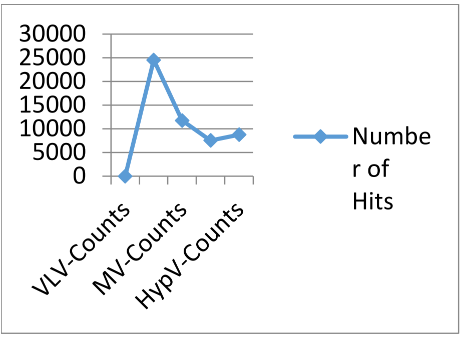


Figure 15 (b): Number of Hits on YouTube

As shown in figures 15 (a) and (b) above, during very low vigilance, no records on YouTube usage were captured because there were no network monitoring tools. However at low vigilance level, the system was configure to monitor the applications running on the network and it detected that about 220 GB of data was being consumed by YouTube alone. A QoS policy was written to control this activity. This resulted into a sharp decline in the amount of bandwidth used in the YouTube steaming to less than 10 GB. After a while the consumption started increasing gradually to about 30 GB. This was as result of users discovering that they had been restricted from accessing YouTube videos and starting to deploy tools to help them circumvent these policies. Table 4 below displays some of the tools users deployed to circumvent the QoS policy restricting the use of YouTube in the network. Figure 15 (b) on the other hand shows that the amount of hits on YouTube declined but not by the same proportion as the amount bandwidth consumed. This was because users continued trying to access the service even when it had been restricted. However when they discovered that they could not access it, most of them gave up but a few decided to use software tools to counter the controls. This majorly happened in subnets 192.168.2.0/24 and to a less extent in subnet 192.168.0.0/24.

Table 4: Tools used to Circumvent QoS Policy in the Network

Tool used to Circumvent QoS Policy	Number of Users in 192.168.2.0/24	Number of Users in 192.168.0.0/24	Number of Users in 192.168.1.0/24
Gbridge VPN Proxy	1069	134	17
Tor Proxy	88	29	5
Ultrasurf Proxy	1025	463	98
Operamini Proxy	2231	74	13
Hotspotshield Proxy	22014	1267	21
Secure Socket Layer Protocol	141702	113221	124619

5 Conclusion

In this paper, we clearly described how we systematically developed and tested the SCSV model for corporate information system security. From the empirical data we collected in-situ, analyzed and used to test the model, we demonstrated that using the SCSV model in a corporate network would improve the process of detecting and controlling logical threats and vulnerabilities, enhancing optimal utilization of internet bandwidth and increasing availability of internet bandwidth in an organization due to enhanced link stability.

REFERENCES

[1] Goyal, A. (2011), Systems Analysis and Design. Asoke K. Ghosh, PHI Learning Private Limited: New Delhi

[2] Banday, T. M. (2011). Effectiveness and Limitations of E-mail Security Protocols; International Journal of Distributed and Parallel Systems (IJDPS) Vol.2, No.3, May 2011

[3] Stallings, W. (2011). Network Security Essentials: Applications and Standards, 4th Ed; Pearson Education, Inc: Prentice Hall

[4] Tanenbaum, A. S. & Steen, M. V. (2014), Distributed Systems: Principles and Paradigms, 2nd ed. Edinburg Gate: Pearson Education Limited.

[5] Dean, M. (2008). A risk-based approach to planning and implementing an information security program. Paper presented at PMI® Global Congress 2008—EMEA, St. Julian's, Malta. Newtown Square, PA: Project Management Institute.

[6] Wurzler, J. (2013), Information Risks & Risk Management; SANS Institute InfoSec Reading Room. Retrieved on 2-1-2016 from: http://www.sans.org/reading-room

[7] Tanenbaum, A. S. (2011). Computer Networks; 4th ed. Prentice-Hall, Inc: New Jersey

[8] Reck, R. (2014), CISO Spotlight: Robb Reck on Security Strategies for Financial Services. Retrieved on 31-12-2015 from: http://darkmatters.norsecorp.com/2014/12/10/cisospotlight- robb-reck-on-security-strategies-for-financial-services

[9] Habraken, J. & Hayden, M. (2009), Teach Yourself Networking in 24 Hours, 3rd ed. Sams Publishing: United States.

[10] O'Brien, J. A. & Marakas, G. M. (2011). Management Information Systems, 10th ed. McGrow-Hill/ Irwin: New York

[11] Laudon, K. C. & Laudon, J. P. (2012). Management Information Systems: Managing the Digital Firm, 12th ed. Pearson Education Limited: Edinburgh Gate, Harlow.

[12] Peterson, L. L. & Davie, B. S. (2007). Computer Networks: A systems Approach, 4th ed. Elsevier, Inc.: San Francisco.

[13] Sinha, P. K. (2007). Distributed Operating Systems: Concepts and Design. Asoke K. Ghosh, PHI Learning Private Limited: New Delhi.

[14] Cambridge Advanced Learner's Dictionary (2010), 3rd ed. Cambridge: Cambridge University Press.

[15] Parasuraman, R. (1986). Vigilance, monitoring and search. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of human perception and performance: Vol. II. Cognitive processes and performance* (pp. 41-1–41-49). New York: Wiley.

[16] Pandey, S. K. (2012), Security Vigilance System Through Level Driven Security Maturity Model; *International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol.2, No.2*.

[17] Deloitte whitepaper (2014), Transforming cyber security in the Financial Services Industry New approaches for an evolving threat landscape; retrieved on 17th May, 2016, from *www2.deloitte.com/content/dam/.../ZA_Transforming_Cybersecurity_05122014.pdf*

[18] Greenwald, G. (2014), No Place to Hide: Edward Snowden, the NSA & the Surveillance State; Penguin Random House, UK.

[19] Daily Nation Newspaper (23rd November, 2016), Rising Threat of Cyber-attacks Put Companies on the Edge. Published on Tuesday 23rd November, 2016.

[20] Stewart, J. M., Tittel, E. & Chapple, M. (2005), CISSP: Certified Information Systems Security Professional Study Guide; 3rd ed. Sybex Inc.: London

[21] Ward, J. & Peppard, J. (2002), Strategic Planning for Information Systems, 3rd Ed. John Wiley & Sons Ltd: Cranfield, Bedfordshire.

[22] National Institute of Standards and Technology –NIST (2003), Building an Information Technology Security Awareness and Training Program; *NIST Special Publication 800 50. Retrieved on 13th November, 2015 from:* csrc.nist.gov/publications/drafts/800-16-rev1/draft_sp800_16_rev1_2nd-draft.pdf

[23] CISA Review Manual (2016), Certified Information Systems Auditor (CISA) Review Manual 2016. Retrieved on 2nd June, 2016 *from* https://www.isaca.org/bookstore/.../Bookstore-2016-Audit-Catalog_bro_eng_1215.pd.

[24] Elky, S. (2006), An Introduction to Information System Risk Management; SANS Institute Engineering with DiffServ and MPLS Support: *Proceedings of the 15th International Conference on Telecommunications - ICT, St. Petersburg, Russia, 2008a.*

[25] Moray, N. (1967). Where is capacity limited? A survey and a model. *Acta Psychologica*, *27*, 84-92.

[26] Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice Hall.

[27] Norman, D., & Bobrow, D. (1975). On data-limited and resource-limited processing. *Journal of Cognitive Psychology*, *7*, 44-60.

[28] Navon, D., & Gopher, D. (1979). On the economy of the human information processing system. *Psychological Review*, *86*, 214-255.

[29] Hancock, P. A., and S. G. Hart. (2002). "Defeating Terrorism: What Can Human Factors/Ergonomics Offer?" Ergonomics in Design 10: 6–16.

[30] Hancock, P. A., and J. L. Szalma. 2003. "Vigilance and the Price of Freedom." Gateway: Human Systems Information Analysis Center 13 (5): 20.

[31] Lieberman, H. R., Castellani, J. W. &Young, A. J. (2009). "Cognitive Function and Mood during Acute Cold Stress after Extended Military Training and Recovery." Aviation, Space, and Environmental Medicine 80 (7): 629–636.

[32] Awodele, O., Onuiri, E. E. & Okolie, S. O. (2012), Vulnerabilities in Network Infrastructures and Prevention/ Containment Measures: Proceedings of Informing Science & IT Education Conference (InSITE) 2012.

[33] Abdulganiyu, A. (2012), Managing Micro-computer Systems Vulnerabilities in an Institutional Network – The Case of IBB University, Lapai, Nigeria. *International Journal of Information and Communication Technology Research; Volume 2 No. 3: 227- 234.*

[34] Panneerselvam, R. (2009), Production and Operations Management, 2nd ed. New Delhi: Asoke K. Ghosh. page. 3.

[35] Saleemi, M. A. (2013), Principles and Prectices of Management simplified; Nairobi: Printing Services Ltd page 14, 19.

[36] Parasuraman, R. (1986). Vigilance, monitoring and search. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of human perception and performance: Vol. II. Cognitive processes and performance* (pp. 41-1–41-49). New York: Wiley.

[37] Donald, F. M. (2008): The classification of vigilance tasks in the real world, Ergonomics, 51:11, 1643-1655.

[38] Beigh, B. M. & Peer, M. A. (2012), Intrusion Detection and Prevention System: Classification and Quick Review. *ARPN Journal of Science and Technology*, Vol. 2, No. 7, Pp. 661 - 675

[39] Tanenbaum, A. S. & Steen, M. V. (2014), Distributed Systems: Principles and Paradigms, 2nd ed. Edinburg Gate: Pearson Education Limited.

[40] Hancock, P. A. 2013. "In Search of Vigilance: The Problem of Iatrogenically Created Psychological Phenomenon." American Psychologist 68: 97–109.

[41] Navon, D., & Gopher, D. (1979). On the economy of the human information processing system. *Psychological Review*, *86*, 214-255.

[42] Norman, D., & Bobrow, D. (1975). On data-limited and resource-limited processing. *Journal of Cognitive Psychology*, *7*, 44-60.

[43] Lucey, T. (2005). Management Information Systems, 9th ed. BookPower: Hampshire

[44] Rue, L. W., Ibrahim, N. A. & Byars, L. L. (2013), Management Skills and Application, 4th ed.New York: McGraw-Hill Companies Inc page 5.

[45] Ward, J. & Peppard, J. (2002), Strategic Planning for Information Systems, 3rd Ed. John Wiley & Sons Ltd: Cranfield, Bedfordshire.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM

TNC TRANSACTIONS ON NETWORKS AND COMMUNICATIONS

VOLUME 6, NO. 5
ISSN: 2054 -7420

Web Browser Based Data Visualization Scheme for XBee Wireless Sensor Network

[1]Xinzhou Wei, [2]Li Geng, [3]Xiaowen Zhang
[1,2]Department of Electrical and Telecommunications Engineering Technology,
New York City College of Technology, City University of New York, 300 Jay St, Brooklyn, NY 11201, U.S.A.
[3]Department of Computer Science, College of Staten Island, City University of New York
2800 Victory Blvd., Staten Island, NY 10314, U.S.A.
xwei@citytech.cuny.edu; lgeng@citytech.cuny.edu; xiaowen.zhang@csi.cuny.edu

ABSTRACT

Wireless sensor network (WSN) plays an important role in the infrastructure of Internet of Things (IoT). Data visualization is an essential component in WSN to facilitate data scientists to interpret information clearly and efficiently. In this paper, we conduct a study of XBee based WSN which integrates the DASH data visualization scheme for building a web-browser based application without using HTML or JavaScript. The data collected from wireless sensors in a WSN were displayed in a web browser with interactive functions. The proposed visualization scheme is real-time, cross platform, and hardware independent. Thus, it could be easily employed on any operating system. Experimental results demonstrated that our WSN data visualization scheme using XBee Python package and Plotly's DASH is feasible for IoT applications like smart buildings, environment monitoring, as well as other WSN applications.

Key words: Data visualization, Internet of Things, Smart building, Wireless sensor network, XBee, ZigBee.

1 Introduction

Wireless sensor network (WSN) has been widely used in the infrastructure of Internet of Things (IoT) [1-3]. It has been applied to the fields related to smart building management [4], environment monitoring [5], energy monitoring [6], health monitoring [7], and precision agriculture [8].

Data visualization is an appealing way to interpret the data in an illustrative and graphical form. It typically converts texts and numbers to aesthetically pleasing visual elements, thus makes them easy to be recognized by human beings [9]. This is the most important reason why data visualization is so compelling to data scientists, data engineers, and researchers. Compared to conventional text reading and number processing by human brain, data visualization can take the same information and make patterns more understandable and readily perceptible [10].

With increased demand in the application of WSN, data visualization has become a key component of sensor networks. As dozens or hundreds of sensors generate a large amount of data in WSN, a powerful visualization tool will help data scientists or decision makers to spot a special event or recognize some unique patterns quickly and efficiently. In the past decades, researchers have developed a lot of visualization tools for WSN, such as TinyViz [11], SpyGlass [12], MoteView [13], MeshNetics [14],

DOI: 10.14738/tnc.65.5261
Publication Date: 18th October 2018
URL: http://dx.doi.org/10.14738/tnc.65.5261

MonSense [15], NetTopo [16], WiseObserver [17], Octopus [18], and Surge Network Viewer [19], *etc.* Although these visualization tools are wonderful, most of them are stand-alone computer applications and have high demand for computer hardware and resources. Some of them were developed on a specific operating system or hardware platform. While others cannot display the data in real- time and are only suitable for static data. Furthermore, some of those visualization tools do not provide interactive functionalities to facilitate user-friendly interface [20-23]. Thus, there is a strong demand for a cross platform data visualization tool which shall display the real time data in WSN interactively. In addition, another objective of this study is to develop a powerful data visualization tool for WSN without the need of advanced programming techniques from the users. As a result, the DASH data visualization framework from Plotly is the best solution for the above criteria.

1.1 Current Data Visualization Tools in Python

The evolution of computers and digital technology makes data visualization possible to process large amounts of data in real-time with interactive rendering components [10]. One big challenge, particularly for data scientists or data engineers, is to represent the innovative ideas via visual means without concerning too much about programming skills and the platforms. Python, a simple open source and cross platform script language, fulfills the requirements for non-programming professionals. The users of Python in scientific fields increased exponentially in the past five years. There are a couple of cross platform visualization tools developed in Python in recent years, such as Pandas, Matplotlib, Seaborn, Bokeh, Pygal, PyQtGraph, Plotly, and VisPy [24]. These visualization tools have demonstrated their technological abilities in different technology fields.

1.2 Python DASH Visualization Scheme

The visualization tools mentioned above are excellent for data scientists or data engineers, but sometimes there is a strong need to create interactive visualizations and more dynamically explore data like surfing on the Internet via web browser [25]. Interactive data visualization tools allow users to explore and analyze data with greater freedom and flexibility. DASH is a cross platform, web browser based, and open source framework created by the Plotly team that leverages Flask, Plotly.js and React.js to build custom data visualization applications in Python [26]. DASH provides most attractive interactivity functions without using HTML or JavaScript. Users can manipulate data in a web browser based environment and can seamlessly take advantage of their experience on web surfing. It greatly speeds up the development cycle, simplifies the development difficulty, and shortens the learning curve for data scientists and data engineers [26].

1.3 Characteristics of Plotly DASH

DASH was recently released in June 2017 by the Plotly team as an open source library. Built on top of Plotly.js, React.js, and Flask, DASH is ideal for the users to develop web-based visualization application interactively [26]. Developers can use Python solely for building interactive web browser based applications. No HTML or JavaScript programming skill is required. Other characteristics of DASH include that developers can adopt all the plotting capabilities which are user-friendly via Plotly’s Python framework, access other Python libraries such as Matplotlib or Pandas, and use all of its powers in their visualization tool impeccably. This open source model was adopted widely by data science and industry in recent years. Plotly makes this open source package available publicly on GitHub for those individuals like data scientists, researchers, and college students [27]. The technical supports, such as training and large scale deployments, are also available from Plotly based on developer's requests [26].

The remainder of this paper is organized as follows: the configuration of XBee modules in WSN, WSN topology, the data collection method for the XBee based WSN, and the structure of Python DASH visualization scheme are presented in Section 2. Section 3 reports the experimental results of wireless temperature monitoring system for smart buildings. The conclusion and discussion are provided in Section 4.

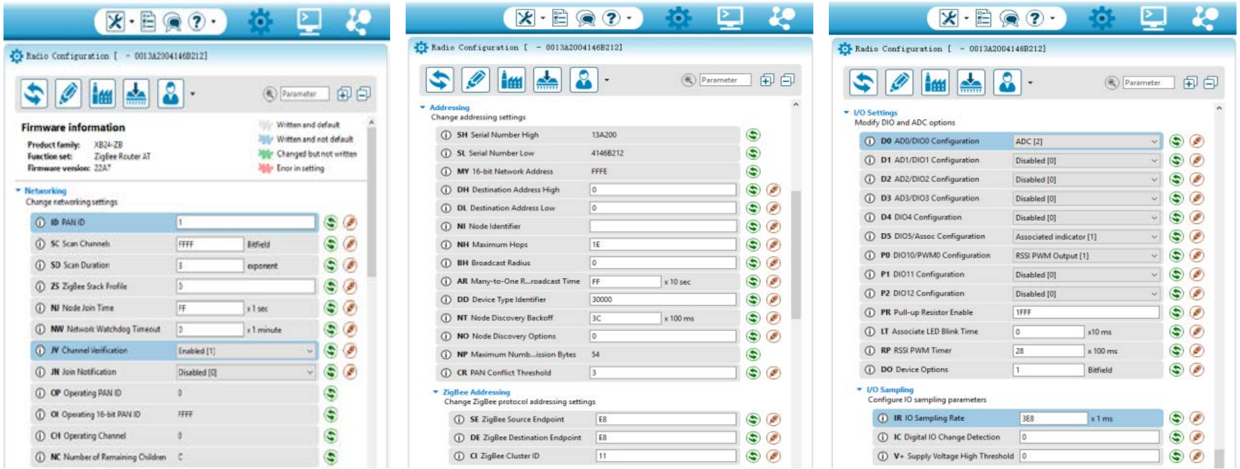
2 Methods

In this study, we propose an XBee based WSN, in which Python DASH data visualization framework has been integrated to build a cross platform, web-browser based interactive WSN environmental monitoring system without using HTML or JavaScript. For the configuration of XBee modules in WSN, one XBee module works as the coordinator that acts like the root of a tree. The coordinator collects all information forwarded by routers in WSN and sends them to a computer via USB connection, whereas a router relays and forwards the information collected by different wireless end nodes in the WSN. Its configuration will be different from that of the coordinator. In addition, all XBee modules in the same WSN should contain the same PAN ID number. Detailed procedure of XBee coordinator and router configuration is described in Faludi's work [28].

Fig. 1 and Fig. 2 show the configurations of an XBee coordinator and a router, respectively, from the XCTU (a free application created by Digi International Inc. for the XBee module programming, configuration, troubleshooting, and network management.) Details of the XCTU application can be found at http://www.digi.com/products/wireless-wired-embedded-solutions/ZigBee-rf-modules/xctu. The parameters of WSN coordinator and router in XCTU configuration interface are shown in Fig. 1 and Fig. 2, respectively. The PAN ID is configured as 1 for both XBee coordinator and routers for the simplicity of demonstration purpose. XCTU is a pure graphic user interface and each parameter of XBee module could be set individually. User can set up a parameter by clicking the textbox on the right side of the parameter label where XCTU will pop up a menu and allow the user to select certain values in the menu. XCTU also uses different colors to indicate the type of the value.

2.1 The XBee Router Configuration

To configure the XBee module as a router, firstly, we need to mount the XBee module on an adapter called XBee explorer board. Developers can obtain the XBee explorer board from Sparkfun Electronic Inc. or Adafruit Inc. Secondly, we attach a mini USB cable from this adapter and connect it to a PC. There is an onboard FT231X USB-to-Serial converter that translates data between XBee and the PC. Finally, we start the XCTU that will automatically detect the XBee module connected via USB port. User could have an option to update the firmware of the XBee module after a successful detection. All routers must have the same PAN ID number with coordinator in the same WSN. Otherwise they will not communicate with each other. In our XBee router, the value of pin DIO0 is configured as ADC input port that has a value 2 as shown on the first textbox of right screen snapshot in Fig. 1.



Radio Configuration [- 0013A2004146B212]

Firmware information
Product family: XB24-ZB
Function set: ZigBee Router AT
Firmware version: 22A7

Networking – Change networking settings
ID PAN ID: 1
SC Scan Channels: FFFF Bitfield
SD Scan Duration: 3 exponent
ZS ZigBee Stack Profile: 0
NJ Node Join Time: FF x 1 sec
NW Network Watchdog Timeout: 0 x 1 minute
JV Channel Verification: Enabled [1]
JN Join Notification: Disabled [0]
OP Operating PAN ID: 0
OI Operating 16-bit PAN ID: FFFF
CH Operating Channel: 0
NC Number of Remaining Children: C

Radio Configuration [- 0013A2004146B212]

Addressing – Change addressing settings
SH Serial Number High: 13A200
SL Serial Number Low: 4146B212
MY 16-bit Network Address: FFFE
DH Destination Address High: 0
DL Destination Address Low: 0
NI Node Identifier:
NH Maximum Hops: 1E
BH Broadcast Radius: 0
AR Many-to-One R...roadcast Time: FF x 10 sec
DD Device Type Identifier: 30000
NT Node Discovery Backoff: 3C x 100 ms
NO Node Discovery Options: 0
NP Maximum Numb...ission Bytes: 54
CR PAN Conflict Threshold: 3

ZigBee Addressing – Change ZigBee protocol addressing settings
SE ZigBee Source Endpoint: E8
DE ZigBee Destination Endpoint: E8
CI ZigBee Cluster ID: 11

Radio Configuration [- 0013A2004146B212]

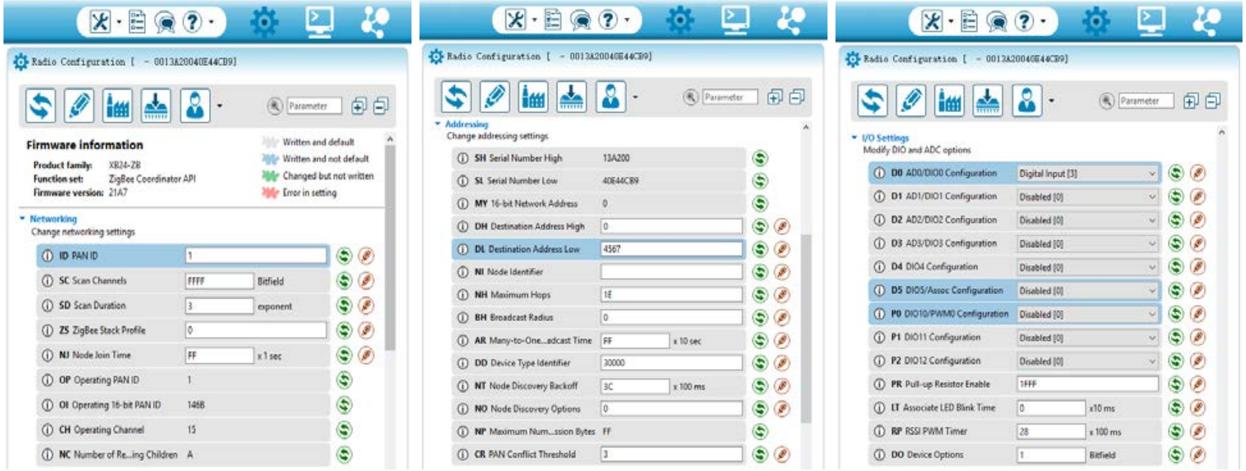
I/O Settings – Modify DIO and ADC options
D0 AD0/DIO0 Configuration: ADC [2]
D1 AD1/DIO1 Configuration: Disabled [0]
D2 AD2/DIO2 Configuration: Disabled [0]
D3 AD3/DIO3 Configuration: Disabled [0]
D4 DIO4 Configuration: Disabled [0]
D5 DIO5/Assoc Configuration: Associated indicator [1]
P0 DIO10/PWM0 Configuration: RSSI PWM Output [1]
P1 DIO11 Configuration: Disabled [0]
P2 DIO12 Configuration: Disabled [0]
PR Pull-up Resistor Enable: 1FFF
LT Associate LED Blink Time: 0 x10 ms
RP RSSI PWM Timer: 28 x 100 ms
DO Device Options: 1 Bitfield

I/O Sampling – Configure IO sampling parameters
IR IO Sampling Rate: 3E8 x 1 ms
IC Digital IO Change Detection: 0
V+ Supply Voltage High Threshold: 0

Fig. 1. Configuration of XBee Router: Network addresses setting (left and middle) and the I/O ports setting (right).

2.2 The XBee Coordinator Configuration

*The XBee coordinator configuration interface in Fig. 2 shows that the XBee module has been set to API mode and the PAN ID of the coordinator must have the same number with the XBee routers in the WSN.



Radio Configuration [- 0013A20040E44CB9]

Firmware information
Product family: XB24-ZB
Function set: ZigBee Coordinator API
Firmware version: 21A7

Networking – Change networking settings
ID PAN ID: 1
SC Scan Channels: FFFF Bitfield
SD Scan Duration: 3 exponent
ZS ZigBee Stack Profile: 0
NJ Node Join Time: FF x 1 sec
OP Operating PAN ID: 1
OI Operating 16-bit PAN ID: 146B
CH Operating Channel: 15
NC Number of Re...ing Children: A

Radio Configuration [- 0013A20040E44CB9]

Addressing – Change addressing settings
SH Serial Number High: 13A200
SL Serial Number Low: 40E44CB9
MY 16-bit Network Address: 0
DH Destination Address High: 0
DL Destination Address Low: 4567
NI Node Identifier:
NH Maximum Hops: 1E
BH Broadcast Radius: 0
AR Many-to-One...adcast Time: FF x 10 sec
DD Device Type Identifier: 30000
NT Node Discovery Backoff: 3C x 100 ms
NO Node Discovery Options: 0
NP Maximum Num...ssion Bytes: FF
CR PAN Conflict Threshold: 3

Radio Configuration [- 0013A20040E44CB9]

I/O Settings – Modify DIO and ADC options
D0 AD0/DIO0 Configuration: Digital Input [3]
D1 AD1/DIO1 Configuration: Disabled [0]
D2 AD2/DIO2 Configuration: Disabled [0]
D3 AD3/DIO3 Configuration: Disabled [0]
D4 DIO4 Configuration: Disabled [0]
D5 DIO5/Assoc Configuration: Disabled [0]
P0 DIO10/PWM0 Configuration: Disabled [0]
P1 DIO11 Configuration: Disabled [0]
P2 DIO12 Configuration: Disabled [0]
PR Pull-up Resistor Enable: 1FFF
LT Associate LED Blink Time: 0 x10 ms
RP RSSI PWM Timer: 28 x 100 ms
DO Device Options: 1 Bitfield

Fig. 2. Configuration of XBee Coordinator: Network addresses setting (left and middle) and the I/O ports setting (Right)

There are two parts in the MAC addresses for both coordinator and routers: 1) The High part is always "13A200" which was assigned by Digi International Inc; 2) The Low address is a hexadecimal number based on that XBee module. The network addressing configuration values and I/O ports configuration values are shown in Fig. 2. In our WSN system, the pin of DIO0 for the XBee coordinator is configured as a digital input port, which has a value of 3 as shown on the first textbox of right screen snapshot in Fig. 2.

2.3 Wireless Sensor Network Topology

After we successfully configure both XBee coordinator and routers, we will test the connections and display the network topology of our XBee based WSN. The details of hardware are described in our earlier paper [29]. XCTU's Network function allows user to discover and visualize the topology and interconnections of a WSN. To display the topology of our WSN, we switch the setting to "Network working mode" and click the "Scan" radio button to start the network discovery process. XCTU will scan the entire XBee based WSN and display the logical connections and link quality as shown in Fig. 3. For the demonstration purposes, in this paper, we only use a couple of XBee modules in our WSN system

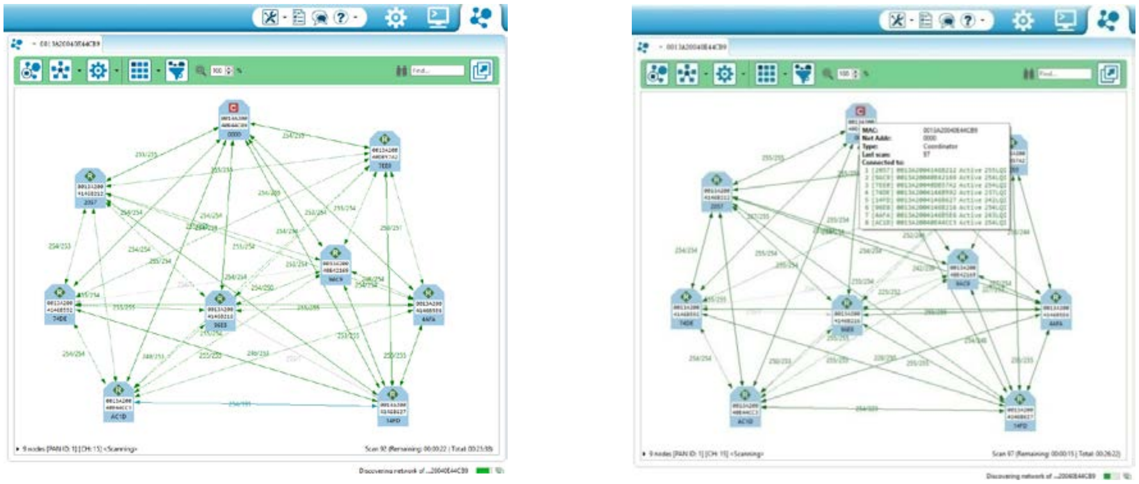


Fig. 3. Wireless sensor network topology and connection table in the XCTU Network Interface: WSN topology (left) and topology and connection table (right).

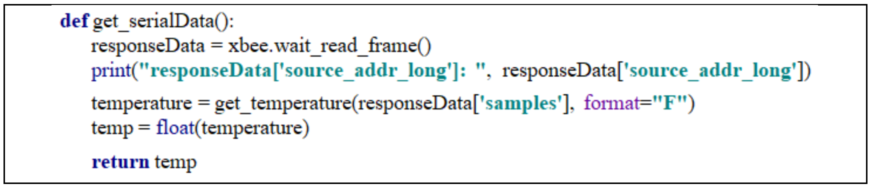
The coordinator is represented by letter 'C' in red color and the router is represented by letter 'R' in green color in the XBee icons. The wireless sensor network shown in Fig. 3 (left) is a real network and the links and their signal qualities are detected by XCTU. To find out the detailed connection information for a specific XBee module, we can click any XBee icon in XCTU network window, a connection table would pop up and display all of the connections with that XBee module as shown in Fig. 3 (right).

All these XBee modules must be XBee S2, XBee S2 Pro version or newer models that make a mesh wireless sensor network. It is noted that XBee S1 version only supports point to point communication and thus cannot be used in a mesh WSN.

After we finish setting up our WSN system, we will collect data and build a data visualization system using Plotly DASH to display data in a web browser.

2.4 Data Collection of XBee Based Wireless Sensor Network

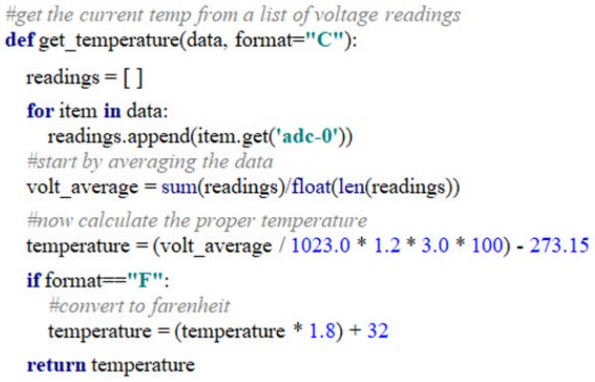
In order to collect data from serial port in our WSN in Python, we need to include XBee, ZigBee, and serial packages. The tool to install these packages in Python IDE is called pip. In this study, we use PyCharm Community 2017, an open source Python IDE to manage our project. As shown in the code snippet of Fig. 4, a Python function called get_serialData() will read XBee data frame collected from the sensors via serial port. First, we called the xbee.wait_read_frame() function in Python XBee package distributed by Digi International Inc. and saved it in a variable "responseData."



```
def get_serialData():
    responseData = xbee.wait_read_frame()
    print("responseData['source_addr_long']: ", responseData['source_addr_long'])
    temperature = get_temperature(responseData['samples'], format="F")
    temp = float(temperature)
    return temp
```

Figure 4. Code snippet of XBee module data collection from serial port.

Then, we applied a user defined function called get_temperature() to extract the temperature and convert it in Fahrenheit. The code snippet of get_temperature() function is listed in Fig. 5.



```
#get the current temp from a list of voltage readings
def get_temperature(data, format="C"):

    readings = [ ]
    for item in data:
        readings.append(item.get('adc-0'))
    #start by averaging the data
    volt_average = sum(readings)/float(len(readings))
    #now calculate the proper temperature
    temperature = (volt_average / 1023.0 * 1.2 * 3.0 * 100) - 273.15
    if format=="F":
        #convert to farenheit
        temperature = (temperature * 1.8) + 32
    return temperature
```

Fig. 5. Code snippet of XBee module temperature conversion from XBee data frame.

In the code snippet of the function get_temperature(), the temperature was calculated by the following temperature sensor's formula:

temperature = (volt_average/1023*1.2*3*100) - 273.15

where volt_average is the analog input of the sensor. Since there are a couple of temperature sensors in our WSN system, we can extract both XBee's MAC addresses and the corresponding temperature values in the XBee frame to distinguish different sensors before performing data visualization.

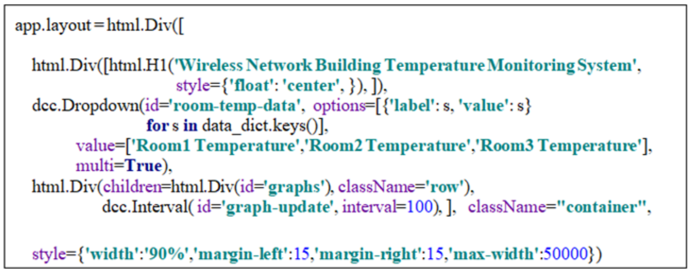
2.5 Structure of the Python DASH Visualization Scheme

In this paper, we employed the DASH visualization scheme, which is an open source library using Python framework to build web browser based applications. There is no HTML or JavaScript needed in this structure. Written on top of Flask, Plotly.js, and React.js, DASH is ideal for building data visualization applications with better user interfaces in pure Python [30]. With DASH, we can create cross platform, web-based interactive applications in pure Python. It is particularly suitable for data scientists or researchers who work in data visualization field in Python. Specifically, Plotly keeps a set of visual components in dash_core_components and dash_html_components library. We imported these packages for developing DASH visualization applications in this study.

2.5.1 DASH layout

There are two parts in the Python Dash Visualization application. The first part is the layout of the application and it decides what the application looks like. The second part decides the interactivity of the DASH application. Furthermore, the DASH layout could be divided in three blocks: Header, Dropdown menu, and Graph. It is composed of a tree of components like html.Div and dcc.Graph. The dash_html_components library has a component for each HTML tag.

In this study, we adopted some codes publicly available to researcher & developer [27][30]. As shown in the code snippet of Fig. 6, the header of the webpage could be created by html.Div and html.H1 with suitable style. The html.H1(children = 'Wireless Network Building Temperature Monitoring System') component generates a HTML element (<h1>Wireless Network Building Temperature Monitoring System </h1>) in the DASH application. The contents in the Dropdown menu will be decided by dcc.Dropdown block with corresponding id value and keys in the dictionary. The display interval and style could also be decided in this step.



```
app.layout = html.Div([

   html.Div([html.H1('Wireless Network Building Temperature Monitoring System',
                     style={'float': 'center', }), ]),
   dcc.Dropdown(id='room-temp-data',  options=[{'label': s, 'value': s}
                for s in data_dict.keys()],
         value=['Room1 Temperature','Room2 Temperature','Room3 Temperature'],
         multi=True),
   html.Div(children=html.Div(id='graphs'), className='row'),
             dcc.Interval( id='graph-update', interval=100), ],   className="container",


   style={'width':'90%','margin-left':15,'margin-right':15,'max-width':50000})
```

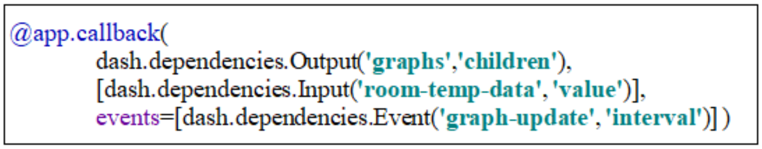
Fig. 6. Code snippet of DASH layout.

2.5.2 DASH core components

The dash_core_components includes a set of higher-level components like dropdown, graph, markdown block, etc. Graph renders interactive data using the open source Plotly.js, a JavaScript graphing library. Plotly.js supports over 35 chart types and renders charts so far in both vector-quality SVG and high-performance WebGL[26]. Graph component is the same figure argument that is used by Plotly.py, a Plotly's open source Python graphing library [26].

2.5.3 DASH interaction method

After the DASH layout is created, we map out the interaction among various DASH components. DASH provided app.callback() decorator to fulfill this requirement. We employed DASH "callback" to bind interactive components such as dropdowns, graphs, sliders, and text inputs in its application. As shown in the code snippet of Fig. 7, the parameters we pass into the app.callback decorator include output components and properties we want to update plus a list of all the input components and properties that can be used to trigger the function.



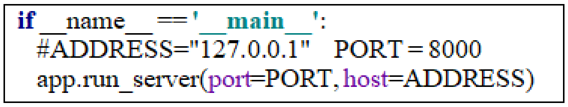
```
@app.callback(
          dash.dependencies.Output('graphs','children'),
          [dash.dependencies.Input('room-temp-data', 'value')],
          events=[dash.dependencies.Event('graph-update', 'interval')] )
```

Fig. 7. Code snippet of DASH callback.

With DASH interactivity, we can dynamically update any of those properties through the callback function. That is to say, we could not only update the children's values of a component and display new text or figure of a dcc.Graph component, but also update the style of the component or even the option values of a dcc.Dropdown component. After laying out all of the above components, we define the figure using a dictionary that contains the figure as well as the data and layout options. Details of these configurations could be found at www.pythonprogramming.net [27].

2.5.4 Launch of the DASH Application

As shown in Fig. 8, we provided the local host's IP address and the port number for the browser to make sure it can find and display the data in the proper location when we launch the DASH application. We launched a web browser and entered address 127.0.0.1:8000 in the address box. The real-time data collected from XBee based WSN will be displayed in web browser shortly and refreshed periodically.



```
if __name__ =='__main__':
  #ADDRESS="127.0.0.1"   PORT = 8000
  app.run_server(port=PORT, host=ADDRESS)
```

Fig. 8. Code snippet of launching project.

3 Experimental Results

The XBee based DASH visualization scheme for WSN could be employed on any web browser like Internet Explorer, Firefox, or Google Chrome. We tested our visualization system displaying the temperature values from different rooms in a smart building with screen snapshots as shown in Fig. 9. For the arrangement of the display, when there's only one graph, DASH will take entire row of the browser. If we select to display two values in the dropdown menu from top, then each value will take half size of the browser horizontally. When we select more values to display, DASH will arrange the graphics in a matrix format.

As shown in Fig. 10, when a user performs a mouse over on a specific graph, DASH will show a group of interactive tools, such as dropdown menu, text box, zoom in, zoom out, auto size, download figure, expand the size of figure, etc. in the browser to display the value in a pop-up text box and show the temperature value. We could zoom in and zoom out over a selected area and check more detailed information. We could also deselect a specific graph from dropdown menu and remove it from visualizing.

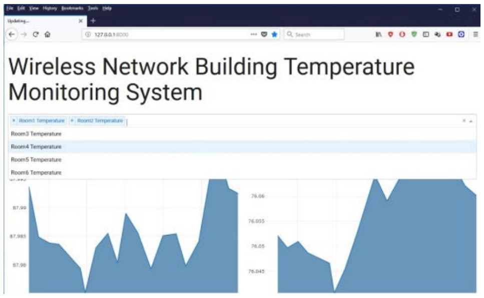


Fig. 9. Visualization system displaying the temperature values from different rooms in a smart building.

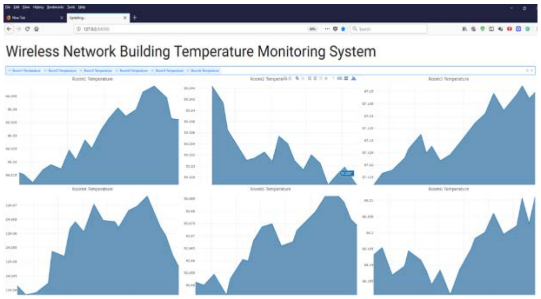
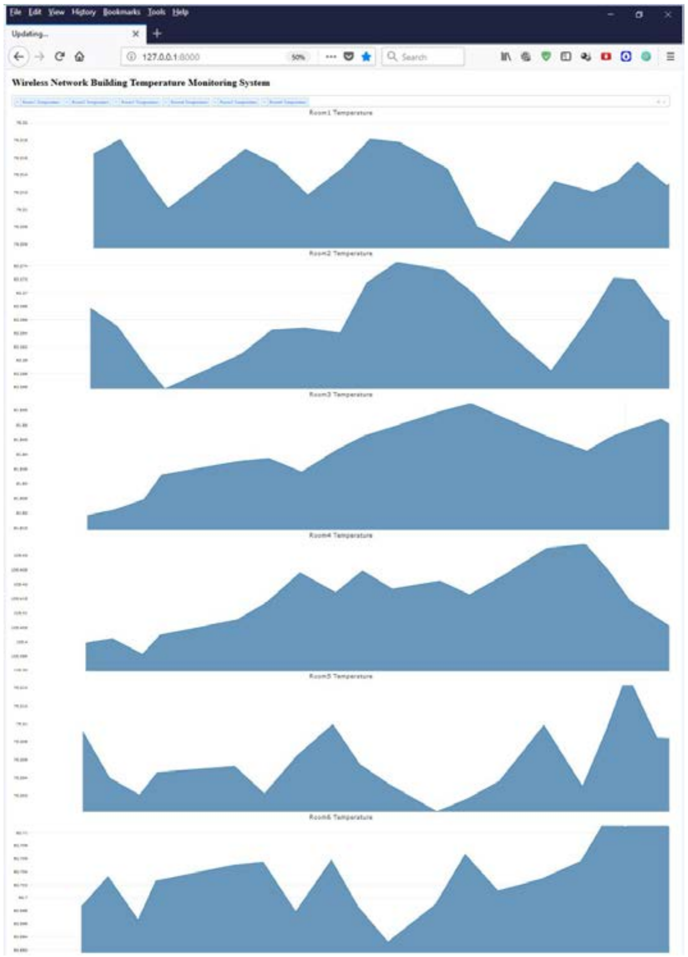


Fig. 10. Visualization system displaying room temperature values with interactivity function.

If we prefer to set up the computer monitor vertically from pivot, we can display all values vertically as shown in Fig. 11.



Wireless Network Building Temperature Monitoring System

Fig. 11. Visualization system displaying values vertically in the browser with a snapshot of full screen.

4 Conclusions and Discussion

In this paper, we present a web browser based data visualization scheme for WSN. Experiment results demonstrated that the data collected from wireless sensors in a WSN were displayed in a web browser with interactive functions. We can select the temperature data from different rooms in a drop-down menu and visualize them in real- time. We have the option to add or remove some graphs to display in the drop-down menu. Finally, we can choose the layout on the page depending on how many charts we want to display.

Our work using this DASH visualization scheme can be extended by further improvement of uploading data of WSN on cloud and then scraping down to any local machine for visualization in a web browser. Future improvements include adoption of X4 Portconnect and wireless nodes from Digi International Inc. to simplify the hardware and improve the reliability of the system.

ACKNOWLEDGEMENT

This work was partly supported by CUNY GRTI grant round 20 and PSC-CUNY grant #61163-00 49, jointly funded by the City University of New York.

REFERENCES

[1] Minoli D, Sohraby K, Occhiogrosso, B. IoT considerations, requirements, and architectures for smart buildings—Energy optimization and next-generation building management systems. IEEE Internet of Things, 2017, 4(1): 269-283.

[2] Zanella A, Bui N, Castellani A, Vangelista L, Zorzi M. Internet of Things for smart cities. IEEE Internet of Things, 2014, pp. 22-32.

[3] Mitton N, Papavassiliou S, Puliafito A, Trived K S. Combining cloud and sensors in a smart city environment. EURASIP Journal on Wireless Communications and Networking, 2012, p. 247.

[4] Ghayvat H, Mukhopadhyay S, Gui X, Suryadevara N. WSN- and IOT-based smart homes and their extension to smart buildings. Sensors, 15(5): 10350–10379.

[5] Jang W S, Healy W M, Skibniewski M J. Wireless sensor networks as a part of a web-based building environmental monitoring system. Automation in Construction, 2008, 17(6):729-736.

[6] Liu X, Chen H, Wang M, Chen S. An XBee-Pro Based Energy Monitoring System. Brisbane, Australasian Telecommunication Networks and Applications Conference (ATNAC), 2012, pp. 1-6.

[7] Othman S B, Trad A, Youssef H. Security architecture for at-home medical care using wireless sensor network.. International Wireless Communications and Mobile Computing Conference (IWCMC), 2014, pp. 304-309.

[8] Davcev D, Mitreski K, Trajkovic S, Nikolovski V, and Koteli N. IoT agriculture system based on LoRaWAN. 14th IEEE International Workshop on Factory Communication Systems (WFCS). 2018.

[9] Shamas N. Why data visualization is important. TechChange, May 19, 2015. https://www.techchange.org/2015/05/19/data-visualization-analysis-international-development/.

[10] Link A. Why create visualizations of your data? Data Visualization, 2017. https://dash.umn.edu/data-visualization/.

[11] Levis P, Lee N, Welsh M, Culler D. TOSSIM: accurate and scalable simulation of entire TinyOS applications. In Proc. the 1st International Conference on Embedded Networked Sensor Systems, 2003, pp. 126-137.

[12] Buschmann C, Pfisterer D, Fischer S, Fekete S P, Kroller A. SpyGlass: a wireless sensor network visualizer. ACM SIGBED Review, 2005, 2(1): 1-6.

[13] Tuton M. MOTE VIEW: A sensor network monigoring and management tool. Second IEEE Workshop on Embedded Networked Sensors (EmNetS-II), 2005, pp. 11-18.

[14] Leonov A. MeshNetics demonstrated integration of wireless sensor data with SCADA system at ZigBee open house. CISION - PRWeb, June 2006. https://www.prweb.com/releases/2006/06/prweb403245.htm.

[15] Pinto J, Sousa A, Goncalves G M, Lebres P, Sousa J. MonSense-application for deployment, monitoring and control of wireless sensor networks. ACM Real Wireless Sensor Network Conference, 2006.

[16] Shu L, Wu C, Zhang Y, Chen J, Wang L, Hauswirth M. NetTopo: Beyond Simulator and Visualizer for Wireless Sensor Networks. IEEE Second International Conference on Future Generation Communication and Networking, 2008, pp. 17-20.

[17] Castillo J A, Ortiz A M, Lopez V, Olivares T, Orozco-Barbosa L. WiseObserver: a real experience with wireless sensor networks. In Proc. the 3nd ACM workshop on Performance Monitoring and Measurement of Heterogeneous Wireless and Wired Networks, 2008, pp. 23-26.

[18] Jurdak R, Ruzzelli A G, Barbirato A, Boivineau S. Octopus: monitoring, visualization, and control of sensor networks. Wireless Communications & Mobile Computing, 2011, 11: 1073-1091.

[19] Surge Network Viewer. By Crossbow Technology, Inc. http://www.hoskin.qc.ca/uploadpdf/ Instrumentation/divers/CrossBow/divers_Surge%20Network%20Viewer_4271286a0135f.pdf.

[20] Parbat B, Dwivedi A K, Vyas O.P. Data visualization tools for WSNs: A glimpse. International Journal of Computer Applications, 2010, 2(1): 14-20.

[21] dAuriol B J, Lee S, Lee Y. K. Visualizations of Wireless Sensor Network Data. In Handbook of Research on Developments and Trends in Wireless Sensor Networks: From Principle to Practice. Hershey : IGI Global, 2010, Chapter 16, pp. 353-370.

[22] ElHakim R, ElHelw M. Interactive 3D visualization for wireless sensor networks. 6-8, June 2010, The Visual Computer, 2010, Vol. 26, Issue 6-8, pp. 1071-1077.

[23] Ravichandranb S, Chandrasekarb R K, Uluagac A S, Beyah R. A simple visualization and programming framework for wireless sensor networks: PROVIZ. Ad Hoc Networks, 2016, 53: 1-16.

[24] Moffitt C. Overview of Python visualization tools. Practical Business Python, 2015. http://pbpython.com/visualization-tools-1.html.

[25] Moffit C, Choosing a Python visualization tool. Practical Business Python, 2018. http://pbpython.com/python-vis-flowchart.html.

[26] Moffitt, C. Creating interactive visualizations with Plotly's Dash framework. Practical Business Python, 2017. http://pbpython.com/plotly-dash-intro.html.

[27] Kinsley H. Introduction to data visualization applications with Dash and Python. Pythonprogramming.net, 2018. https://pythonprogramming.net/data-visualization-application-dash-python-tutorial-introduction/.

[28] Faludi R. Building Wireless Sensor Networks: with ZigBee, XBee, Arduino, and Processing. O'Reilly Media, 2010.

[29] Wei X, Geng L, Zhang X. An open source data visualization system for wireless sensor network. Journal of Computer Science and Information Technology, 2017,5(2): 10-17.

[30] Parmer C. Introduction to DASH. Build beautiful web-based interfaces in Python. 2016. https://dash.plot.ly/introduction.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM

TNC TRANSACTIONS ON NETWORKS AND COMMUNICATIONS

VOLUME 6, NO. 5
ISSN: 2054 -7420

Consumer Perceptions towards Online Retailing in Botswana: A Case Study of the University of Botswana

Mashoko T. Dzimiri, Katlego A. Thamage, Mogotsinyana Mapharing, Elang Basuhi, Ishmael Radikoko
University of Botswana, Department of Accounting and Finance
mashokodzimiri@gmail.com; thamagekatlego@gmail.com; mogotsinyana.mapharing@mopipi.ub.bw
basuhie@ub.ac.bw; radikokoi@ub.ac.bw

ABSTRACT

The study was premised on assessing the existing customer perceptions towards online retailing in Botswana. In particular, the study used University of Botswana staff and students as a case study. Data for this study was collected through questionnaires given to students and staff of the University of Botswana. Descriptive statistics and Independent t-tests were used to analyse the results. The results from this study indicate that consumers in the University of Botswana utilise traditional shopping more than online shopping. Though accessible by people of all income, online shopping is perceived to be risky and needs a skillful internet user. Delivery concerns, technology specific innovativeness and financial risk were found to negatively influence online shopping behavior. On the other hand, subjective norms, good return policy and convenience were found to positively influence online shopping behavior. The findings from this study provide a first glance at existing customer perceptions in Botswana which should encourage further and more extensive research to yield more generalizable results that reflect consumer perceptions. Local retailers or businesses will find these results useful as these will help them narrow their focus on how to sway consumers to use their online shops.

Key words: consumer perceptions, online retailing, factors, descriptive analysis, University of Botswana

1 Introduction

The internet has without doubt been considered one of the 50 Greatest breakthroughs since the invention of the Wheel [1]. In that light, the internet has changed several aspects of human life particularly how we interact, learn and govern ourselves and especially how business is conducted[2]. A great example of such technological change in business conduct is the emergence of online retailing which started off as a mere alternative method of selling products. Consequently, it has grown phenomenally over the years from a feasible sales channel to a lucrative retail sector that is redefining the retail landscape. For example, through online retailing, traditional retail businesses have managed to tap into other markets beyond geographical limitations, reduce overhead costs, generate more revenues, increase market share and increase online presence [3] . Additionally, customers benefit more from the improved customer service delivered through online retailing.

The prominence of online retailing is especially evident in more developed countries particularly China and the United States. In 2014, the two countries jointly controlled 55% of global internet retail sales,

DOI: 10.14738/tnc.65.5292
Publication Date: 20th October 2018
URL: http://dx.doi.org/10.14738/tnc.65.5292

which in total were expected to top an astounding estimate of $22 trillion with the figure being expected to grow in the following years[4]. Examples of such online retail businesses pioneering this exemplary performance included brick and mortar companies which adopted online retailing such as Wal-Mart stores Inc. (USA) and Tmall (China); as well as self-operating website businesses such as Amazon (USA) and Alibaba (China), all of which earned billions in remarkable sales and revenue performance through online retailing [5, 6]. Further to that, [7] posits that online retailing shops have substantially benefitted from this sales alternative, such that companies such as Amazon.com have achieved superior performance making it the 8th largest retailer in the world.

In the African context, the ever increasing development in internet connectivity infrastructure and online usage has increased phenomenally in developing countries, particularly those in the African continent. The number of internet users in Africa was estimated to be close to 453 million, conveying an astounding growth rate of 9,942% from 2000 to 2017 as at 31 December 2017 [8]. Such increase in internet usage has led to the recent adoption and emergence of internet retailing albeit its infancy stage. Despite such, online retailing has been adopted by a majority of African countries such as Egypt, Nigeria, South Africa, Angola, and Kenya which stand as Africa's leaders in online retail, with a combined total online retail sales of close to $4 billion [9] .

Early adopters of the online retailing concept include a majority of the top 10 retailers in Africa which happen to be notable brick and mortar companies such as Shoprite Holdings Ltd, Woolworths Holdings Ltd, and Pick n Pay Stores Ltd. The three retailers performed exceptionally well in sales and revenue in 2013. Another interesting fact is these all originate from South Africa[9]. Notable online retail website companies such as Jumia, Kaymu and Konga have also seen prosperity in online retailing growing beyond their country of origin (Nigeria) and expanding across the African continent[9]. Similarly, internet usage in Botswana has significantly increased over the years due to much improvement in internet connectivity by government and internet providing companies. This is conveyed by 36.7% of the population using the internet[10]. By extension, the researchers intend to explore online retail usage perceptions in Botswana context. We extend on the conceptual model used by [11] in assessing factors affecting Indian consumers' online buying behavior by incorporating other perceptual dimensions such as demographics.

1.1 Problem Statement

Recently, companies in Botswana have started to adopt online retailing which includes Sefalana Holdings, which is the first brick and mortar retail group to launch an online shopping site in Botswana. Online shopping companies such as shop360 and Skymartbw (Pty) Ltd have also established themselves in an effort to pioneer and benefit from the prospect of online retailing in Botswana. However, despite such emergence of such promising online retailers, online usage statistics indicate that out of the entire internet user population in 2014, only 7.4% used the internet to purchase goods and services[10]. As a result, the key question is what are the factors influencing or deterring consumers to utilise online retailing stores in Botswana?

1.2 Research Objectives

The overall objective of this paper is to explore consumer perceptions on online retailing in Botswana. In particular, this study aims to:

1. Identify and assess the existing customer perceptions towards online retailing

1.3 Significance of the Study

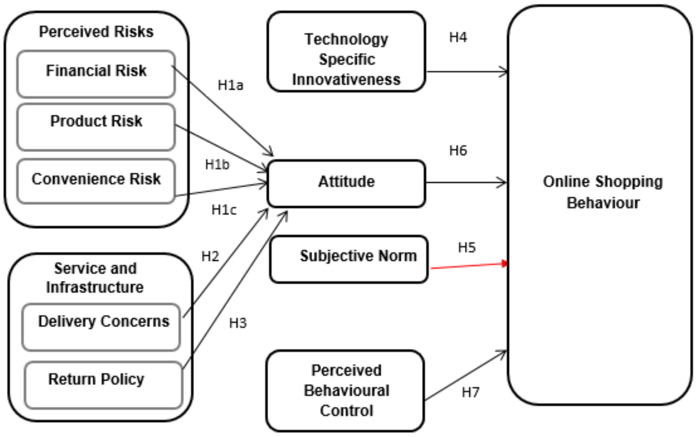
The study is focused on the consumer perceptions towards online retailing among University of Botswana students and employees. As already outlined, online retailing is still at its infancy stage in Botswana. Nevertheless, Botswana retail businesses could greatly benefit from such a lucrative online presence. Therefore it is imperative that online retail businesses understand their target market in order to succeed in online retailing. Thus, this study aims to provide valuable insight into the perceptions that customers hold towards online retailing. Businesses can then use such valuable knowledge to better market themselves to their target market and increase their customer base, revenues and profitability. More importantly, the Botswana economy could greatly benefit from the success of these online businesses as this will provide needed tax revenue and thus contribute towards the GDP, promote diversity, and increase export presence in the global market.

This study is divided into five sections. Section 1 is the Introduction and section 2 covers Literature review. Methodology is covered in section 3 and section 4 covers the Analysis of data, findings and discussions. Meanwhile, conclusions and recommendations are covered in section 5.

2 Literature Review

2.1 Theoretical Review

The Theory of Planned Behaviour (TPB) has been used in prior studies to assess consumer perceptions and online retailing (Azjen, 1985, 1991; Azjen & Fishbein, 1980), as cited in [11]. However, conceptual model proposed to assess perceptions of consumers towards online retailing in Botswana is the modified Theory of Planned Behavior (TPB). This conceptual framework was particularly used by [11] in their study assessing factors affecting Indian consumers' online buying behavior. The theory is appropriate for this study as it explains behaviors over which individuals have incomplete voluntary control. In other words, the theory predicts deliberate behavior due to the fact that behavior can be deliberate and planned [12] The conceptual model is presented in Figure 1.



Source: Sinha & Kim, 2012

Figure 1: Conceptual model of factors influencing Botswana shoppers' online shopping behavior

This modified Theory of Planned Behavior assumes that online shopping behavior is influenced by several factors, the four most common being attitude, subjective norm, perceived behavioral control and technology specific innovativeness as discussed below:

Attitude: [11]), note that a consumer attitude towards performing a behavior has been proven as a strong predictor of behavior. The researchers further state that the attitude can be considered in several contexts which include customers' acceptance of the internet as a shopping channel, customer attitudes in preference of a specific internet retailer as well as attitude towards online shopping as a whole. In addition, [11]) indicates that consumer attitude is further influenced by other factors such as (1) perceived risks and (2) service and infrastructure.

A perceived risk refers to the nature and amount of risk perceived by a consumer in contemplating a particular purchase decision. Basically, this refers to all the risks the consumer considers in making the decision to purchase online. Key risks in the online retailing context include:

Financial risk: Risk involved in engaging in a financial transaction through the internet such as risk of being hacked, or being a victim of fraud etc.

Product risk: Risk involved in receiving the undesired product different from the one displayed on the online retailing stores.

Convenience risk: Referred to as the discontent arising from shopping via online retailing stores.

Service and infrastructural variables on the other hand includes all variables related to service and infrastructural deficiencies in the country that affect consumers' attitude towards online retailing.

Subjective norm: Subjective norms reflect a consumer's perceptions influenced significantly by others such as friends, family, authority figures and media [11].

Perceived behavioral control: Another influential factor in online consumer behavior, the perceived behavioral control accounts for consumers' perceptions in their ability to perform a given behavior [11]. Subsequently, this sheds light on the possible internal and external constraints which a consumer faces and which in turn affect their behavior towards online retailing.

Technology specific innovativeness: According to [11], this refers to the degree to which an individual is relatively earlier in adopting an innovation than other members of their system. Basically, this conveys the extent to which an individual goes outside their usual shopping routine and rather actively pursues engaging in new technological innovations.

Essentially, the modified theory of Planned Behavior though used in a different but similar research is a feasible conceptual framework for this study which is aimed at evaluating the customer perceptions towards online retailing in Botswana. According to [13], perception is the first impression that an individual draws and on the basis of it selects and interprets information to form a meaningful picture of the world. Subsequently, as observed by the framework, perceptions play important role in influencing the attitudes as well as other factors crucial to motivating online consumer behavior as well as online shopping rate [14, 15].

2.2 Empirical review

This sub-section explores previous empirical studies carried out on the perceptions of consumers on online retailing in other countries, and in particular, focusing on the factors identified in the conceptual framework(attitude, subjective norms, perceived behavioral control, technology specific innovativeness) and any other factors identified in relevant literature.

2.2.1 Attitude

Attitude towards a behavior refers to the degree to which a person has a favorable or unfavorable evaluation of the behavior in question[16]. In the context of online shopping, [17]consider attitudes to be the consumers' positive or negative feelings related to accomplishing the purchasing behavior on the internet. A study by [18] highlights that an individual's shopping choice is influenced by four key factors which they identify to be motivation, perception, learning and beliefs, and attitudes. Attitude thus serves as a bridge between consumers' background characteristics and the consumption that satisfies their needs[19]. [11] in their modified version of the Theory of Planned behavior, convey that attitude can be influenced by the following key perception factors, in particular, (1)perceived risk and (2) Service and Infrastructure variables:

Perceived risk: [20] defines risk as the measure of probability and the weight of undesired consequences.[21] defines risk as the effect of uncertainty on objectives. [22] identifed perceived risk to be the degree to which a person expresses uncertainty about a service or good and particularly the consequence. In the context of online retailing, [11] consider perceived risk as the risk or uncertainty inherent in a decision to purchase from online retailers. Perceived risk is considered to directly affect online consumer behavior and their intention to purchase [23-26]. Thus, in doing so, the researchers convey that there is a negative relationship or correlation between consumer perceived risk and consumer intention to purchase online. This indicates that when consumer perceived risk is high, the consumer intention to purchase online is low [23]. As noted in [27]," risk is a word that has various meanings to various people". Therefore, due to the differing viewpoints of perceived risk, there is need to contextualize the risk to reflect the key factors being assessed. Possibly in view of this and to best aid the purpose of their study, [11] broke down perceived risk into 3 key subdimensions which are (1)finacial risk, (2) product risk and (3) convenience risk as reviewed below:

Financial Risk: defined as the likelihood of suffering a monetary loss from a purchase [28, 29]. [30] note that there are different reasons why online shoppers may suffer financial loss such as price and security.

Price can be defined as the consumer's perceptual representation or subjective perception of the objective price of the product[31]. The promise of greater savings is one of the major motives drawing consumers to shop online[32]. According to [33] price is a critical factor in customers using online shopping. Online shoppers find it difficult to determine whether the price offered by online retailing sites is indeed the lowest at that point in time [30]. [34] further add that the perception of such financial risk explains why online shoppers abandon carts. Additional costs associated with shopping online such as shipping and delivery costs had been also identifed to deter consumers from shopping online[35].

On the other hand, [36] highlights that security is another critical factor for any success to occur in online retailing. The statement was further cemented by [37][37][37]who state that without proper security, online shoppers would lack significant confidence in using online shopping which would ultimately affect these online retailing stores. [30] highlights that the main security concern and cause of financial loss highlighted was credit card fraud. [38] shares this view as they highlight that the primary reason indicated by most shoppers who preferred not to shop online was due to the fear of revealing personal credit card information to retailers over the internet. The view is further emphasized by [39], whose findings indicate that close to 70% of the US internet shopping population limit their online purchases because of concerns related to the privacy and safety of their personal information. Despite privacy concerns being frequently cited for not purchasing online, it did not have any significant influence on purchasing behaviors suggesting that privacy concerns may not deter shopping among current online shoppers[40] .

Further, findings from previous research indicate that perceived financial risk is the most consistent predictor of internet patronage behavior [40].[40] further indicate that perceived financial risk was a significant predictor for searching with intent to buy, amount spent on the Web and frequency of purchasing online. Thus, this further conveys that financial risk is likely to deter Internet shoppers from initiating the Internet shopping process or cause them to be more selective regarding which online retailing sites they patronize as well as further reducing the amount they could have spent had they not been concerned with this financial risk [40] .

Product Risk also known or considered as performance risk, is defined as the probability of the item failing to meet the performance requirements originally intended[41]. [30]indicates that product risk has been reported as the most frequently cited reason for not shopping online. Furthermore, product risk has been identifed to have significant impact on the frequency of purchasing online[40].[40] note that product performance risk was frequently cited as the reason for not purchasing online, however the impact of that risk was limited to only the frequency of purchasing online. The researchers further indicate that the limited effect of product risk on shopping behaviors suggests that concerns regarding product performance may be less of a deterrent to internet patronage behavior among current Internet shoppers than previously thought [40].

Convenience risk: A distinct characteristic of online shopping is its convenience and it has been found to be the major motive for consumers to shop online [42, 43].[23] defines convenience risk as the risk associated with how the consumer perceives the delay in delivery time, the quality delivered, the after sale services and the relationship with the online vendors. However, out of the noted key points, the risk is mostly focused on the quality provided through the online service [23]. Convenience risk was noted to be a significant predictor of the frequency of searching with intent to buy and frequency of purchasing online but not the amount spent on the web [40]. This suggests some Internet shoppers may hesitate to shop online due to concerns about inconvenience or delays in receiving merchandise [40].

Service and infrastructure variables: This refers to the additional challenges facing e-commerce diffusion particularly in developing countries such as internet connectivity [11]. According to [11], developing countries such as India face challenges such as lack of telecomunications infrastructure throughout the country which ultimately affects use and support of online retailing stores. Further, there are Delivery Concerns which encompasses all the concerns associated with delivery of the product ordered, such as shipping fees, delayed delivery and/or not receiving a product ordered [11]. The researchers further highlight that in India this has proven to be a significant concern as postal couriers are unreliable and relatively risky forcing online shoppers to choose the government postal courier which is more reliable. However, this result in increased costs as it is more expensive. In addition to that, there are issues with Return Policy. [36] in their study note that online retailers should have refund policies to convince online consumers that they easily return products and get refunds if they are dissatisfied or exchange products for free within a reasonable timeframe. However, the ease of return policies is frequently cited as a concern to online shoppers[44].

2.2.2 Subjective norms

[23] note that consumer buying decisions are highly affected by the opinions and decisions of others. As was further cited by [45], word-of-mouth marketing is a fundamental part of the marketing process as consumers greatly rely on advice of other people when they make potential purchase decisions.

2.2.3 Perceived behavioral control

This refers to the additional construct added to extend the Theory of Reasoned Action (TRA) into the Theory of Planned Behavior as a determinant of behavioral intention and behavior [11]. Perceived behavioral control refers to individuals' perceptions of their ability to perform a given behavior[46]. Such perceptions can influence the behavior directly or indirectly through behavioral intentions[47]. [48] note that consumer willingness and preference for adopting online shopping was positively related to income, household size and innovativeness. This correlates with the view by [11] that perceived behavioral control reflects perceptions of internal constraints (self-efficacy) as well as external constraints on behavior like availability of resources. In addition, the researchers indicate that unless control over a behavior exists, intentions will not be a sufficient as the predictor of the behavior[47].[49] highlights that Perceived Behavioral Control directly affects online shopping behavior.

2.2.4 Technology specific innovativeness (TSI)

[48] posits that consumer willingness to and preference for adopting online shopping was positively related to innovativeness amongst other factors. [11] further highlights that Technology specific innovativeness conveys the extent to which an individual goes outside their usual shopping routine and rather actively pursues engaging in new technological innovations. However in order to use online shopping, individuals require necessary computer skills and thus those uncomfortable with the use of a computer will likely find shopping at a traditional store easier and faster than through online retailing sites[50].

2.2.5 Other perceptual dimensions Satisfaction:

[51] found consumer satisfaction as essential in order to gain better financial performance of services in a company which is considered the second most important thing to such companies, next to gaining profit.

Demographics:

[52] established that motivational factors as well as age and gender impacted the likelihood of online purchasing.

Gender is an important factor in explaining several differences in consumers' shopping behaviors and perception of goods[53]. [52] highlight that older males were found to have the highest online purchasing behavior. This might be explained further by [40] who found that women perceived more financial risks associated with online shopping than men. Women also portrayed more privacy concerns and perceived risks, all which may be viewed as reasons explaining the unfavorable attitude by women towards online shopping [54, 55]. However, [56] concluded that gender had no significant influence on shopping behavior.

In terms of age, [52] highlights that older internet users were more likely to buy online as compared to younger users. This might be better explained by [57] who notes that internet shoppers tend to be older and have high levels of income. This is so because younger users often do not have credit cards and thus are unable to purchase online[58]. Further, older generations were found to comparatively be more likely to purchase because they spend less time by searching for fewer products [59]. However, [59] further highlighted that younger internet users spend more time searching for many products as much as they purchased almost the same quantity of products as older generations. Meanwhile, [40] note that age was found to be a predictor of amount spent online, however did not predict other shopping behaviors.

Experience is viewed as a complex and developing structure as there are no two exactly similar experiences, but there are various experiences perceived differently[60]. Experience in this case refers to undergoing or encountering online shopping. Intention to shop online is related to past purchasing experience and directly influences internet shopping behavior[50]. Such past online shopping experiences have been identified to have significant influence on purchasing intention [61, 62]. However, [63] conducted a study in Indonesia, the direct effect of experience on purchasing intention was found to be insignificant. However, [40] note that despite the potential impact of such demographical characteristics, they are found to have little consistent impact on online shopping behaviors.

Based on the reviewed literature, the following seven hypotheses were formulated;

H1a: There is a negative relationship between financial risk and attitude towards online retailing

H1b: There is a negative relationship between product risk and attitude towards online retailing

H1c: There is a negative relationship between convenience risk and attitude towards online retailing

H2: The fear of product delivery will negatively affect attitudes towards online shopping

H3: A good return policy will positively affect attitudes towards online shopping

H4: There is a positive relationship between technology specific innovativeness (TSI) and online retailing

H5: There is a positive relationship between subjective norms (opinion and decision of significant others such as relatives, friends, peers) and online retailing

H6: There is direct relationship between attitude and online retailing

H7a: Older generations are perceived to use more online retailing than younger ones

H7b: Males are perceived to use more online retailing than female counterparts

3 Methodology

The study adopted a positivism philosophy as credible data could only be derived through quantitative analysis of phenomena observed [64]. Due to the positivist nature of the research, a deductive approach was adopted. This approach represents the most common view of the relationship between theory and research and results received from this approach are developed through logical reasoning [65]. A survey was conducted using a questionnaire to collect data on consumer perceptions towards online retailing.

The target population for the study was the University of Botswana community. Thus a sample was taken comprising of university students and staff, academic and non-academic staff (administration staff). Justification for University of Botswana community was that:

i. Internet Access: The University of Botswana provided year-round unlimited internet and Wi-Fi connectivity to staff and students at the campus.

ii. Computer literate: The University of Botswana staff is highly qualified and proficient in the use of computers. Students during the first academic year must take an Information and Communication Technology course that provide sufficient basic understanding of how to use the computer and internet.

iii. Financial Resources: The salaries and wages for employees are credited into their bank accounts thus meaning that such employees all had either a debit or credit card. Students on the other hand, had debit cards made possible by the government sponsorship agreement which resulted in debit cards being issued to students in order to access their monthly allowance stipend.

Based on the gender proportions of the internet users population in Botswana which stood at 240 871 males and 252 913 females as at 31 December 2014 [8], a quota and convenient sampling technique was applied to determine the study sample. Using 0.01% proportion from each quota of the population the sample size of the study was determined. Effectively, meaning that 24 males and 26 females were selected to make up the study sample.

The time horizon for the study was a cross sectional one, as such this study was conducted once off and represented the consumer perceptions towards online retailing prevalent during the period April-May 2017.

The study used primary data that was collected using a questionnaire that had two sections, the first section being for the background information of respondents and second section having questions of their perceptions towards online retailing. They were closed ended questions based on Likert scale and respondents were asked to state their opinion or preference for in response to the particular question on a scale given.

IBM SPSS Statistics 24 software was used to analyse the data from the questionnaire through descriptive statistics, frequency tabulations, independent t-tests and as well as cross-tabulations.

4 Empirical Findings and Discussion

Background information of the respondents is shown on Table 1. A total of forty eight (48) questionnaires were returned by the respondents which was 96% response rate. In terms of respondents occupation 24 (50%) were students whilst the other 24 (50%) were university employees and in terms of their gender 25 (52.1%) were male and 23 (47.9%) were female. 52.08% of respondents fell within the 21-25 age range followed by the 32 and above age group that had 39.58% of respondents whilst 26-32 years and, 20 and under years both had 4.17% of respondents. The majority of the respondents, 52.08% earned a monthly income in the range P1, 000 – P10, 000 followed by 41.67% of respondents that earned more than P10, 000 per month. Only 6.25% of the respondents earned a monthly income less than P1, 000.

Table 1 Background Information of Respondents

	Number	Percentage
Gender		
Male	25	52.10
Female	23	47.90
Total	48	100.00
Age		
20 and under	2	4.17
21-25	25	52.08
26-32	2	4.17
32 and above	19	39.58
Total	48	100.00
Occupation		
Student	24	50.00
Employed	24	50.00
Total	48	100.00
Monthly income		
Less than P1000	3	6.25
P1000-P10000	25	52.08
More than P10000	20	41.67
Total	48	100.00

Perceptions of respondents regarding online retailing

The perceptions of respondents regarding online retailing are depicted in Table 2. In particular, respondents were asked to state their degree of agreement of the identified statements regarding their perceptions. Out of the twenty six opinions evaluated from the respondents only two had means below the expected mean of 3 and these were “online shopping is for individuals with high income” and “online shopping is as secure as traditional shopping” with means of 2.75 and 2.88 respectively.

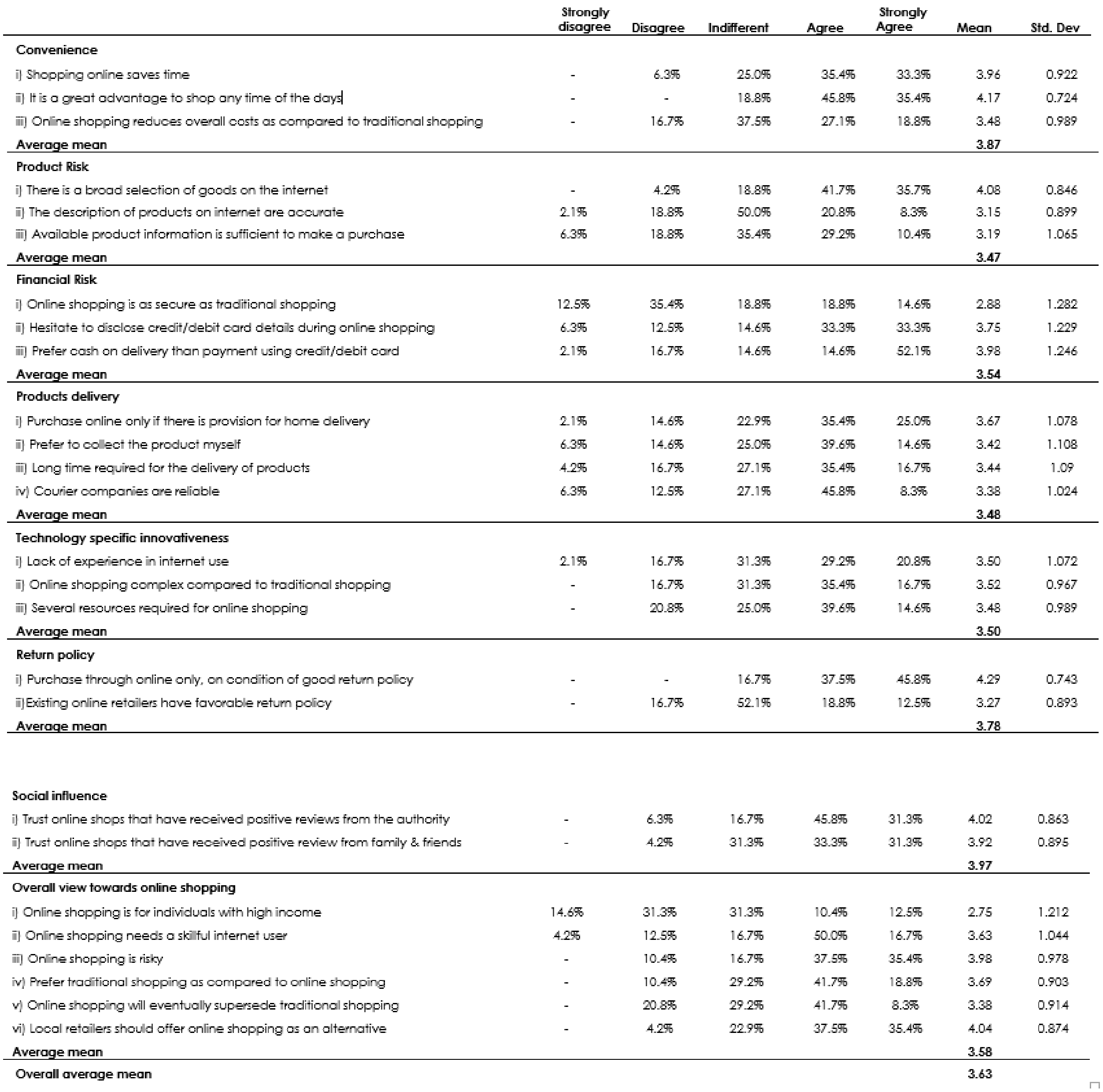
The overall mean of the respondents’ perceptions was 3.63 and greater than the expected mean of 3. Notably, “Purchasing through online only, on condition of good return policy” had the highest mean of 4.29 followed by “It is a great advantage to shop anytime of the day” with a mean of 4.17. The other perceptions with means above 4 were “There is a broad selection of goods on the internet”, “Local retailers should offer online shopping as an alternative” and “Trust online shops that have received positive reviews from the authority”.

In order to generalize the perception of respondents, the twenty six perceptions were sub-divided into eight groups being convenience, product risk, financial risk, product delivery, technology specific innovativeness, return policy, social influence and overall view (attitude) towards online shopping. Social influence had the highest average mean of 3.97 in which respondents were agreeing that they trusted online shops with positive reviews from authorities, friends and family.

The three perceptions that were grouped together to represent “Financial risk” had an average mean of 3.54 above the expected average mean of 3. The respondents were not in agreement that online shopping was secure as compared traditional shopping with a mean of 2.88. They preferred to pay cash on delivery as compared to using credit/debit cards when purchasing online purchases with a mean of 3.98. The respondents hesitated to disclose details of their debit/credit cards when making online purchases as confirmed by the mean of 3.75. This shows that the respondents were afraid of the financial risk they could face when purchasing online. Therefore hypothesis H_{1a}: is accepted which states that "There is a negative relationship between financial risk and attitude towards online retailing".

Respondents' perception towards product risk was determined by enquiring on the selection of goods on the internet, accuracy of products description and whether sufficient information was available to make a purchase decision. The average mean of product risk was 3.47 entailing that respondents agreed there was a broad selection of products with accurate descriptions and sufficient information to make a buy decision on the websites. Hypothesis H_{1b}: is rejected which states that, "There is a negative relationship between product risk and attitude towards online retailing".

Table 2: Perceptions of respondents regarding online retailing



	Strongly disagree	Disagree	Indifferent	Agree	Strongly Agree	Mean	Std. Dev
Convenience							
i) Shopping online saves time	-	6.3%	25.0%	35.4%	33.3%	3.96	0.922
ii) It is a great advantage to shop any time of the days	-	-	18.8%	45.8%	35.4%	4.17	0.724
iii) Online shopping reduces overall costs as compared to traditional shopping	-	16.7%	37.5%	27.1%	18.8%	3.48	0.989
Average mean						**3.87**	
Product Risk							
i) There is a broad selection of goods on the internet	-	4.2%	18.8%	41.7%	35.7%	4.08	0.846
ii) The description of products on internet are accurate	2.1%	18.8%	50.0%	20.8%	8.3%	3.15	0.899
iii) Available product information is sufficient to make a purchase	6.3%	18.8%	35.4%	29.2%	10.4%	3.19	1.065
Average mean						**3.47**	
Financial Risk							
i) Online shopping is as secure as traditional shopping	12.5%	35.4%	18.8%	18.8%	14.6%	2.88	1.282
ii) Hesitate to disclose credit/debit card details during online shopping	6.3%	12.5%	14.6%	33.3%	33.3%	3.75	1.229
iii) Prefer cash on delivery than payment using credit/debit card	2.1%	16.7%	14.6%	14.6%	52.1%	3.98	1.246
Average mean						**3.54**	
Products delivery							
i) Purchase online only if there is provision for home delivery	2.1%	14.6%	22.9%	35.4%	25.0%	3.67	1.078
ii) Prefer to collect the product myself	6.3%	14.6%	25.0%	39.6%	14.6%	3.42	1.108
iii) Long time required for the delivery of products	4.2%	16.7%	27.1%	35.4%	16.7%	3.44	1.09
iv) Courier companies are reliable	6.3%	12.5%	27.1%	45.8%	8.3%	3.38	1.024
Average mean						**3.48**	
Technology specific innovativeness							
i) Lack of experience in internet use	2.1%	16.7%	31.3%	29.2%	20.8%	3.50	1.072
ii) Online shopping complex compared to traditional shopping	-	16.7%	31.3%	35.4%	16.7%	3.52	0.967
iii) Several resources required for online shopping	-	20.8%	25.0%	39.6%	14.6%	3.48	0.989
Average mean						**3.50**	
Return policy							
i) Purchase through online only, on condition of good return policy	-	-	16.7%	37.5%	45.8%	4.29	0.743
ii)Existing online retailers have favorable return policy	-	16.7%	52.1%	18.8%	12.5%	3.27	0.893
Average mean						**3.78**	
Social influence							
i) Trust online shops that have received positive reviews from the authority	-	6.3%	16.7%	45.8%	31.3%	4.02	0.863
ii) Trust online shops that have received positive review from family & friends	-	4.2%	31.3%	33.3%	31.3%	3.92	0.895
Average mean						**3.97**	
Overall view towards online shopping							
i) Online shopping is for individuals with high income	14.6%	31.3%	31.3%	10.4%	12.5%	2.75	1.212
ii) Online shopping needs a skillful internet user	4.2%	12.5%	16.7%	50.0%	16.7%	3.63	1.044
iii) Online shopping is risky	-	10.4%	16.7%	37.5%	35.4%	3.98	0.978
iv) Prefer traditional shopping as compared to online shopping	-	10.4%	29.2%	41.7%	18.8%	3.69	0.903
v) Online shopping will eventually supersede traditional shopping	-	20.8%	29.2%	41.7%	8.3%	3.38	0.914
vi) Local retailers should offer online shopping as an alternative	-	4.2%	22.9%	37.5%	35.4%	4.04	0.874
Average mean						**3.58**	
Overall average mean						**3.63**	

The perceptions of respondents regarding convenience brought up as result of online purchasing was determined based on three opinions, “Shopping online saves time”, “It is a great advantage to shop anytime of the day” and “Online shopping reduces overall costs as compared to traditional shopping” with means of 3.96, 4.17 and 4.48 respectively. The average mean of convenience was 3.87 implying that online shopping enabled the respondents to shop anytime of the day, spent less time shopping and reduced overall cost of shopping. Hypothesis H1c: is rejected which states that “There is a negative relationship between convenience risk and attitude towards online retailing”. Respondents were of the opinion that online shopping will make shopping less time consuming and convenient (shop anytime of the day).

The four opinions considered regarding product delivery had an average mean of 3.48. The respondents expressed their feelings on the following issues “Purchase online only, if there is provision for home delivery (mean=3.67)”,” Prefer to collect the product myself (mean=3.42)”, “Long time is required for the delivery of products (mean=3.44)” and “Courier companies are reliable (mean=3.38)”. The respondents were concerned about the delivery of the products they purchase online. Hypothesis H2: is accepted, “The fear of product delivery will negatively affect attitudes towards online shopping”.

Respondents expressed their opinions regarding online shopping return policy based on two factors “Purchase through online only, on condition of good return policy (mean=4.29)” and “Existing online retailers have favourable return policy (mean=3.27)”. The average mean for return policy was 3.78 and this highlight the fact that respondents will shop online on condition of good return policy. Therefore Hypothesis H3: is accepted, “A good return policy will positively affect attitudes towards online shopping”.

The view of respondents on Technology specific innovativeness was determined based on three issues, “Lack of internet use experience (mean=3.50), “Online shopping complex as compared to traditional shopping (mean=3.52)” and “Several resources required for online shopping (mean=3.48)”. The average mean of technology specific innovativeness was 3.50 implying that the respondents were in agreement it was complicated to purchase online, internet experience was a necessity and there was need of several resources for online shopping. Hypothesis H4: is rejected, “There is a negative relationship between technology specific innovativeness (TSI) and online retailing”.

Social influence effect on purchasing online was determined based on two issues, “Trust online shops that have received positive reviews from the authority (mean=4.02)” and “Trust online shops that have received positive review from family and friends (mean=3.92)”. 3.97 was the average mean of Social influence implying that respondents were influenced positively by good reviews from authorities, friends and family regarding online shops. Hypothesis H5: is accepted, “There is a positive relationship between subjective norms (opinion and decision of significant others such as relatives, friends, peers) and online retailing”.

Perceptions regarding the overall view of respondents towards online shopping individually had a mean above the expected mean of 3.00 except for “Online shopping is for individuals with high income” that had a mean of 2.75. The respondents believed that anyone can make use of online shopping and an individual’s social standing in the society was not a deterrent. The majority of the respondents were of the view that online shopping was risky (mean = 3.98), however, it was supposed to be offered as shopping alternative to customers (mean = 4.04) and required experienced internet users (mean = 3.63). Nonetheless, the respondents preferred to use traditional shopping as compared online shopping (mean=3.69) although they were of the opinion that at some point online shopping will supersede traditional shopping (mean = 3.38). The overall view of respondents represented their attitude toward online shopping and based on the average mean of 3.58, respondents were of the opinion that although online retailing may seem risky and not preferred method of shopping at the moment gradually everyone will have to make use of it. Therefore Hypothesis H6: is accepted, "There is direct relationship between attitude and online retailing".

Comparison of students and employees perceptions toward online shopping

Table 3 shows a comparison of students and employees perceptions toward online shopping. Students had an average mean of 3.65 as compared to employees' average mean of 3.58 implying that students were more influenced by the perceptions considered in making a decision to purchase online. When an independent t-test was conducted for the means of the twenty six opinions considered for the students and employees, only one opinion "Local retailers should offer online shopping as an alternative" had statistically significant results with a t= 2.429, p < 0.05.

There was a significant difference in terms of opinion on whether local retailers should offer online shopping as an alternative. The students were of the opinion that online retailing should be offered as an alternative with a higher mean of 4.33 as compared to the mean of 3.75 for employees.

"Social influences" had the highest mean of 4.15 for students followed by "Return policy" with mean of 3.86. The other perceptions of students that had average means above 3.5 were Convenience (mean=3.73), Overall view towards online shopping (mean=3.72), Financial risk (mean=3.57) and Technology specific innovativeness (mean=3.503). Products delivery and products risk had average means of 3.49 and 3.36 respectively above the expected average mean of 3.

For employees "Convenience" had the highest mean of 4 followed by "Social influence" with average mean of 3.80. The other perceptions that had higher average means for employees were Return policy (mean=3.71), Products risk (mean=3.58), Financial risk (mean=3.50) and Technology specific innovativeness (mean=3.50). Products delivery and Overall view (attitude) toward online shopping had average means of 3.46 and 3.43 respectively above the expected average mean of 3.

Table 3: Comparison of students and employees perceptions towards online shopping

	Students		Employees		t-test for equality of means	
	Mean	Std. Dev	Mean	Std. Dev	t	Significance
Convenience						
i) Shopping online saves time	3.83	0.87	4.08	0.97	(0.939)	0.353
ii) It is a great advantage to shop any time of the days	4.08	0.65	4.25	0.79	(0.794)	0.431
iii) Online shopping reduces overall costs as compared to traditional shopping	3.29	1.04	3.67	0.92	(1.324)	0.192
Average mean	**3.73**		**4.00**			
Products Risk						
i) There is a broad selection of goods on the internet	4.04	0.96	4.13	0.74	(0.338)	0.737
ii) The description of products on internet are accurate	3.00	0.98	3.29	0.81	(1.127)	0.266
iii) Available product information is sufficient to make a purchase	3.04	1.08	3.33	1.05	(0.948)	0.348
Average mean	**3.36**		**3.58**			
Financial Risk						
i) Online shopping is as secure as traditional shopping	2.63	1.17	3.13	1.36	(1.363)	0.179
ii) Hesitate to disclose credit/debit card details during online shopping	3.83	1.20	3.67	1.27	0.466	0.644
iii) Prefer cash on delivery than payment using credit/debit card	4.25	1.19	3.71	1.27	1.527	0.134
Average mean	**3.57**		**3.50**			
Products delivery						
i) Purchase online only if there is provision for home delivery	3.92	0.88	3.42	1.21	1.634	0.109
ii) Prefer to collect the product myself	3.17	1.13	3.67	1.05	(1.589)	0.119
iii) Long time required for the delivery of products	3.63	0.97	3.25	1.19	1.198	0.237
iv) Courier companies are reliable	3.25	1.11	3.50	0.93	(0.843)	0.403
Average mean	**3.49**		**3.46**			
Technology specific innovativeness						
i) Lack of internet use experience	3.460	1.179	3.540	0.977	(0.267)	0.791
ii) Online shopping complex compare to traditional shopping	3.630	1.013	3.420	0.929	0.742	0.462
iii) Several resources required for online shopping	3.420	1.100	3.540	0.884	(0.434)	0.666
Average mean	**3.503**		**3.500**			
Return policy						
i) Purchase through online only, on condition of good return policy	4.46	0.66	4.13	0.80	1.580	0.121
ii)Existing online retailers have favorable return policy	3.25	0.79	3.29	1.00	(0.160)	0.874
Average mean	**3.86**		**3.71**			
Social influence						
i) Trust online shops that have received positive reviews from the authority	4.13	0.95	3.92	0.78	0.834	0.409
ii) Trust online shops that have received positive review from family & friends	4.17	0.96	3.67	0.76	1.995	0.052
Average mean	**4.15**		**3.80**			
Overall view towards online shopping						
i) Online shopping is for individuals with high income	2.79	1.10	2.71	1.33	0.236	0.815
ii) Online shopping needs a skillful internet user	3.83	1.05	3.42	1.02	1.396	0.169
iii) Online shopping is risky	4.17	0.82	3.79	1.10	1.339	0.188
iv) Prefer traditional shopping as compared to online shopping	3.83	0.92	3.54	0.88	1.122	0.268
v) Online shopping will eventually supersede traditional shopping	3.38	1.01	3.38	0.82	0	1.000
vi) Local retailers should offer online shopping as an alternative	4.33	0.76	3.75	0.90	2.429	0.019
Average mean	**3.72**		**3.43**			
Overall average mean	**3.65**		**3.58**		**0.245**	**0.384**

The independent t-test results for the twenty six perceptions considered for the students and employees were insignificant ($p > 0.05$) except for one perception “Local retailers should offer online shopping as an alternative”. The average t-test score for all the perceptions was $t = 0.245$, $p > 0.05$. This implies that there was an insignificant difference amongst the means of the perceptions influencing students and employees to purchase online. Therefore hypothesis H7a: is rejected which states that “Older generations (staff) are perceived to use more online retailing than younger ones (students)" because employees and students opinions regarding the use of online purchasing were more or less the same.

Comparison of males and females perceptions toward online shopping

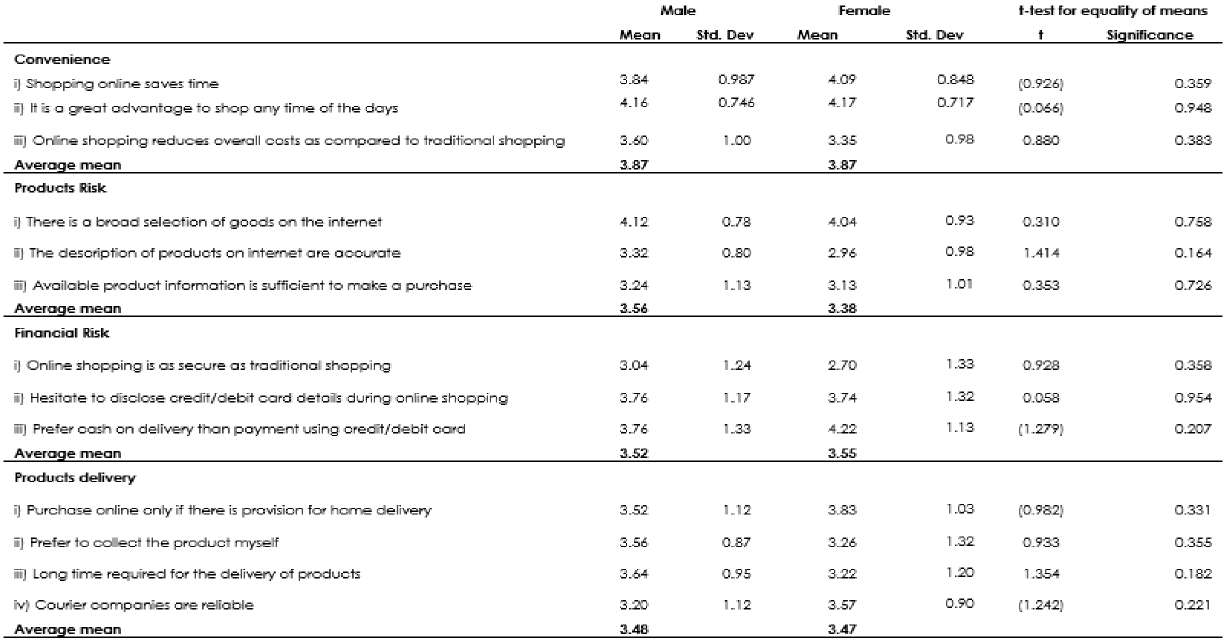
Table 4 shows a comparison of males and females perceptions toward online shopping. Males had an average mean of 3.64 as compared to females' average mean of 3.59 implying that males were more influenced by the factors considered in making a decision to purchase online as compared to females. When an independent t-test was conducted for the means of the twenty six opinions considered they all had statistically insignificant results.

Convenience had an equal average mean of 3.87 for both groups implying that the perceptions of both males and females were the same regarding the convenience of using online retailing. There were no major differences in terms of average means for "Financial risk", "Return policy" and "Product delivery" implying that their perceptions towards the three factors were almost the same for both groups.

Reasonable differences were noticed for the average means of "Technology specific innovativeness" males had an average mean of 3.31 whilst their female counterparts had 3.71 and "Social influences" in which males had an average mean 4.14 and females an average mean of 3.79.

The independent t-test results for the twenty six perceptions considered for the males and females were insignificant ($p > 0.05$). The average t-test score for all the perceptions was $t = 0.182$, $p > 0.05$. This implies that there was an insignificant difference amongst the means of the perceptions influencing males and females to purchase online. Therefore hypothesis H7b: is rejected which states that "Males are perceived to use more online retailing than female counterparts" because males and females perceptions towards the use of online purchasing were almost the same.

Table 4: Comparison of males and females perceptions toward online shopping



	Male		Female		t-test for equality of means	
	Mean	Std. Dev	Mean	Std. Dev	t	Significance
Convenience						
i) Shopping online saves time	3.84	0.987	4.09	0.848	(0.926)	0.359
ii) It is a great advantage to shop any time of the days	4.16	0.746	4.17	0.717	(0.066)	0.948
iii) Online shopping reduces overall costs as compared to traditional shopping	3.60	1.00	3.35	0.98	0.880	0.383
Average mean	**3.87**		**3.87**			
Products Risk						
i) There is a broad selection of goods on the internet	4.12	0.78	4.04	0.93	0.310	0.758
ii) The description of products on internet are accurate	3.32	0.80	2.96	0.98	1.414	0.164
iii) Available product information is sufficient to make a purchase	3.24	1.13	3.13	1.01	0.353	0.726
Average mean	**3.56**		**3.38**			
Financial Risk						
i) Online shopping is as secure as traditional shopping	3.04	1.24	2.70	1.33	0.928	0.358
ii) Hesitate to disclose credit/debit card details during online shopping	3.76	1.17	3.74	1.32	0.058	0.954
iii) Prefer cash on delivery than payment using credit/debit card	3.76	1.33	4.22	1.13	(1.279)	0.207
Average mean	**3.52**		**3.55**			
Products delivery						
i) Purchase online only if there is provision for home delivery	3.52	1.12	3.83	1.03	(0.982)	0.331
ii) Prefer to collect the product myself	3.56	0.87	3.26	1.32	0.933	0.355
iii) Long time required for the delivery of products	3.64	0.95	3.22	1.20	1.354	0.182
iv) Courier companies are reliable	3.20	1.12	3.57	0.90	(1.242)	0.221
Average mean	**3.48**		**3.47**			

Technology specific innovativeness						
i) Lack of internet use experience	3.32	1.145	3.70	0.974	(1.219)	0.229
ii) Online shopping complex compare to traditional shopping	3.32	1.03	3.74	0.864	(1.520)	0.135
iii) Several resources required for online shopping	3.28	1.021	3.70	0.926	(1.472)	0.148
Average mean	**3.31**		**3.71**			
Return policy						
i) Purchase through online only on condition of good return policy	4.28	0.74	4.30	0.77	(0.112)	0.911
ii)Existing online retailers have favorable return policy	3.40	1.00	3.13	0.76	1.046	0.301
Average mean	**3.84**		**3.72**			
Social influence						
i) Trust online shops that have received positive reviews from the authority	4.16	0.80	3.87	0.92	1.170	0.248
ii) Trust online shops that have received positive review from family & friends	4.12	0.78	3.70	0.97	1.672	0.101
Average mean	**4.14**		**3.79**			
Overall view towards online shopping						
i) Online shopping is for individuals with high income	3.00	1.23	2.48	1.16	1.511	0.138
ii) Online shopping needs a skillful internet user	3.76	0.97	3.48	1.12	0.933	0.356
iii) Online shopping is risky	3.88	1.05	4.09	0.90	(0.729)	0.470
iv) Prefer traditional shopping as compared to online shopping	3.80	0.91	3.57	0.90	0.898	0.374
v) Online shopping will eventually supersede traditional shopping	3.52	0.87	3.22	0.95	1	0.256
vi) Local retailers should offer online shopping as an alternative	4.00	0.96	4.09	0.79	(0.341)	0.735
Average mean	**3.66**		**3.49**			
Overall average mean	**3.64**		**3.59**		**0.182**	**0.398**

Table 5: Summary of accepted/rejected hypothesis.

Hypothesis	Statement	Result
Ho_{1a}:	There is a negative relationship between financial risk and attitude towards online retailing	Accepted
Ho_{1b}:	There is a negative relationship between product risk and attitude towards online retailing	Rejected
Ho_{1c}:	There is a negative relationship between convenience risk and attitude towards online retailing	Rejected
Ho_2:	The fear of product delivery will negatively affect attitudes towards online shopping	Accepted
Ho_3:	A good return policy will positively affect attitudes towards online shopping	Accepted
Ho_4:	There is a positive relationship between technology specific innovativeness (TSI) and online retailing	Rejected
Ho_5:	There is a positive relationship between subjective norms (opinion and decision of significant others such as relatives, friends, peers) and online retailing	Accepted
Ho_6:	There is direct relationship between attitude and online retailing	Accepted
Ho_{7a}:	Older generations (staff) are perceived to use more online retailing than younger ones (students)	Rejected
Ho_{7b}:	Males are perceived to use more online retailing than female counterparts	Rejected

5 Conclusion and Recommendations

The paper set out to establish the factors influencing or deterring consumers to purchase from online retailing stores in Botswana. Specific focus was on identifying and assessing the existing customer perceptions towards online retailing among University of Botswana students and employees. We adopted the modified Theory of Planned Behavior (TPB) particularly used by [11] as our conceptual framework. The model assumes that online shopping behavior is influenced by several factors, the four most common being attitude, subjective norm, perceived behavioral control and technology specific innovativeness.

The expectation was that the way consumers perceived financial risk in terms of online shopping has a direct impact on the decision to partake in online retailing. The results of this study are in line with literature [23] that found that the higher the percieved financial risks the more unlikely it is for consumers to take part in online retailing. When consumers choose to buy products online there is an expectation that the product displayed and received match the given description. Researchers have differents opinions regarding the impact of product risk on consumers decisions to carry out online retailing. However results of this study indicate that users are not deterred by perceived product risk. One of the reasons consumers choose online retailing over traditional methods is the convenience offered by these platforms. On the other hand, some Internet shoppers may hesitate to shop online due to concerns about inconvenience or delays in receiving merchandise [40]. In this particular study the results indicate that consumers believe online retailing is more favourable because of its potential to being more convenient this is in line with the findings of [42, 43].

In developing countries such as Botswana, internet accessibility is not fully established and is often expensive. These concerns as similar to those of consumers in India ultimately affect use and support of online retailing stores. Furthermore, there is the aspect of delivery which encompasses all the concerns associated with delivery of the product ordered, such as shipping fees, delayed delivery and/or not receiving a product ordered [11]. It is no wonder the results of this study indicate that product delivery plays a major role in their decision to shop online. Consumer’s attitude towards online retailing is also influenced by retailers having a good return policy in place to easily allow consumers to return goods if dissatisfied. Not everyone is comfortable or conversant with online retailing hence the results of this study support the view that those uncomfortable with the use of a computer will likely find shopping at a traditional store easier and faster than through online retailing sites[50]. The results also confirm the fact that many consumers rely on reviews or recommendations from other customers which is in line with the findings of [23], in their study they noted that consumer buying decisions are highly affected by the opinions and decisions of others.

In the context of online shopping, [17] consider attitudes to be the consumers’ positive or negative feelings related to accomplishing the purchasing behavior on the internet. The respondents of this study do not have a negative attitude towards online shopping despite the fact many currently use the traditional method of shopping in physical stores. Instead consumers have an appreciation that online retailing will eventually be the preferred method of shopping. There is a difference in opinion among researchers regarding demographic factors such as age and gender on online retailing. Findings by [59] found that older generation were more likely to participate in online retailing while [40] note that age was found to be a predictor of amount spent online; this is not in line with the findings of this study that age plays an insignificant role in consumers choice in participating in online retailing. In another study by [56], concluded that gender had no significant influence on shopping behavior which is in line with the findings of this study that found that gender plays an insignifant role on online retailing.

As part of recommendations we suggest that local retailers provide online retailing as an alternative method of shopping, as there is still opportunity for growth in the Botswana market based on the results of this study. Future research can be carried out beyond the University of Botswana community to assess consumer perceptions towards online retailing at a national level.

Local retailers should aim to fully integrate both the physical and virtual stores in order to give consumers a seamless experience. This includes offering a wide variety of products, providing online and after sales support. Delivery costs and the time period between placing an order and receiving the goods is also an essential component of online retailing that often deters consumers from completing an online purchase [38]. In addition retailers should ensure that they provide a secure platform in order to build trust between themselves and consumers by protecting customer details such as personal credentials and credit card details; this way reducing the perceived risk associated with online retailing.

REFERENCES

[1] Fallows, J., The 50 greatest breakthroughs since the wheel. The Atlantic, 2013. 3.

[2] Macdougald, J.J., Internet Use and Economic Development: Evidence and Policy Implications, 2011, University of South Florida Scholar Commons: South Florida.

[3] Ajeet, K., E-commerce Basics: Advantages and Disadvantages of Ecommerce, 2014.

[4] eMarketer, in eMarketer Inc.2014.

[5] Zaczkiewicz, A., Amazon, Wal-Mart lead top 25 e-commerce retail list. WWD, March, 2016. 7.

[6] Li, C., China Online Payment Market Overview for 2014. China Internet Watch, 2014.

[7] Gensler, L., The world’s largest retailers 2016: Wal-mart dominates but amazon is catching up. Retrieved from Forbes website: http://www. forbes. com/sites/laurengensler/2016/05/27/global-2000-worldslargest-retailers, 2016.

[8] Botswana, s., Botswana Household access and individual use of Information and Communication Technologies- 2014, 2016, Statistics Botswana: Gaborone.

[9] Deloitte. Deloitte Touche Tohmatsu Limited, 2015.

[10] Population & Housing Census 2011, 2015, Statistics Botswana: Gaborone.

[11] Sinha, J. and J. Kim, Factors affecting Indian consumers’ online buying behavior. Innovative Marketing, 2012. 8(2): p. 46-57.

[12] Ajzen, I., The theory of planned behavior. Organizational behavior and human decision processes, 1991. 50(2): p. 179-211.

[13] Munnukka, J., Customers' purchase intentions as a reflection of price perception. Journal of Product & Brand Management, 2008. 17(3): p. 188-196.

[14] Cheung, C.M. and M.K. Lee, An integrative model of consumer trust in internet shopping. ECIS 2003 Proceedings, 2003: p. 48.

[15] Wu, S.-I., The relationship between consumer characteristics and attitude toward online shopping. Marketing intelligence & planning, 2003. 21(1): p. 37-44.

[16] Grandom, E. and P. Mykytyn, Theory-based Instrumentation to Measure the Intention to use Electronic Commerce in Small and Medium Sized Businesses. Journal of Computer Information Systems, 2004: p. 44-57.

[17] Chiu, Y.-B., C.-P. Lin, and L.-L. Tang, Gender differs: assessing a model of online purchase intentions in e-tail service. International journal of service industry management, 2005. 16(5): p. 416-435.

[18] Armstrong, G. and P. Kotler, Marketing–An Introduction, (2000), Pearson Education, Asia.

[19] Delafrooz, N., et al., Factors affecting students' attitude toward online shopping. African Journal of Business Management Vol. 3(5), 2009: p. 200-209.

[20] Lawrence, W.W., Of acceptable risk. William Kaufmann, Los Altos, CA, 1976.

[21] Standardization, I.O.f., ISO 31000: Risk Management: Principles and Guidelines. 2009: ISO.

[22] Bauer, R.A. Consumer behavior as risk taking. in Proceedings of the 43rd National Conference of the American Marketing Assocation, June 15, 16, 17, Chicago, Illinois, 1960. 1960. American Marketing Association.

[23] Arshad, A., et al., The Impact of Perceived Risk on Online Buying Behavior. International Journal of New Technology and Research (IJNTR) ISSN, 2015: p. 2454-4116.

[24] Bhatnagar, A. and S. Ghose, Segmenting consumers based on the benefits and risks of Internet shopping. Journal of Business Research, 2004. 57(12): p. 1352-1360.

[25] Doolin, B., et al., Perceived risk, the Internet shopping experience and online purchasing behavior: A New Zealand perspective. Journal of Global Information Management (JGIM), 2005. 13(2): p. 66-88.

[26] Kuhlmeier, D. and G. Knight, Antecedents to internet-based purchasing: a multinational study. International Marketing Review, 2005. 22(4): p. 460-473.

[27] Adams, J., Managing Risk: framing your problems. BoeringerIngelheim Alumni, 2014.

[28] Jacoby, K. Leon. The Components of Perceived Risk Proceedings. 1972. Third Annual Conference Association for Consumer Research. Chicago: University of Chicago.

[29] Horton, R.L., The structure of perceived risk: Some further progress. Journal of the Academy of Marketing Science, 1976. 4(4): p. 694-706.

[30] Dai, B., S. Forsythe, and W.-S. Kwon, The impact of online shopping experience on risk perceptions and online purchase intentions: does product category matter? Journal of Electronic Commerce Research, 2014. 15(1): p. 13.

[31] Jacoby, J. and J.C. Olson, Consumer response to price: An attitudinal information processing perspective, in Moving ahead in attitudinal research. 1977, American Marketing Association: Chicago.

[32] Chiang, K.-P. and R.R. Dholakia, Factors Driving Consumer Intention to Shop Online: An Empirical Investigation. Journal of Consumer Psychology Vol.13(1&2), 2003: p. 177-183.

[33] Heim, G.R. and K.K. Sinha, Operational drivers of customer loyalty in electronic retailing: An empirical analysis of electronic food retailers. Manufacturing & Service Operations Management, 2001. 3(3): p. 264-271.

[34] Egeln, L.S. and J.A. Joseph, Shopping cart abandonment in online shopping. Atlantic Marketing Journal, 2012. 1(1): p. 1.

[35] Bhatnagar, A., S. Misra, and H.R. Rao, On Risk, Convenience and Internet Shopping Behavior. Communications of the ACM (Association for Computing Machinery) Vol. 43(11), 2000: p. 98-105.

[36] Akbar, S. and P.T. James, Consumers' attitude towards online shopping Factors influencing employees of crazy domains to shop online. Journal of Management and Marketing Research, 2014. 14: p. 1.

[37] Kesh, S., S. Ramanujan, and S. Nerur, A framework for analyzing e-commerce security. Information Management & Computer Security, 2002. 10(4): p. 149-158.

[38] Bhatnagar, A., S. Misra, and H.R. Rao, On risk, convenience, and Internet shopping behavior. Communications of the ACM, 2000. 43(11): p. 98-105.

[39] Chapell, A., Eye on privacy. Target Marketing, 2005. 28(10): p. 27.

[40] Forsythe, S. and B. Shi, Consumer Patronage and Risk Perceptions in Internet Shopping. Journal of Business Research Vol.56, 2003: p. 867-875.

[41] Peter, J.P. and L.X. Tarpey, A Comprehensive Analysis Three Consumer Decision Strategies. Journal of Consumer Research Vol. 2, 1975: p. 29-37.

[42] Delhagen, Retailer revs up, 1997. p. 4.

[43] Jarvenpaa, S.L. and P.A. Todd, Is there a future for retailing on the Internet. Electronic marketing and the consumer, 1997. 1(12): p. 139-154.

[44] Teo and S.H. Thompson, Attitude toward online shopping and the Internet. Behavior and Information Technology Vol. 21(4), 2002: p. 259-271.

[45] Cheema, A. and A.M. Kaikati, The effect of need for uniqueness on word of mouth. Journal of Marketing Research, 2010. 47(3): p. 553-563.

[46] Ajzen, I., The Theory of Planned Behavior. Organizational Behavior and Human Decision Processes, 1991: p. 179-211.

[47] Khatimah, H. and F. Halim, The effect of attitude and its decomposed, perceived behavioral control and its decomposed and awareness on intention to use e-money mobile in Indonesia. Journal of Scientific Research and Development, 2016: p. 39-50.

[48] Sultan, F. and R.B. Henrichs, Consumer preferences for Internet services over time: initial explorations. Journal of consumer marketing, 2000. 17(5): p. 386-402.

[49] George, J.F., The theory of planned behavior and Internet purchasing. Internet research, 2004. 14(3): p. 198-212.

[50] Monsuwe, T.P.y., B.G.C. Dellaert, and K.d. Ruyter, What drives consumers to shop online? A literature review. International Journal of Service Industry Management, 2004.

[51] Heskett, J.L., et al., Putting the service-profit chain to work. Harvard business review, 1994. 72(2): p. 164-174.

[52] Korgaonkar, P.K. and L.D. Wolin, A multivariate analysis of web usage. Journal of advertising research, 1999. 39: p. 53-68.

[53] Wan, Y., M. Nakayama, and N. Sutdiffe, The Impact of age and shopping experiences on the classification of search, experience, and credence goods in online shopping. Journal of Information Systems, 2012: p. 135-148.

[54] Graeff, T.R. and S. Harmon, Collecting and using personal data: consumers' awareness and concerns. Journal of consumer marketing, 2002. 19(4): p. 302-318.

[55] Milne, G.R. and A.J. Rohm, Consumer privacy and name removal across direct marketing channels: Exploring opt-in and opt-out alternatives. Journal of Public Policy & Marketing, 2000. 19(2): p. 238-249.

[56] Stafford, T.F., A. Turan, and M.S. Raisinghani, International and cross-cultural influences on online shopping behavior. Journal of Global Information Technology Management, 2004. 7(2): p. 70-87.

[57] Donthu, N. and A. Garcia, The internet shopper. Journal of advertising research, 1999. 39(3): p. 52-52.

[58] Thompson, B., The selling of the clickeati: Today's kids have a visceral connection to computers, in Now the shopping Gods are starting to connect as well. 1999, Wash Post.

[59] Sorce, P., V. Perotti, and S. Widrick, Attitude and age differences in online buying. International Journal of Retail & Distribution Management, 2005. 33(2): p. 122-132.

[60] Giantari, G.A.K., et al., The role of perceived behavioral control and trust as mediator of experience on online purchasing intentions relationship: A study on youths in Denpasar City (Indonesia). International Journal of Business and Management Invention Vol.2 (1), 2013: p. 30-38.

[61] Kwon, W.-S. and M. Noh, The influence of prior experience and age on mature consumers' perceptions and intentions of internet apparel shopping. Journal of Fashion Marketing and Management: An International Journal, 2010. 14(3): p. 335-349.

[62] Weisberg, J., D. Te'eni, and L. Arman, Past purchase and intention to purchase in e-commerce: The mediation of social presence and trust. Internet research, 2011. 21(1): p. 82-96.

[63] Giantari, I., et al., The role of perceived behavioral control and trust as mediator of experience on online purchasing intentions relationship a study on youths in denpasar city (Indonesia). International Journal of Business and Management Invention, 2013. 2(1): p. 30-38.

[64] Saunders, M., P. Lewis, and A. Thornhill, Research methods for business students. 2007. England: Pearson Education Limited, 2009.

[65] Bryman, A. and E. Bell, Business research methods. 2015: Oxford University Press, USA.

Society for Science and Education
United Kingdom

TNC Transactions on Networks and Communications

Volume 6, No. 5
ISSN: 2054 -7420

Enhanced Intelligent Model for NCD in Wireless Sensor Networks

Manyam Thaile, O.B.V. Ramanaiah
Department of Computer Science & Engineering
JNTUH-College of Engineering, Hyderabad
manyamthaile@gmail.com, obvramanaiah@gmail.com

ABSTRACT

The serious security threat for Wireless Sensor Network comes from compromised nodes. Node Compromise (NC) effects of two types: independent and dependent. In independent type, the NC effect is limited to that node only; whereas in dependent, it will spread to all the nodes across the network. In literature, there is an intelligent model which predicts the spread of node compromise. This model suffers from false positives as it trusts the communication from neighbouring nodes. To address this issue, our Parameter Grouping (PG)†mechanism is used in association with the existing Intelligent Model (IM). This Extended Intelligent Model (E-IM) performs better than the IM. The E-IM is studied through NS-2 based simulation and it’s performance is analyzed.

Keywords: WSN security; Uniform Model;Gradient Model; NCD;

1 Introduction

Wireless Sensor Networks (WSN) are used in different areas. A WSN is a self-configuring network which consists of a large number of sensor nodes and are scattered either in regular or random manner. WSNs measure environmental conditions like temperature, wind, sound, pollution levels, and so on. Due to the unattended nature of a WSN, it becomes vulnerable as an attacker can physically capture nodes to make them compromised.

Usually, the attacks are of two types: outsider and insider attacks. Outsider attacks find no extraordinary access of the deployed sensor network yet wants to harm the network. These are also known as external attacks [1]. The attacker nodes which participate and execute this type of attack are not the part of network but still authorize themselves to harm the network. In insider attacks nodes situated in the network are compromised. These attacks from inside are generated by the network nodes rather than from outside nodes, and they are truly a part of the sensor network. These types of attacks are more dangerous than that of outside attacks as the insider knows sensitive information, and has all types of access rights. Hence, to detect compromise nodes is of paramount importance in WSN security.

Compromise node effects are of two types: Independent and Dependent. In case of Independent type, the compromised node does not effect its neighbours. A lot of research work is available on independent node compromise detection (NCD). In case of dependent type, the compromise node effects its neighbours and then it spreads across the network. Little work exists in literature for dependent NCD.

DOI: 10.14738/tnc.65.5396
Publication Date: 31st October, 2018
URL: http://dx.doi.org/10.14738/tnc.5396

The work in [2] represents an intelligent model for dependent NCD. This model estimates the compromise probability of a node based on its compromised neighbour nodes. This model suffers from the risk of false positives. This model is augmented with our Parameter Grouping (PG) model to mitigate false positivies.

The rest of the paper is organized as follows: The related work is discussed in Section-II. The network and attacker models are explained in Section-III. The Proposed Extended-Intelligent Model is presented in Section-IV, and Section-V concludes the paper.

2 Related Work

In independent NCD, behaviour of a node/zone/network is analyzed with the help of different parameters such as packet arrival rate, packet sending rate, packet arrival time, node energy, and node location [4, 5, 6]. The paper [7] introduced a Reputation-Based trust management scheme in which a Bayesian formulation is used to compute an individual node's trustworthiness. The trustworthiness evaluation frame work proposed in [8] with the help of probability and entropy concepts. The compromised nodes are usually not revoked by these schemes due to the likelihood of false positive reports.

Software attestation is another method to detect Independent NCD. This method checks integrity of software code of the sensor node. The papers [9, 10, 11, 12] present work related to software attestation method. All these schemes require each and every sensor node in that network to be attested, whereas in real time scenario all the nodes may not be compromised. Thereby, benign nodes become part of the attestation unnecessarily. This results in wastage of resources of the benign nodes. The works in papers [13, 3] have combined Reputation-based and Attestation schemes. They follow a two-step procedure:

The first step is Identifying an untrustworthy zone/ node, and the next step is Software attestation of that zone/node.

2.1 Intelligent Model

For dependent NCD, a different approach named Intelligent Model (IM) is reported in [2]. The probability of a node for compromise is estimated based on compromised neighbour nodes. This model is further classified into Uniform and Gradient.

- Uniform: Each node has the equal probability for node compromise irrespective of the node position.
- Gradient: Each node has the different probability for node compromise. The far away nodes from the Base Station (BS) have more probability for node compromise.

An intelligent means use current compromised node probability then estimate future compromise node probability.

1) Intelligent uniform model: In this model, a compromised node will have its impact on its neighbour for compromise. But the position of the node is immaterial. The compromise probability of a node is estimated with the help of all the compromised neighbour nodes of it [2]. The mathematical model can be expressed with the following:

$$NCP_k = 1 - \prod_{i=1}^{N} \prod_{j=1}^{M_i} (1 - \frac{P_i}{NG_{ij} - C_{ij}}) \tag{1}$$

- *NCPk* =Node Compromise Probability of *kth* node positioned at (*x, y*) (position is immaterial)

- N=Maximum number of hops of the node being considered
- *Mi*=Number of compromised nodes in *ith* hop of the node being considered
- *Pi*=probability of compromise of the *ith* node using PG model
- NGij =Total number of 1-hop neighbours of jth compromised node in ith hop
- Cij =Total number of 1-hop COMPROMISED neighbours of jth compromised node in ith hop

2) Intelligent gradient model: This model can be adapted in application environments namely, military where the probability of node compromise will be more if the it is far away from BS. The mathematical model is given by:

$$NCP_{(x,y,k)} = (S_k)(1 - \prod_{i=1}^{N}\prod_{j=1}^{M_i}(1 - \frac{P_i}{NG_{ij} - C_{ij}})) \qquad (2)$$

A node gradient distance is calculated from BS by using the Euclidean Distance. If a node has more than specified distance from the BS then we conclude that the node is far away from the BS. It is defined as:

$$S_k = \begin{cases} 1, if(d(BS,k) > Th) \\ 0, otherwise. \end{cases} \qquad (3)$$

Where d(BS, k) is distance between BS and k nodes. The localization techniques will be used to find out the position of sensor nodes [14,15]. These models generate more positives that are false and not detect false negatives effectively.

3 Network and Attacker Models

In this section, we explained our model of wireless sensor networks and the attacker model under which we observe our models.

3.1 Network Model

We assume a static WSN in which the sensor nodes do not change their positions after deployment. We also assume that all direct communication links between the sensor nodes are bidirectional. We also assume that the Base Station (BS) is a trusted entity. If the BS is compromised, the entire mission of the sensor network can be easily undermined. We assume that every sensor node is able to obtain its location information and identify its placement zone by using an existing secure localization scheme such as [14], [15]. We considered two types of networks: Flat and Hierarchical (Zone). The flat network consists of all sensor nodes have same capabilities. The sensor nodes sense environment features and process them and report to BS through multi-hop communication. The Hierarchical network is divided into number of zones. Each zone has some of the normal nodes and then from that one of the node selected as Zone Head (ZH). ZH collects all information from it's zone members and then send to the BS.

3.2 Attacker Model

An attacker is captured physically sensor nodes with the purpose to steal secret data stored. The attacker may inject malicious code onto the nodes and make them as compromised. An attacker redeploys the nodes back them into the network to launch further attacks such as Routing Attacks (e.g., Gray Hole, Black

Hole, Sybil, Wormhole and Hello Flood), Identity Replication, False Data Injection and Passive Data Gathering. It is necessary to detect compromise nodes are very important.

4 Extended intelligent NCD model

We extended IM with our PG model to reduce false positives and detect false negatives effectively. An E-IM model is further classified into Extended-Intelligent Uniform (EIU) NCD model and Extended-Intelligent Gradient (EIG) NCD model. These two models implemented on flat network. We also extended E-IM model to Hierarchical network (zone).

The IM model estimates compromise probaility for a node and it uses only compromised neighbour nodes information. The compromised neighbour nodes may not give correct report about it's neighbour nodes sometimes.

The E-IM model estimates compromise probability of a node with combination of two things: Based on compromised neighbour nodes and behaviour of the node (PG).

The behaviour of the node is evaluated with the help of five parameters [3]. They are Packet sending rate, Depletion of node energy, Node location, False information, and Non-availability. These parameters generates binary values (0,1). The five parameters generate total of 32 (25 = 32) binary patterns, which are five bit in length. We considered these binary patterns as node compromise probability. The node compromise probabilities are 0, 0.2, 0.4, 0.6, 0.8, 1 as shown in the Table-I.

Table 1: Probabilities of compromise node

Pattern	1^ts	0^ts	Probability
00000	0	5	0
00001	1	4	0.2
00011	2	3	0.4
00111	3	2	0.6
01111	1	4	0.8
11111	5	0	1

4.1 Flat Networks

1) Extended Intelligent Uniform NCD Model: This model accommodate the application environment where the attackers attacking with same probability. The following mathematical model uses to estimate node compromise probability P_k:

$$P_k = \frac{P_k + NCP_k}{2} \tag{4}$$

Where

- P_k= k^{th} node compromise probability
- P_k=current compromise probability of k^{th} node using PG
- NCP_k=compromise probability of k^{th} node with respect to all compromised neighbour nodes.

To describe the EU model clearly, we use Fig. 1 in order to calculate compromise probability of the node ’j’. In Fig. 1, nodes e, d, c, i, o, p and k are 1-hop neighbors of node ’j’; nodes d, c, b, h, n, o, p and j are 1-hop neighbors of node ’i’; nodes i, d, e are compromised nodes. In Fig. 1, for node j, N=4, i,e., node j can reach all the sensors in the network within 4 hops, node j has one 1-hop neighbor node (node i), one 2-hops neighbor node (node l) which have been recently compromised. So that M_1=1, M_2=1. Node i has eight 1-hop neighbors, thus n_{11}=8. Node i has one 1 hop compromised neighbor, i.e., node d, then k_{11}=1. Node l has five 2-hops neighbors (nodes d, p, j, v, and w) and one 2-hops compromised neighbor (node d), consequently n_{21}=5, k_{21}=1. Suppose p_1=0.8 p_2=0.4, and P_j=0.2. We calculate the compromise probability of the node j’s as follows:

$$P_{(x,y,j)} = \frac{0.2+1-(1-\frac{1}{8-1}*0.8)(1-\frac{1}{5-1}*0.4)}{2} = 0.20$$

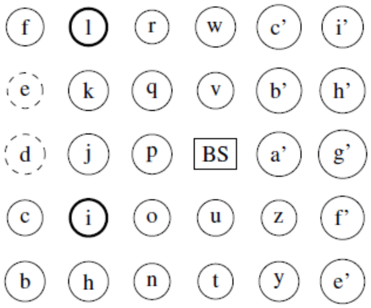


Fig. 1: Extended Intelligent Uniform Model

2) Extended Intelligent Gradient NCD Model: The given mathematical model used to estimate compromise node probability which is far away from BS.

$$P_{(x,y,k)} = \frac{P_k + S_{(x,y,k)} * NCP_{(x,y,k)}}{2} \qquad (5)$$

where $S_{(x,y,k)}$= k^{th} node is far away from Base Station (1 or 0) and P_k is current compromised probability.

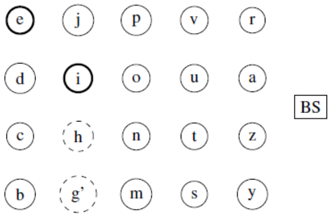


Fig. 2: Extended Intelligent Gradient Model

To describe the EG model clearly, we use Fig. 2 in order to calculate compromise probability of node ’c’. In Fig. 2, compromised nodes will not be recovered after they detected as compromised nodes. In Fig. 2, node i and e are recently compromised nodes that have been compromised in the last time period; Nodes h, g’ are compromised nodes previously. In Fig. 2, for node ’c’, N=4, i,e., node ’c’ can reach all the sensor nodes in the network within 4 hops, node ’c’ has one 1-hop neighbor node (node i), one 2-hops neighbor node (node e), that have been recently compromised. So that M1=1, M2=1, M3=0 and M4=0. Node i has eight 1-hop neighbors, thus n11=8. Node i has two one hop compromised neighbor, i.e., node e and node h, then k11=2. Node e has five 2-hops neighbors (nodes p, o, n, h, and c) and one 2-hops compromised neighbor (node h), consequently n21=5, k21=1. Suppose p1=0.8, p2=0.6, p3=0, p4=0, S(x,y,c)=1 and Pi=0.8. We calculate compromise probability of node c’s as follows:

$$P_{(x,y,c)} = \frac{0.8+1*[1-(1-\frac{1}{8-2}*0.8)(1-\frac{1}{5-1}*0.6)]}{2} = 0.53$$

4.2 Hierarchical Network

We estimated compromise probability for entire network and zone-wise, for which Extended Intelligent Uniform model uses only. The network is divided into ’N’ number of zones in which a node acts as Zone Head (ZH). Every ZH can measure compromise probability of it’s zone. The Base Station (BS) measure compromise probability for the total network. In Fig.3, zh1, zh2, zh3 and zh4 are ZHs, remaining nodes are normal nodes. The dashed circle (node:g) is previously compromised node and thick circle (node:a) is currently compromised node. The compromise probability of total network and zone-wise is tabulated in Table-IV and V respectively.

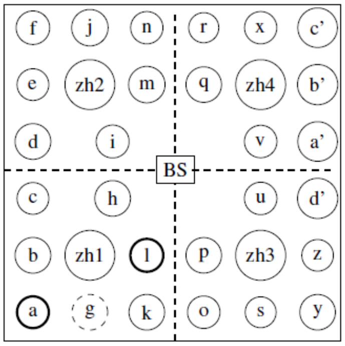


Fig. 3: Zone Network

5 Simulation Study

In this section, we describe our simulation experimental environment and then discuss the simulation results.

5.1 Simulation Environment

We considered NS2 open source simulator for the analysis of the E-IM models. We simulated flat and hierarchical (zones) networks. The network size is scaled from 20, 40, 60, 80, and 100 nodes.

5.2 Simulation Results

The performance of the E-IM models (EU, EG) are evaluated. The simulation results are compared with IM models. We considered following performance metrics.

- False Positive (FP): A trusted node is identified as untrusted.
- False Negative (FN): An untrusted node is identified as trusted.
- False Positive Ratio (FPR): $\frac{FP}{FP+FN}$.
- False Negative Ratio (FNR): $\frac{FN}{FP+FN}$.

Table 2: Flat network

No.of Nodes	IM		E-IM		IM		E-IM	
	False Positives	False Negatives	False Positives	False Negatives	FPR	FNR	FPR	FNR
20	4	1	1	2	0.80	0.20	0.33	0.66
40	8	2	3	4	0.80	0.20	0.42	0.57
60	14	3	7	6	0.82	0.17	0.53	0.46
80	19	4	10	8	0.82	0.17	0.55	0.44
100	26	5	14	9	0.83	0.16	0.59	0.40

The Table-II shows results of false positives, false negatives, false positive ratio, and false negative ratio. We tested three cases for each network of size 20, 40, 60, 80, and 100 nodes. The E-IM model reports less false positives when compare with IM model. An E-IM model is detected false negatives effectively when compare with IM model.

The Fig.4 shows false positive rates between IM and E-IM models. The x-axis indicates the number of nodes, and y-axis indicates false positive rates. An E-IM model got less false positives when compare with IM model.

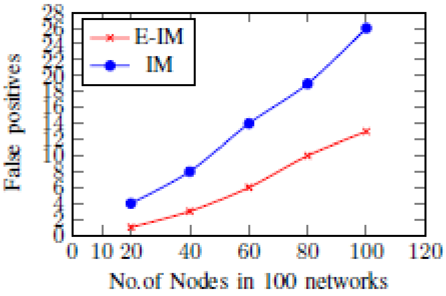


Fig. 4: False positives

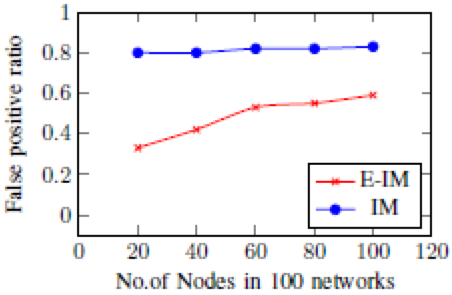


Fig. 5: False positive ratio

The Fig.5 shows the false positive ratio between IM model and E-IM model. The x-axis indicates the number of nodes, and y-axis indicates false positive ratio. The figures shows an E-IM model got less false positive ratio when compare with IM model.

The Fig.6 and 7 show the false negative rates and false negative ratio respectively. Fig.6 represents an E-IM model detected false negatives effectively.

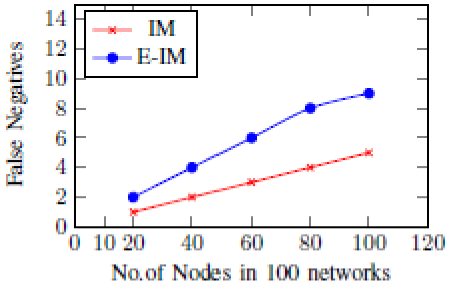


Fig. 6: False Negatives

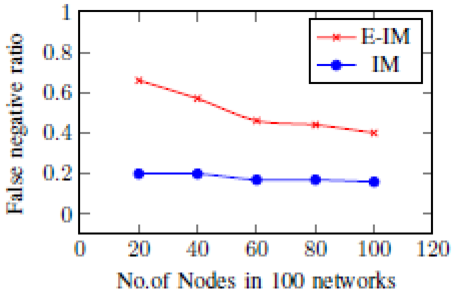


Fig. 7: False negative ratio

We calculated node probabilities for EU and EG. The IM model has given high probabilities but it leads to more false positivies. The E-IM provides less probabilities when compare with IM model and it gives less false positives. The values are shown in Table-III.

Table 3: Node probability

No.of Nodes	Uniform		Gradient	
	IM	E-IM	IM	E-IM
20	0.10	0.05	0.19	0.10
40	0.26	0.12	0.25	0.12
60	0.42	0.31	0.31	0.15
80	0.60	0.36	0.35	0.21
100	0.73	0.52	0.40	0.30

The Fig.8, 9 shows node compromise probability between IM model and E-IM model. An E-IM model gives less compromise probability.

Table-IV indicates the entire network compromise probability. We can predict whether network is going to be compromised or not in the future. We tested three cases for finding network probability by using zones probability. In the table $z_1 = 6 - 1$

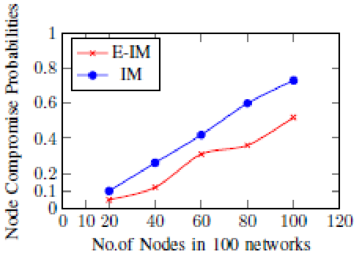


Fig. 8: Uniform model

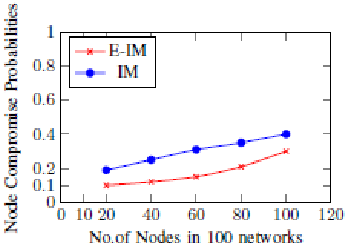


Fig. 9: Gradient model

indicates the difference between total number of nodes (6) in zone z_1 and number of compromised nodes (1) in zone z_1.

Table 4: Network probability by using zones in hierarchical networks

No.of Nodes	No.of Zones	No.of Nodes in each zone and compromised nodes			Probability			Average
		Case1	Case2	Case3	Case1	Case2	Case3	
20	2	z_1=6-1,z_2=10-1	z_1=6-1,z_2=9-1	z_1=6-1,z_2=10-1	0.06	0.06	0.06	0.06
40	4	z_1=9-1,z_2=8-1, z_3=10-2,z_4=7-0	z_1=9-1,z_2=19-2, z_3=2-0,z_4=3-1	z_1=11-2,z_2=6-1, z_3=5-0,z_4=11-1	0.11	0.15	0.12	0.12
60	6	z_1=10-2,z_2=14-2, z_3=9-1,z_4=3-0, z_5=8-1,z_6=4-0	z_1=1-0,z_2=4-0, z_3=5-1,z_4=11-2, z_5=14-2,z_6=5-1	z_1=7-0,z_2=10-2, z_3=3-0,z_4=11-2, z_5=7-1,z_6=10-1	0.15	0.18	0.16	0.16
80	8	z_1=3-0,z_2=8-1, z_3=17-3,z_4=4-0, z_5=6-1,z_6=14-2, z_7=5-0,z_8=6-1	z_1=14-2,z_2=5-1, z_3=17-2,z_4=6-1, z_5=9-1,z_6=8-1, z_7=3-0,z_8=6-0	z_1=6-1,z_2=6-1, z_3=1-0,z_4=9-1, z_5=10-1,z_6=12-2, z_7=11-2,z_8=3-0	0.19	0.21	0.21	0.20
100	10	z_1=11-2,z_2=11-1, z_3=4-0,z_4=3-0, z_5=6-0,z_6=14-2, z_7=10-1,z_8=17-4, z_9=6-0,z_{10}=3-0	z_1=1-0,z_2=9-1, z_3=10-1,z_4=6-1, z_5=4-0,z_6=4-0, z_7=10-1,z_8=5-1, z_9=11-2,z_{10}=16-3	z_1=6-0,z_2=12-2, z_3=9-1,z_4=8-1, z_5=9-1,z_6=4-0, z_7=3-0,z_8=22-4, z_9=11-1,z_{10}=1-0	0.20	0.31	0.19	0.23

Table 5: Zone-Wise probability in hierarchical net- works

No.Of Nodes	No.Of Zones	Zone-Wise Probability
20	2	z_1=0.96, z_2=0.97
40	4	z_1=0.96,z_2=0.96,z_3=0.31, z_4=0.62
60	6	z_1=0.31,z_2=0.63,z_3=0.64, z_4=0.63,z_5=0.96,z_6=0.64
80	8	z_1=0.64,z_2=0.96,z_3=0.64, z_4=0.64,z_5=0.96,z_6=0.94, z_7=0.31, z_8=0.32
100	10	z_1=0.31,z_2=0.97,z_3=0.64, z_4=0.64,z_5=0.32,z_6=0.32, z_7=0.64,z_8=0.94,z_9=0.64, z_{10}=0.31

The Fig.10 shows network compromise probability and it can be observed that when compromised nodes are increased in zones then compromised probability is also increased.

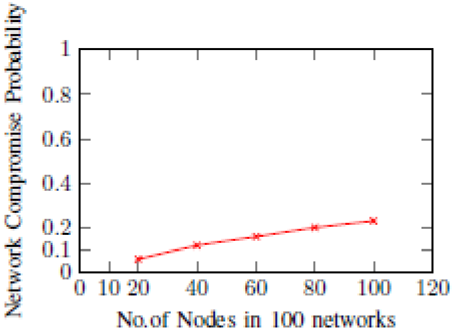


Fig. 10: Network compromise probability

We measured compromise probability zone-wise with the help of number of nodes in a zone and also number of compromised nodes. If any zone consist of compromised nodes then that zone has more compromised probability as shown in Table-V.

6 Conclusion

We implemented an E-IM models to measure compromise probability of a node in flat network. We also extended an E- IM model to hierarchical network and analyzed compromised probability for zone-wise and network wise. The E-IM model reduced false positivies and detected false negatives effectively. We compared performance of E-IM model with IM model and got better performance.

REFERENCES

[1] Kalpana Sharma and M K Ghose, ”Wireless Sensor Networks: An Overview on its Security Threats”, IJCA Special Issue on Mobile Ad-hoc Networks (MANETs), Volume 1, Issue 8, pp.42-45, 2010.

[2] Xiangqian Chen, Kia Makki, Kang Yen, Niki Pissinou, ”Node Compromise Modeling and its Applications in Sensor Networks,” in the proceedings of IEEE ISCC 2007, IEEE Symposium on Computers and Communications, Aveiro, Portugal, July 2007.

[3] Manyam Thaile, and O.B.V Ramanaiah, ”Node Compromise Detection Based on Parameter Grouping in Wireless Sensor Networks,” SECURWARE 2016:The Tenth International Conference on Emerging Security Information, Systems and Technologies, July. 24 – 28, 2016, pp. 14-20, ISBN: 978-1- 61208-493-0, Nice, France.

[4] F. Li and J. Wu, “Mobility Reduces Uncertainty in MANET,” May 2007, Proc. IEEE International Conference on Computer Communications, pp. 1946-1954.

[5] D. Estrin, R. Govindan, J. Heidemann, S. Kumar, ”Next century challenges: scalable coordination in sensor networks,” ACM MobiCom’99, Washingtion, USA, 1999, pp. 263-270.

[6] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, ”A survey on sensor networks,” WEEE communications Magazine, Volume: 40 Issue: 8, pp. 102-114, August 2002.

[7] S. Ganeriwal and M. Srivastava, ”Reputation-Based Framework for High Integrity Sensor Networks,” Oct. 2004, Proc. ACM Workshop Security of Ad Hoc and Sensor Networks (SASN), pp. 66-77.

[8] Y. Sun, Z. Han, W. Yu, and K. Liu, ”A Trust Evaluation Framework in Distributed Networks: Vulnerability Analysis and Defense against Attacks,” Apr. 2006, Proc. IEEE INFOCOM, pp. 1-13.

[9] A. Seshadri, A. Perrig, L. van Doorn, and P. Khosla, ”SWATT:SoftWare-Based Attestation for Embedded Devices,” Proc. IEEE Symp. Security and Privacy, May 2004.

[10] T. Abuhmed, N. Nyamaa, and D. Nyang, ”Software-Based Remote Code Attestation in Wireless Sensor Network,” Proc. of IEEE GLOBECOM, December. 2009.

[11] T. Park and K. G. Shin, ”Soft Tamper-Proofing via Program Integrity Verification in Wireless Sensor Networks,” IEEE Trans.Mobile Computing, May/June 2005, vol. 4, no. 3, pp. 297-309.

[12] Y. Yang, X. Wang, S. Zhu, and G. Cao, ”Distributed Software-Based Attestation for Node Compromise Detection in Sensor Networks,” Proc.IEEE 26th Int’l Symp. Reliable Distributed Systems (SRDS), Oct. 2007.

[13] Jun-Won Ho, Matthew Wright, and Sajal K. Das, ”ZoneTrust: Fast Zone-Based Node Compromise Detection and Revocation in Wireless Sensor Networks Using Sequential Hypothesis Testing,” IEEE Transactions on Dependable and Secure Computing, July/August 2012, vol. 9, no. 4, pp. 494-511.

[14] S. Capkun and J.P. Hubaux, ”Secure Positioning in Wireless Networks,” IEEE J. Selected Areas in Comm., vol. 24, no. 2, pp. 221-232, Feb. 2006. [15] Z. Li, W. Trappe, Y. Zhang, and B. Nath, ”Robust Statistical Methods for Securing Wireless Localization in Sensor Networks,” Proc. Fourth Int’l Symp. Information Processing in Sensor Networks (IPSN), Apr. 2005.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM

TNC TRANSACTIONS ON NETWORKS AND COMMUNICATIONS

VOLUME 6, NO. 5
ISSN: 2054 -7420

Big Data and Machine Learning Driven Open5GMEC for Vehicular Communications

Luong-Vy Le[1], Do Sinh[2], Bao-Shuh Paul Lin[2,3], Li-Ping Tung[3]
[1]*College of Electrical and Computer Engineering, National Chiao Tung University, Hsinchu, Taiwan*
[2]*Department of Computer Science, National Chiao Tung University, Hsinchu, Taiwan*
[3]*Microelectronics & Information Research Center, National Chiao Tung University, Hsinchu, Taiwan*
leluongvy.eed03g@nctu.edu.tw; dosinhuda.cs04g@nctu.edu.tw; bplin@mail.nctu.edu.tw; lptung@nctu.edu.tw

ABSTRACT

Mobile Edge Computing (MEC) is an emerging technology and an essential component of 5G networks to bring cloud services closer to users. That means data collection, storage, processing, computing, communication, and network control are implemented at network edges. MEC is expected to be able to satisfy a variety of delay-sensitive services and applications. On the other hand, the development of vehicles to everything (V2X) communication brings many requirements to future networks to guarantee full intelligence, automatic, and faster computation, management, and optimization to fulfill network QoS (quality of service) and QoE (quality of experience). To deal with those requirements, recently, software-defined networking (SDN), network functions virtualization (NFV), big data, and machine learning (ML) have been proposed as emerging technologies and the necessary tools for MEC and vehicular networks. This study aims to integrate those technologies to build a comprehensive architecture and an experimental framework for future 5G MEC called Open5GMEC. Moreover, the authors analyzed challenges and proposed relevant solutions for future vehicular communications in 5G networks. Finally, based on this framework, we successfully implemented several powerful ML-based applications for V2X such as object detection, network slicing, and migration services, which are executed at Broadband Mobile Lab (BML), National Chiao Tung University (NCTU).

Keywords: 5G, Vehicular communication, V2X; Automotive driving, SDN/NFV; Machine Learning; Big Data; MEC.

1 Introduction

In Mobile Edge Computing (MEC) has recently been proposed as a promising paradigm to overcome the requirements of future networks that enabling a wide range of benefits such as high bit rate, high availability, low latency (less than 1ms), and high mobility in heterogeneously converged connectivity environments by shifting computational efforts from the centralized cloud computing to edge servers. The servers are usually deployed and co-hosted at base stations or near mobile users to eliminate and reduce a tremendous amount of data routing through the core network. As a result, the core network can be simplified, and the end-to-end (E2E) latency is reduced [1][2]. Furthermore, due to the exponential increase of IoT applications and the massive deployment of new vertical business services, MEC is

DOI: 10.14738/tnc.65.5410
Publication Date: 31st October, 2018
URL: http://dx.doi.org/10.14738/tnc.5410

expected to be a flexible and efficient framework, in which network service providers can deploy their applications quickly and efficiently. With particular reference to vehicular communication networks, it is expected to meet various communication requirements of future Intelligent Transportation Systems (ITS) such as low latency, location awareness, and real-time response applications (e.g., driving safety applications and real-time warning on the road). Moreover, the IEEE vehicular communication standard also provided many requirements for vehicular communications like congestion control mechanisms, fairness in accessing resources, and the availability of infotainment services [3]. On the other hand, vehicle-to-everything (V2X) is a promising solution to improve road safety, traffic efficiency, and meet various QoS requirements in different application scenarios [4]. Therefore, recently, 5G-based V2X has been actively conducted by the Third Generation Partnership Project (3GPP) to provide solutions for vehicular communications [5][6]. For example, research [6] explored MEC for 5G-enabled software defined vehicular networks where SDN was exploited and combined with MEC to strengthen vehicular systems. However, MEC for vehicular communications is still in its early stage with many unresolved challenges ranging from reliability, flexibility, scalability architecture to data management and integration, even security issues and so on.

This study, firstly, explores various emerging technologies, such as big data, ML, SDN/NFV, and cloud computing, and then integrates them to propose a comprehensive platform called Open5GMEC for 5G and vehicular networks. SDN/NFV are considered as the most critical technologies for 5G networks to provide the full power of programmability, interoperability, agility, short time to market, and low-cost solution by virtualizing network components and creating multiple logical end-to-end networks [7]. Moreover, recently, ML and big data have been exploited as the key technologies to empower computing components in 5G networks. For example, they make the 5G MEC and SON better integration and more intelligent capabilities [8] [9][10]. Therefore, they are considered as promising solutions deciding the success of 5G-based vehicular communication in term of network reconstruction, virtual-network cooperation, and resource optimization [11]].

The remainder of the paper is organized as follows: Section II reviews V2X communication and the experimental platform based on Big Data, SDN/NFV, and ML; Section III proposes Open5GMEC experimental architecture; Section IV introduces and analyzes ML-based applications for V2X communication; Section V introduces and applies ML and Open5GMEC for V2X applications; Section VI proposes service migration for V2X; finally, Section VII concludes the present study.

2 Overview of V2X Communication and Experimental Network

2.1 Overview V2X communication

V2X is considered as a crucial service of 5G networks to support a variety of ITS applications, each with a specific set of requirements about data rates, latency, mobility, ubiquity, and reliability. A vehicular communication scenario usually consists of a vast number of smart vehicles and roadside units. Generally, an intelligent vehicle integrates many data generators, storage and communication components for real-time sensing and computing (e.g., cameras, radars, and GPS). Roadside units are commonly deployed along the roads to collect and process data sent by smart vehicles. Generally, V2X can be divided into four main categories: Vehicle-to-Vehicle (V2V), Vehicle-to-Pedestrian (V2P), Vehicle-to-Infrastructure (V2I), and Vehicle-to-Network (V2N) communications [4] [12]. They include a massive number of connections among a large number of sensors to offer a wide variety of versatile IoT services for ITS consumers, ranging from accident-free transportation, road safety, parking infrastructure, and greener transport to mobile broadband services like video streaming. They also integrate various communication technologies and protocols. For example, V2V can directly communicate with one another to exchange their information through 5G-based networks or new D2D (device to device) communication interface known as PC5 for sharing safety and critical information; V2N can communicate through heterogeneous networks such as WSNs, non3GPPRANs, 3GPPRANs; V2I communication allows vehicles to communicate with eNodeBs to provide various traffic efficiency and entertainment services.

2.2 Experimental platform

In the past few years, big data, ML, SDN/NFV, and cloud have been integrated into the experimental network testbed for 4G/LTE and beyond 5G, located at MIRC/BML (Microelectronics and Information Research Center/Broadband Mobile Lab) in the campus of National Chiao Tung University to build an open architecture for developing future network applications [7][11][13][14][15]. For example, the integrated architecture of SDN/NFV, cloud, IoT, and big data in 5G and their roles were introduced in [15]; in studies [8][9][16], we explored and implemented many big data and ML algorithms for 5G applications. Most recently, the collaboration between NCTU and Open Networking Lab (ON.Lab) has accomplished an SDN-IP global peering deployment and established primitive primary CORD (Central Office Re-architected as a Datacenter). Besides, we have also focused on developing P4, ONOS, and CORD applications that are considered as the essential solutions of SDN/NFV technologies for 5G.

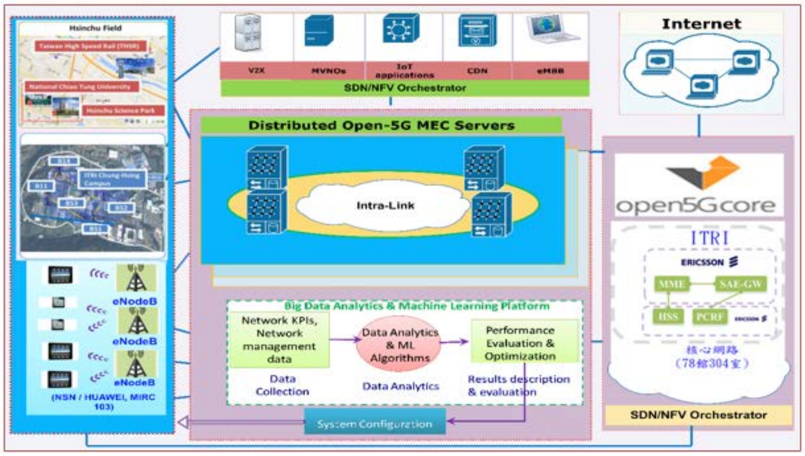


Figure 1. Experimental architecture of 5G network at BML

Fig.1 describes the abstract architecture and physical components of the current developed 5G testbed at BML. In general, it involves four main parts, the RANs, Open5GCore, Open5GMEC, and applications. The RAN of 5G integrates new technologies such as massive MIMO and optical fiber to support high-speed connections for wide-area wireless connectivity to various types access devices, such as 3GPP (e.g., LTE-E-UTRAN), non-3GPP (e.g., WiMAX), Wireless Sensor Network (WSN). The Open5GCore component was described in the study [7]. Its elements such as S-GW, P-GW, home subscriber server (HSS), and mobility management entity (MME) are virtualized and run on commodity data center under the control of SDN/NFV orchestrations. SDN/NFV orchestrations are also used to manage 5G-based services and applications, such as network management as a service, V2X, IoT applications, etc. These applications are considered as virtual entities deployed in Docker containers, and they are created and controlled by SDN/NFV applications. The Open5GMEC component will be introduced in the next section.

3 Open5GMEC Architecture based on SDN/NFV, Big Data, and ML

Fig.1 and Fig.2 illustrate the Open5GMEC in the network architecture, in which big data, ML, cloud computing, and SDN controller applications work on NFV environments, and they are considered as the brain of Open5GMEC. As can be seen in Fig.1, Open5GMEC can interact with both the RANs and the SON to move the computing functions to the proximity of UEs and eNodeBs. Therefore, it is considered as a new intermediate layer responsible for data collection, transformation, filtering, aggregation, and processing and then building both online and offline applications for the SON or RAN, even for third-party applications and services. Furthermore, the integration of those state-of-the-art technologies exhibits as breakthrough approaches and promising solutions that deciding the success of the MEC concept in solving new requirements for 5G and V2X networks regarding network coordination, configuration, management, and optimization to support various new services. For example, to build a V2X application, the Open5GMEC firstly collects necessary data such as real-time GPS to determine the position of the vehicle, traffic on the road, and other necessary data from all sources of the network. Next, it extracts and preprocesses the collected data for applying to big data and ML models. And then, the

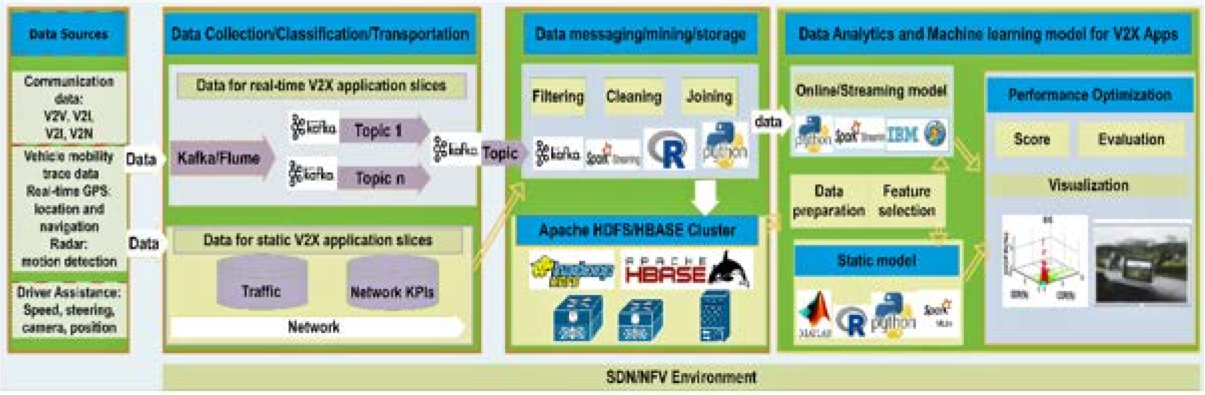

Data Sources
Communication data: V2V, V2I, V2I, V2N
Vehicle mobility trace data
Real-time GPS: location and navigation
Radar: motion detection
Driver Assistance: Speed, steering, camera, position
Data
Data
Data Collection/Classification/Transportation
Data for real-time V2X application slices
Kafka/Flume
Topic 1
Topic n
Data for static V2X application slices
Traffic
Network KPIs
Network
Data messaging/mining/storage
Filtering
Cleaning
Joining
Topic
Apache HDFS/HBASE Cluster
data
Data Analytics and Machine learning model for V2X Apps
Online/Streaming model
Data preparation
Feature selection
Static model
Performance Optimization
Score
Evaluation
Visualization
SDN/NFV Environment

Figure 2. Open5GMEC computing platform for V2X applicationsation intelligently detect which

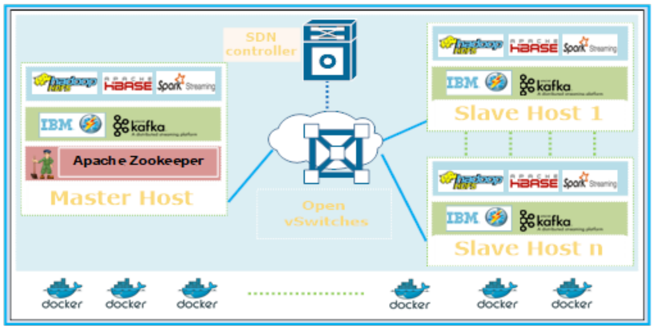


Figure 3. Distributed computing platform

application model is built and sent the result to the SON or RAN. Finally, the network performance is analyzed, evaluated. The following subsection introduces several essential characteristics of the Open5GMEC platform.

Open platform: this is a crucial requirement in deciding the success of Open5GMEC-based ecosystems for IoT and V2X communications. As described in Fig.3, the platform is opened for different state-of-the-art technologies and software can be easily integrated into it for data processing and ML algorithms. For example, Kafka, Flute, and Python are used to collect and transform data from different resources such as network KPIs, sensors of smart vehicles; programming platforms such as Matlab, R, Spark, and InfoSphere can be used for data analytics, visualization, data mining, and ML algorithms.

Distributed platform: Open5GMEC is a distributed computing system in which a host works as the master, and multiple hosts work as workers or executors as illustrated in Fig.3. Each host (master/slave) is a virtual machine working on SDN/NFV environment, and its components are Docker containers running on Linux systems and sharing the OS' kernel; therefore, these containers can be created and started instantly while consuming small resources. These software components usually run independently and concurrently on multiple virtual or physical machines. As a result, it is easy to deploy and upgrade Open5GMEC components. Moreover, in this framework, SDN controller applications are used to manage and control connections among virtual hosts through configuring OpenvSwitch.

Reliability: In the Open5GMEC platform, each component can utilize different methods such as partition, replication, and fault tolerance to improve the reliability of the whole system. For example, in a ZooKeeper cluster, when the primary master is a failure, the backup master takes over the role of primary master. Another example, Kafka stores critical information such as information about topics, consumer offsets, and brokers in the Zookeeper, which generally replicates this data across its ensemble. As a result, failure of Kafka broker or Zookeeper does not affect their clusters and achieve zero data loss, zero downtime. That means failure of a single or a few parts does not affect the system.

Flexibility: An application can be submitted their jobs to the master or any worker in the system. Once the master receives a job, it distributes the workload to its executors, optimizes and controls the number of executors based on the job computing load and the available worker resources.

Security: Security and privacy of the IoT services are critical challenges in the current study. A common threat is cyber- attacks, such as distributed denial of service (DDoS), Jamming attacks, privacy leakage, and man-in-the-middle. In the case of MEC, the data processing and computing are performed at the edge of a network close to the data source and mobile subscribers. Therefore, it bridges the gap between remote data centers and IoT devices to enable a wide range of security advantage. For example, ML can be utilized to detect abnormal attacks in networks such as learning-based IoT malware detection and learning-based authentication.

4 Applying Machine Learning to Open5GMEC and Potential V2X Applications

Recently, ML and big data have been utilized for empowering the SON of 5G, future MEC, and cloud computing to push their performances to the next level of full intelligence and automation [8][9][10][17]. The LTE/4G&5G network testbed, located at MIRC/BML has integrated big data and ML as the vital enabler to develop 5G applications. Generally, there are four categories of ML algorithms, namely, supervised learning, unsupervised learning, reinforcement learning, and deep learning.

4.1 ML categories

Supervised learning, the majority type of practical ML methods used in most current research, is a type of learning that requires a supervisor to learn the model parameters. Its models use a labeled dataset, which contains both input and output information, called training samples to find the relationship that maps from the input attribute space to the labels. As a result, the model gives the expected output for new coming input.

Unsupervised learning, on the other hand, is given an unlabeled input dataset. That means it does not have a supervisor. Therefore, it must investigate the similarity and the relationship among the unlabeled data samples, and then groups them into different clusters.

Reinforcement learning is an area of ML, where an agent learns the optimal behaviors in a trial-and-error manner by interacting with its environment, senses its current state and the state of the surroundings, and chooses an action to achieve a goal. It learns how to map from situations to actions, and therefore, the learner must identify which operation obtains the most reward.

Deep learning is a state-of-the-art and powerful algorithm with the sophistication of self-learning capability. It provides a significant improvement in various fields such as object detection, speech recognition, computer vision, and vehicle trajectory. In general, it is considered as a more in-depth version of neural networks (NN), which consist a series of multiple layers of neurons, the input layer, the hidden layers, and the output layer. Deep learning usually breaks down a very complicated problem into several simple issues to provide more accurate and faster processing.

4.2 Potential ML applications and implementation platforms

Fig.4 summarizes the most popular ML-based applications and the appropriate ML algorithms that can be exploited to empower the Open5GMEC with a full capacity of intelligence and automation.

Prediction and forecasting models are used in dynamic systems to precisely predict and estimate trends of events such as mobile traffic, vehicle mobility, vehicle tracking, and then, the system can keep track of those event's behaviors in changing environments. The suitable ML algorithms for these applications are Hidden Markov Models (HMMs), Linear/Non-linear Dynamical Systems.

Clustering is the most popular and powerful application of unsupervised learning to group a set of more similar data samples together, such as cluster traffic density, V2V neighbor, and abnormal detections. Typical ML algorithms for clustering model are K-means, Mixtures of Gaussians, Automatic relevance determination (ARD), etc.

Classification is the most typical ML models used in vehicular applications, for example, in this study, deep learning is used to classify and detect the object on the roads; Random Forest, SVM, NN, etc. are used to classify mobile broadband applications at the early state.

Diagnosis and decision making are used to analyze the current condition of a system or individual element in a network or vehicle to take timely controlling actions to ensure these components working at their peak performance even under complex situations. Moreover, rapid and relevant controls are essential for vehicular systems to develop safety and driver-assistance applications like power control, adjustment car direction, car speed. Applicable ML algorithms for these applications are Bayesian networks (BN), Reinforcement learning, logistic regression, decision tree, Gaussian Processes.

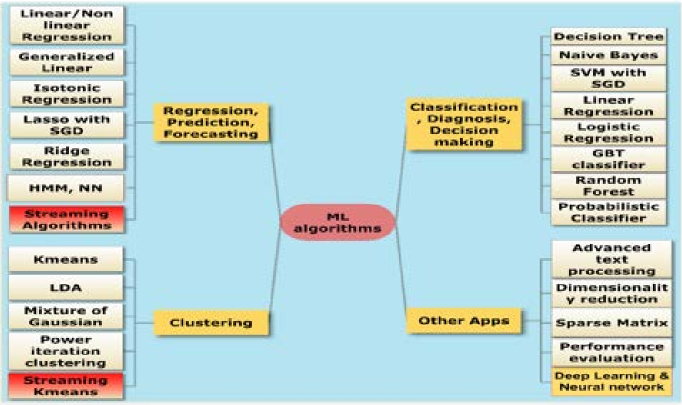


Figure 4. ML algorithms and Applications

Since the majority models of ML application in 5G and IoT are stream mining and distributed computing forms, it is necessary to build ML and big data on powerful platforms. Fortunately, emerging platforms like InfoSphere and Apache Spark are considered as comprehensive platforms supporting most of the popular ML algorithms and tools for big data analytics with various capacities

4.3 ML-based applications for vehicular communication

This subsection identifies the possible ML-based V2X applications in 5G networks. Fig.5 summarizes new applications and services that enable higher mobility, better coordinated, more enjoyable driving experience, more reliable for vehicular communication. For example:

Traffic prediction aims to accurately predict or forecast the future traffic of a road, road sections, even for an area. It has significant roles in improving traffic control, management and other ITS applications, such as traffic congestion avoidance, public vehicle deployment, and road hazard warning. Traffic prediction model usually uses real-time data collected by various roadway sensors or cameras. The prevalent model for traffic prediction is Time Series models, which are based mainly on the historical traces of traffic to forecast the future one. The suitable ML algorithms are dynamic models like HMMs, Deep Learning, and Gaussian Process (GP) [8][9]. LOBECOM 2011), 2011 IEEE, pp. 1–6, 2011.

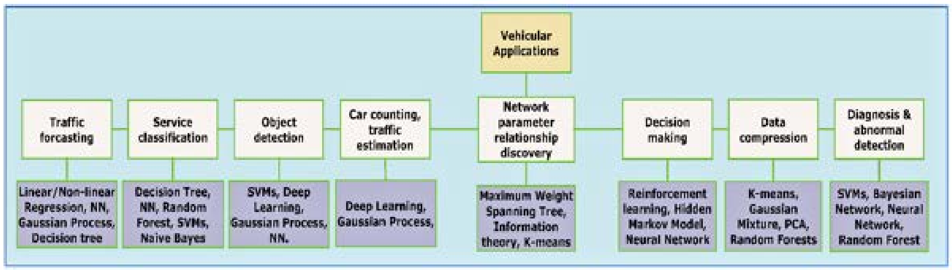

Vehicular Applications
Traffic forcasting
Service classification
Object detection
Car counting, traffic estimation
Network parameter relationship discovery
Decision making
Data compression
Diagnosis & abnormal detection
Linear/Non-linear Regression, NN, Gaussian Process, Decision tree
Decision Tree, NN, Random Forest, SVMs, Naive Bayes
SVMs, Deep Learning, Gaussian Process, NN.
Deep Learning, Gaussian Process,
Maximum Weight Spanning Tree, Information theory, K-means
Reinforcement learning, Hidden Markov Model, Neural Network
K-means, Gaussian Mixture, PCA, Random Forests
SVMs, Bayesian Network, Neural Network, Random Forest

Figure 5. ML-Based Applications for V2X Communication

Vehicle trajectory prediction is crucial for developing advanced driver-assistance systems and autonomous vehicles by providing a better understanding of the traffic environment to perform various functions such as criticality assessment in advance, collision avoidance, trajectory planning, and vehicle tracking [18]. Moreover, it also plays a vital role in mobile network planning and optimization, such as handover policies; therefore, it has recently received extensive research interest from both academia and industry. For example, research [19] introduced two approaches for vehicle trajectory prediction, the physics-based motion models and maneuver-based methods. These methods observed several running parameters of the vehicles, for example, acceleration, velocity, and direction rate to predict driving behavior using a dynamic Bayesian network. The most suitable ML algorithms are dynamic ML algorithms like KF, dynamic Bayesian networks, HMM, and deep learning.

5 V2X communication Application based on ML and open5G MEC

MEC for vehicular communication applications at BML/NCTU is described in Fig.6. MEC servers were deployed in BML of MIRC building, near the eNodeBs to reduce the delay caused by propagation for real-time applications, like MAR (mobile augmented reality)[14][20]. Smart devices in the vehicle collect and send data directly to the MEC servers through the 5G RAN. The MEC servers process and send the result to their clients to display [14]. The following section analyzes several V2X applications deployed in our platform.

5.1 Computer vision for V2X Based on Deep Learning

Vehicular applications based on computer vision techniques such as object identification and classification, lane-change detection, object density estimation, and vehicle-trajectory prediction have essential roles in developing Intelligent Transportation Systems. These applications enable vehicles to understand the road conditions and the vehicle's precise position; therefore, they can be applied for different purposes like obstacle avoidance, situational awareness, driving safety, roads maintenance, etc. The traditional ML algorithms for object classification are SVM, dynamic Bayesian networks, and K-Nearest Neighbor Classifier (KNN). However, recently, they have gradually been replaced by DL models (e.g., region-based convolutional neural networks (R-CCNs)) to provide real-time detectors satisfying the requirements of autonomous driving and V2X communications [21][22]. For example, research [22] proposed a high-accuracy model using DL for counting the objects in the crowded scene like a number of the car in traffic jam scenarios; research [21] applied the faster R-CNN for real-time object detection. Notably, there are available datasets for current object detection such as

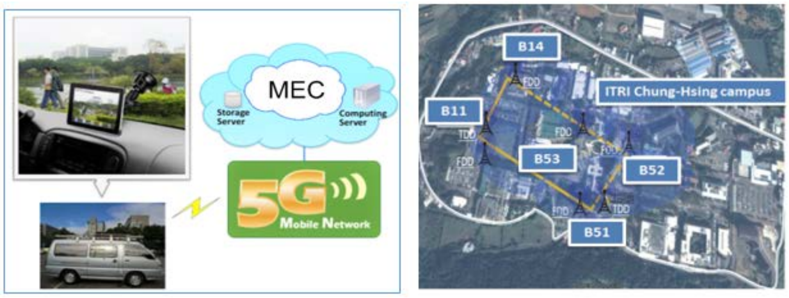


Figure 6. Open-5G MEC for V2X Application at BML/NCTU

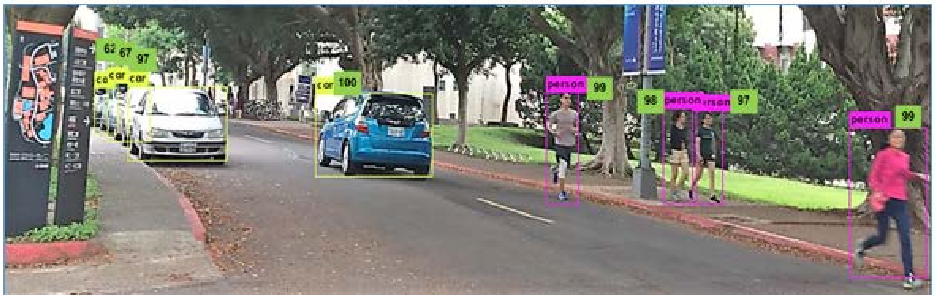


Figure 7. Object detection using deep learning at NCTU

CNN for real-time object detection. Notably, there are available datasets for current object detection such as Microsoft COCO datasets [23], which contain large-scale object detection with more than 91 common object categories, each category has more than 5,000 labeled instances. Up to the year 2018, the dataset has 2,500,000 labeled instances in 330,000 images.

In the past few years, computer vision has been applied to the 4G/LTE network testbed at NCTU; many applications were introduced [14][24][25]. For example, in research [24], we proposed a platform to recognize tactic patterns in broadcast basketball videos. This system used Kalman filter to automatically detects the court lines, tracks the players and ball, captures and analyzes ball trajectory, calibrates the players' positions to the real-world, etc. In research [25], we developed a preliminary system called YogaST, which utilized C++ with OpenNI 1.5.4.0 and OpenCV to assist the Yoga practitioner in self-training. It aims at instructing him/her to adjust his/her posture correctly and prevents injury caused by improper postures. Notably, research [14] performed MAR for outdoor vehicular navigation applications, such as corner detection, after that it also addressed the real-time challenge and estimate the performance of 4G/LTE and 5G based on outdoor navigation system at BML. This study applies R-CCN, which is one of the most popular DL models comprising many pooling and convolutional layers that resembles the human visual system, into object detection of autonomous vehicles. Fig.7 shows a result of object detection implemented at NCTU campus. As can be seen, it identified precisely cars and pedestrians on the road, the confidences of the detected objects were also shown.

In summary, R-CCN is robust algorithms, enabling computer vision to support many useful applications for automotive driving with high accuracies like object recognition, car counting, and vehicle tracking under different traffic conditions.

5.2 Two stages slicing for V2X applications

This section analyzes and applies network slicing, which can be considered as one of the most significant innovations and evolutions in 5G architectures, to support various V2X services by providing flexible subscription models and multiple end-to-end virtual networks that share the resources of a network operator. Generally, smart devices in a vehicle usually work as multi-slice devices, that means they can simultaneously attach to multiple slices for different purposes. For example, a driver could use a self-driving service controlled by an autonomous driving slice based on V2I and V2V communications. In the meanwhile, he/she opens an HD streaming of a

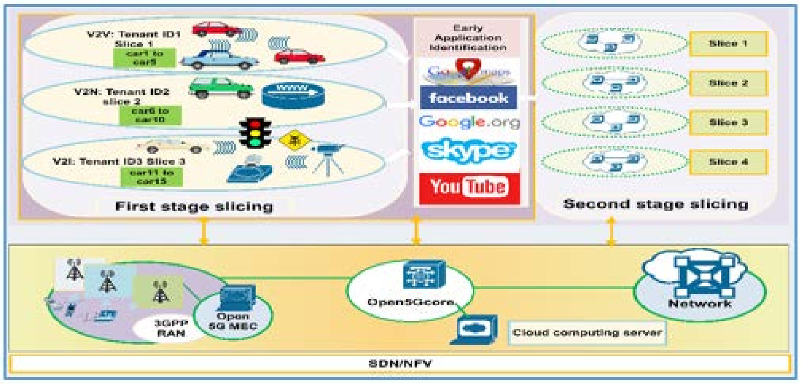


Figure 8. Two stages network slicing for V2X

different purposes. For example, a driver could use a self-driving service controlled by an autonomous driving slice based on V2I and V2V communications. In the meanwhile, he/she opens an HD streaming of a football game offered by an infotainment slice. The following subsection proposes a comprehensive framework called two-stage network slicing for V2X applications as shown in Fig.8. The first stage slices the network for different tenants or purposes, the second stage implements network slicing for various applications with different QoS requirements.

5.2.1 The first stage network slicing

Multi-tenancy is typical characteristics of V2X ecosystems, different tenants representing different services so that they should be mapped into different slices, for example, the slice for tele-operated driving and the slice for autonomous driving. These slices may be offered and managed by different service providers with diverse sets of performances and service requirements. For example, autonomous driving slices require latency less than 1ms, data rate 10Mb/s, and reliability nearly 100%, while vehicular internet and infotainment slices require latency around 10ms for web browsing, data rate 25Mb/s for UHD video streaming, and they do not concern about reliability.

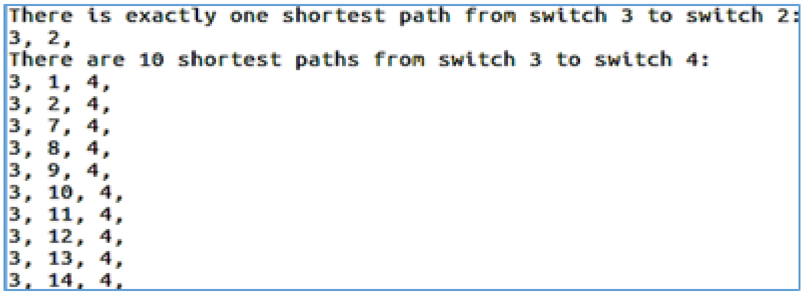
In this study, we assume that there are 3 slices representing for V2V, V2N, and V2I communications working on 5G network infrastructure under the control and management of open5Gcore and open5GMEC as described in Fig.8. For example, when a vehicle detects a dangerous obstacle on the road, it will send the warning information and the location information towards all the cars around it. Firstly, the car can directly send this information to cars very nearby using V2V communication (PC5 interface). In the meantime, it passes the information through eNodeB to MEC, and then, the SDN controller in MEC platform will flood the information to cars inside its slice. In this case, the slice is defined and managed based on the distance between cars and the obstacle. That means when a car moves into the range, it will be added to the slice's list or become a slice's member until it moves out of the range.

Experimental implementation

In this experiment, we assume that there are 16 cars, car1 to car15 belong to 3 slices, each slice contains 5 cars, and car16 does not belong to any slice. They are connected together by Open vSwitches controlled by the SDN controller. An SDN application was built on the controller to handle the first stage slicing that allows only cars in the same slice to communicate with one another. The SDN application has two main functions:

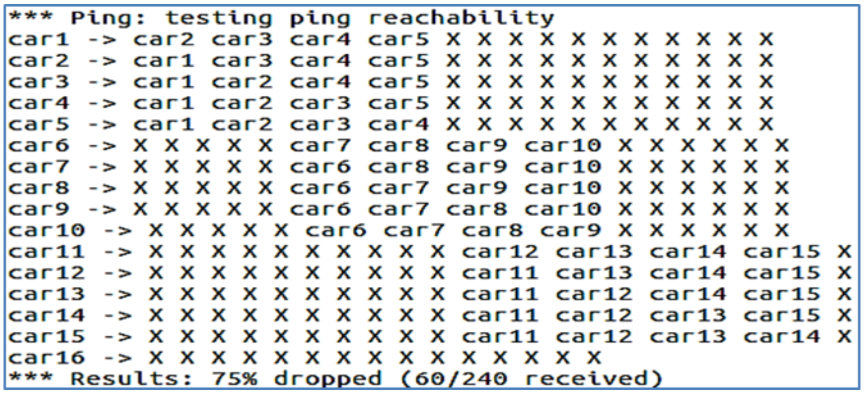
The primary function is to build the network topology, calculate and install all the shortest paths between all pair of sources and destinations in the network. For example, Fig.9 shows several couples of switches in the network and the shortest paths between them. Two hosts in the network can communicate through a random route in the shortest paths or by applying a load balancing algorithm.

The second function applies network slicing policies to isolate a slice from others by checking the source and destination IP addresses of the communicating packets. If they belong to hosts (cars) of the same slice (tenant),



```
There is exactly one shortest path from switch 3 to switch 2:
3, 2,
There are 10 shortest paths from switch 3 to switch 4:
3, 1, 4,
3, 2, 4,
3, 7, 4,
3, 8, 4,
3, 9, 4,
3, 10, 4,
3, 11, 4,
3, 12, 4,
3, 13, 4,
3, 14, 4,
```

Figure. 9 Example shortest paths of couples of switches



```
*** Ping: testing ping reachability
car1 -> car2 car3 car4 car5 X X X X X X X X X X X
car2 -> car1 car3 car4 car5 X X X X X X X X X X X
car3 -> car1 car2 car4 car5 X X X X X X X X X X X
car4 -> car1 car2 car3 car5 X X X X X X X X X X X
car5 -> car1 car2 car3 car4 X X X X X X X X X X X
car6 -> X X X X X car7 car8 car9 car10 X X X X X X
car7 -> X X X X X car6 car8 car9 car10 X X X X X X
car8 -> X X X X X car6 car7 car9 car10 X X X X X X
car9 -> X X X X X car6 car7 car8 car10 X X X X X X
car10 -> X X X X X car6 car7 car8 car9 X X X X X X
car11 -> X X X X X X X X X X car12 car13 car14 car15 X
car12 -> X X X X X X X X X X car11 car13 car14 car15 X
car13 -> X X X X X X X X X X car11 car12 car14 car15 X
car14 -> X X X X X X X X X X car11 car12 car13 car15 X
car15 -> X X X X X X X X X X car11 car12 car13 car14 X
car16 -> X X X X X X X X X X X X X X X
*** Results: 75% dropped (60/240 received)
```

Figure 10. Ping result after applying slicing for V2X

destination IP addresses of the communicating packets. If they belong to hosts (cars) of the same slice (tenant), they are allowed to communicate. Otherwise, the packets are dropped. Fig.10 shows the ping result after acting the slicing function. As can be seen, only the cars in the same tenant can communicate with others, and since car16 does not belong to any tenant, the communicating packets between it and others are dropped. This function works as a firewall to logically isolate network functions and resources among slices.

5.2.2 The second stage network slicing

The second stage of network slicing is to guarantee QoS, QoE, user experience, and network resource efficiency for different V2X services. For example, in the slice of vehicular infotainment services, each application requires a QoS guarantee. Generally, a high-definition video streaming requires up to gigabytes per second peak data rate, small delay and jitter to provide seamless communication for end devices in a vehicle with high mobility, while other data services like web browsing, Gmail, and Google are more sensitive to packet loss. Hence, the network operator must provide different policies for them, for instance, preserve a relevant bandwidth for each application. To solve this problem, in previous research

[17], we proposed and implemented a QoS control model for different mobile broadband applications. The model's process is described in Fig.8 and Fig.11; it involves two steps, namely, the early-state traffic flow classification and network QoS control for traffic flow using an SDN application.

Firstly, we applied several state-of-the-art classification algorithms like Naïve Bayes, Gradient Boosted Tree (GBT), Random Forest, SVM, and NN to identify traffic flows at the early stage [17]. Moreover, in this study, we applied several methods: cross-validation, parameter optimizations, and prediction fusion that combines SVM classifier and Naïve Bayes classifier to improve the performance of classification models. The result shows that all models achieved high accuracy to identify traffic flows at the early-state as summarized in Table 1.

Next, the second step utilizes the result of the traffic flow classification to build the SDN application for QoS control. Currently, SDN/NFV have been considered as emerging technologies enabling new robust methods for QoS control, such as queue management, the utilization of meter tables, ingress policing, etc. These approaches are more powerful, flexible, and easier than the traditional technologies like Integrated Service (IntServ) and Differentiated Service (DiffServ). The IntServ method is too complicated and not scalable, while DiffServ is less complicated but it does not provide strong QoS guarantees. Moreover, among the SDN-based approaches, using meter tables and QoS queue is the most popular methods. They install flow table entries for each connection to instruct Open vSwitches how to execute the flows (packets). For example, in the QoS queue approach, which is an egress packet queuing mechanism in Switch ports, each application is assigned a QoS ID in which we can set both min data rate and max data rate. Also, we can easily create, modify, delete a queue through the Open vSwitch configuration protocol like OF-Config (or ovs-vsctl command). On the other hand, the meter table method allows monitoring the ingress rate of flows by meter tables, which are attached directly to flow table entries, to measure and control the data rate of packets. Fig.11 illustrates the abstract process of implementing the second stage network slicing for uplink directions. In this framework, the classification model is deployed on the Open5GMEC close to the eNodeBs to identify and label traffic flows. When the SDN controller receives an identified traffic flow, it extracts and stores the necessary information of the connection, such as application label or type, source IP address, and destination IP address. After that, it installs flow table entries and QoS policies on OpenFlow Switch to guarantees bandwidth for that connection on both uplink and downlink directions

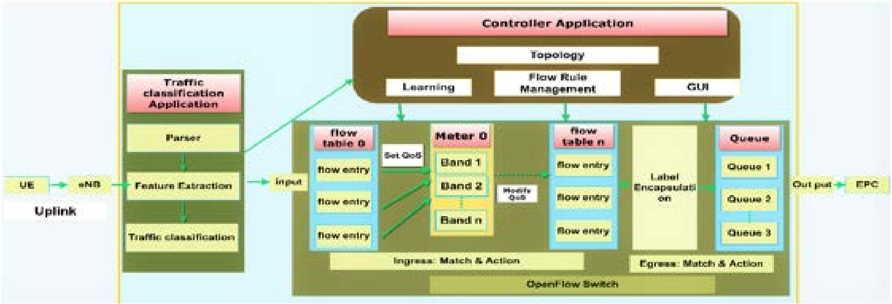
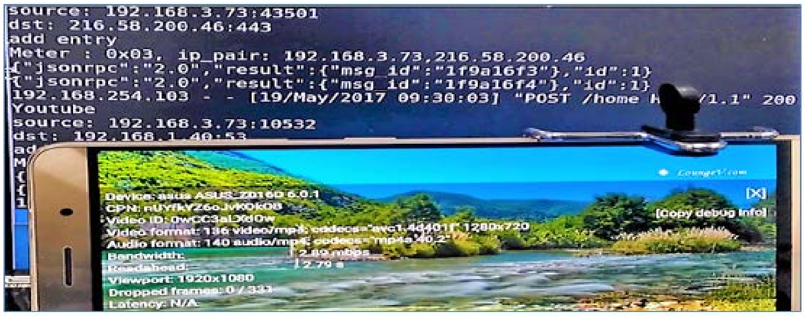


Figure 11. The second stage network slicing process



```
source: 192.168.3.73:43501
dst: 216.58.200.46:443
add entry
Meter : 0x03, ip_pair: 192.168.3.73,216.58.200.46
{"jsonrpc":"2.0","result":{"msg_id":"1f9a16f3"},"id":1}
{"jsonrpc":"2.0","result":{"msg_id":"1f9a16f4"},"id":1}
192.168.254.103 - - [19/May/2017 09:30:03] "POST /home H   /1.1" 200
Youtube
source: 192.168.3.73:10532
dst: 192.168.1.40:53
```

```
Device: asus ASUS_Z016D 6.0.1
CPN: nUYfkYZ6oJvKOkO8
Video ID: 0wCC3aLXd0w
Video format: 136 video/mp4; codecs="avc1.4d401f" 1280x720
Audio format: 140 audio/mp4; codecs="mp4a.40.2"
Bandwidth:        2.89 mbps
Readahead:        2.79 s
Viewport: 1920x1080
Dropped frames: 0 / 331
Latency: N/A
```

Figure 12. QoS control based on application identification

In our experiment, YouTube was preserved a bandwidth of 3Mbps in the QoS control, and then we used a UE to play a video from YouTube. Fig 12 shows the result of the test. It is clear that YouTube traffic flows were identified and classified precisely at the early state, and the bandwidth for YouTube at that time was 2.88Mbps. This bandwidth guarantees for QoS of the video to play smoothly without any dropped frame. In summary, SDN/NFV have crucial roles in creating and managing on-demand and wide-range network slicing services for V2X and 5G to build multiple logical end-to-end networks.

6 SDN/NFV Driven Service Migration for V2X

V2X communication is characterized by highly dynamic network topologies providing services for vehicles while they are moving quickly in and out of cells. The critical requirements of V2X services are not only about real-time services but also the transparent of connections on the IP layer, and TCP sessions, which are unable to be solved by applying fast-handover approaches. Moreover, a real network always encounters abnormal problems that may affect network QoS while operating due to equipment, nodes, and links failures. It becomes a big challenge for modern V2X communications because their services usually require incredibly fault-tolerant systems that can guarantee fast recovery. This section proposes a practical solution called service migration based on SDN aspects to deal with those challenges by moving a running application from a server to another one without any disruption. The process of service migration can be divided into three steps The first step: once the controller application receives a handover request from a service due to the movement of a user or due to a failure at a server called the old server, it triggers the service migration process to select an available server called migration server by using an optimization algorithm (e.g., based on load balancing).

Table 1. Accuracy of the early-state traffic classification models (%)

Application	SSL.Gmail	Amazon	WhatsApp	Facebook	Snapchat	Skype	Google	Instagram	YouTube	Apple
Naïve Bayes	90.1	89.5	93.4	91.6	87.8	94.8	85.8	87.9	94.1	93.4
SVM	93.6	91.3	95.7	94.9	92.6	97.2	90.6	92.8	97.8	94.8
SVM+Naïve Bayes	94.5	92.1	96.2	96.8	94.3	98.4	92.1	93.7	98.5	96.4
NN	93.4	94.5	97.9	95.1	96.4	97.2	93.8	97.5	99.2	98.5
GBT	98.6	98.7	99.5	98.1	99.2	99.8	99.4	99.6	100	99.4
Random Forest	98.9	99.2	99.6	98.4	99.8	100	99.6	99.2	100	99.7

After that, it calculates all the shortest paths between the client and migration servers (this function was described in the first-stage slicing section).

The second step: Delete all routing flows or table entries that relate to the old server and the current client in Open vSwitches. So that, when a switch receives a new packet from the user sent to the old server, it will ask the controller through Packet-In messages.

The third step: Once the SDN controller receives the Packet-In message sending to the old server, it sends Packet-Out messages to set rules for redirecting all flows between the client and the old server on both directions: Modify the destination IP and MAC of flow packets that are sending from the client to old server to the IP and MAC addresses of the migration server; modify the source IP and MAC addresses of flow packets that are sending from the migration server to the client to the IP and MAC addresses of the old server.

Experimental implementation: The experimental topology and scenario are illustrated in Fig.13. Worker1 and worker3 work as operating and available servers, respectively, running the same application. Worker8 works as a client (or car) served by worker1. In this test, the worker1 was turned down, SDN application selected server3 as migration server to provide service for the client. We performed a Ping test and TCP test services to evaluate the results.

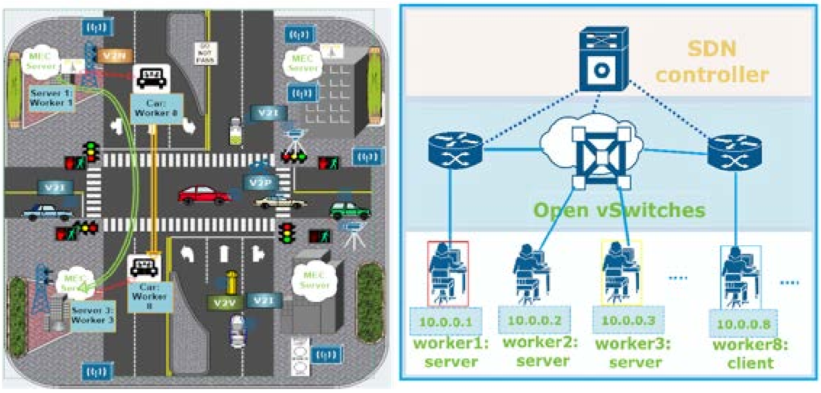
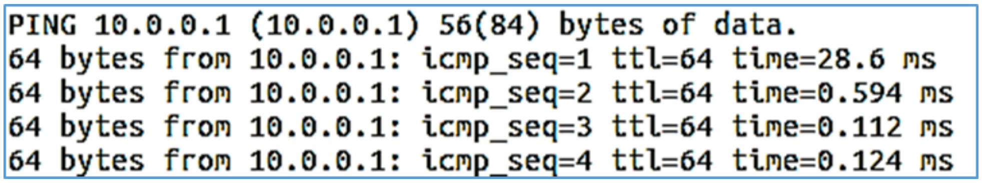


Figure 13. Experimental scenarios

Ping test result: Fig.14 shows the result of ping from worker8 to worker1. As can be seen, the client smoothly received reply messages with IP of worker1 even though worker1 was out of service. The time to receive the first message was 28.6ms much longer than the time of the following messages. It involves the time when a switch receives the first message of the connection; it must send packet-in to ask the controller, which will check the packet headers and then install the rules in table flow entries to instruct the switch how to handle this message and later messages. For more information, Wireshark was used to capture Ethernet packets at worker8 and worker3; the results are shown in Fig.15 and Fig.16, respectively. It is clear that all the messages sent from worker8 to worker1 were redirected to worker3 as shown in Fig.15; all the messages from worker3 to worker8 were modified the source IP address from those of worker3 to those of worker1 as shown in Fig.15.

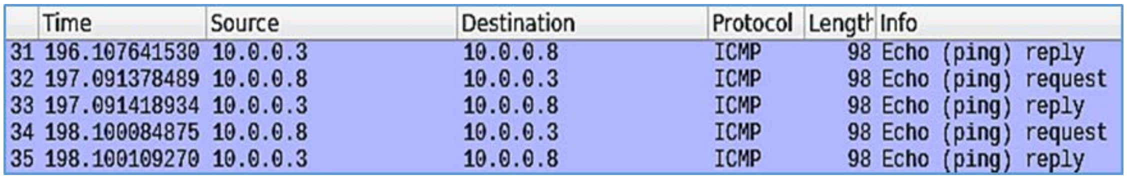
TCP service test result: Fig.16 describes our experiment, worker1 and worker3 run as TCP servers listening at port 4444, worker8 run as a TCP client connecting to port 4444 of server1. The results in Fig.17 shows that the TCP service is served by worker3, not worker1

In summary, the Open5GMEC platform supports transparent migration, which is essential in V2X communication to enable more robust applications regarding resource management, service recovery, computing scaling, handover for high mobility services



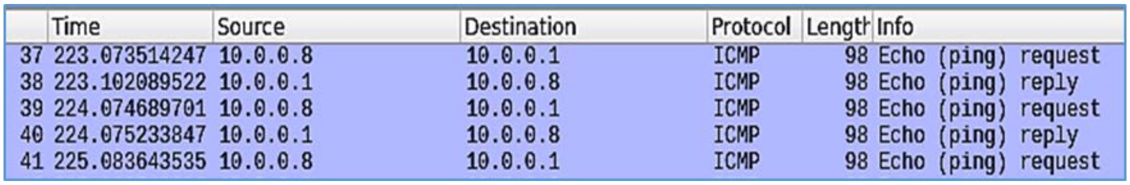
```
PING 10.0.0.1 (10.0.0.1) 56(84) bytes of data.
64 bytes from 10.0.0.1: icmp_seq=1 ttl=64 time=28.6 ms
64 bytes from 10.0.0.1: icmp_seq=2 ttl=64 time=0.594 ms
64 bytes from 10.0.0.1: icmp_seq=3 ttl=64 time=0.112 ms
64 bytes from 10.0.0.1: icmp_seq=4 ttl=64 time=0.124 ms
```

Figure 14. Ping test of MEC server migration



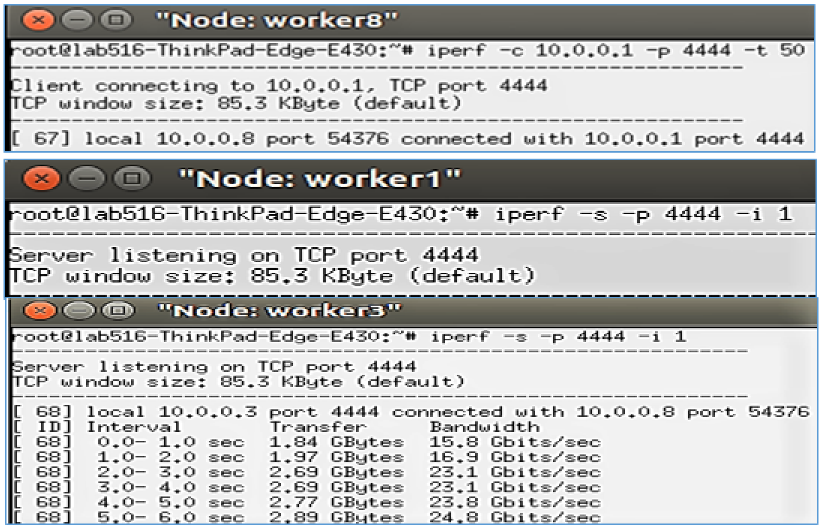
	Time	Source	Destination	Protocol	Length	Info
31	196.107641530	10.0.0.3	10.0.0.8	ICMP	98	Echo (ping) reply
32	197.091378489	10.0.0.8	10.0.0.3	ICMP	98	Echo (ping) request
33	197.091418934	10.0.0.3	10.0.0.8	ICMP	98	Echo (ping) reply
34	198.100084875	10.0.0.8	10.0.0.3	ICMP	98	Echo (ping) request
35	198.100109270	10.0.0.3	10.0.0.8	ICMP	98	Echo (ping) reply

Figure 15. Captured packets of the Ping test at worker3



	Time	Source	Destination	Protocol	Length	Info
37	223.073514247	10.0.0.8	10.0.0.1	ICMP	98	Echo (ping) request
38	223.102089522	10.0.0.1	10.0.0.8	ICMP	98	Echo (ping) reply
39	224.074689701	10.0.0.8	10.0.0.1	ICMP	98	Echo (ping) request
40	224.075233847	10.0.0.1	10.0.0.8	ICMP	98	Echo (ping) reply
41	225.083643535	10.0.0.8	10.0.0.1	ICMP	98	Echo (ping) request

Figure 16. Captured packets of the Ping test at worker8



```
root@lab516-ThinkPad-Edge-E430:~# iperf -c 10.0.0.1 -p 4444 -t 50
------------------------------------------------------------
Client connecting to 10.0.0.1, TCP port 4444
TCP window size: 85.3 KByte (default)
------------------------------------------------------------
[ 67] local 10.0.0.8 port 54376 connected with 10.0.0.1 port 4444
```

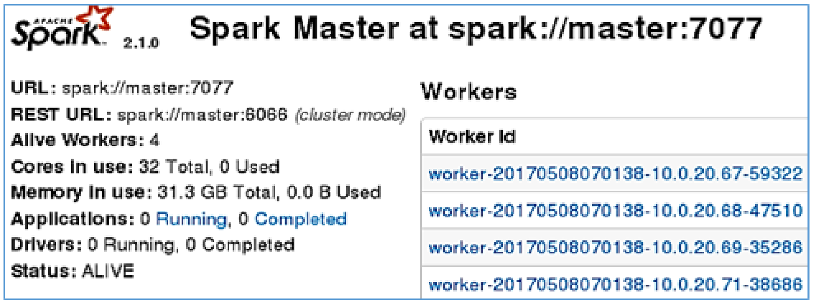
```
root@lab516-ThinkPad-Edge-E430:~# iperf -s -p 4444 -i 1
------------------------------------------------------------
Server listening on TCP port 4444
TCP window size: 85.3 KByte (default)
------------------------------------------------------------
```

```
root@lab516-ThinkPad-Edge-E430:~# iperf -s -p 4444 -i 1
------------------------------------------------------------
Server listening on TCP port 4444
TCP window size: 85.3 KByte (default)
------------------------------------------------------------
[ 68] local 10.0.0.3 port 4444 connected with 10.0.0.8 port 54376
[ ID] Interval       Transfer     Bandwidth
[ 68]  0.0- 1.0 sec  1.84 GBytes  15.8 Gbits/sec
[ 68]  1.0- 2.0 sec  1.97 GBytes  16.9 Gbits/sec
[ 68]  2.0- 3.0 sec  2.69 GBytes  23.1 Gbits/sec
[ 68]  3.0- 4.0 sec  2.69 GBytes  23.1 Gbits/sec
[ 68]  4.0- 5.0 sec  2.77 GBytes  23.8 Gbits/sec
[ 68]  5.0- 6.0 sec  2.89 GBytes  24.8 Gbits/sec
```

Figure 17. TCP test result at client and servers

To evaluate the computing capacity of the platform, we run the classification application (described in the first stage slicing section) with a specific configuration. Fig.18 shows the Spark computing cluster, which contains four computing workers with 32 cores and 31.3 GB memory. A stream of 210.000.000 data samples was generated with a different data rate as in Fig. 19. After receiving the data, Apache Kafka created a topic and then produced the data as the classification testing dataset to Apache Spark. After that, the incoming samples were classified by SVM classifier built from Spark MLlib 2.0. Generally, when a new input data is received, Spark streaming model continually records, processes, and updates model parameters. However, to attain better performance, this study used the mini-batch method in which the stream of data is discretized into a sequence mini-batches called sequence of RDD (Resilient Distributed

Datasets). The mini-batch length can be defined by either a number of records or time interval. Here, we used the time duration of 3 seconds for each mini-batch. Fig. 23 shows the input and the output data rate. Fig. 20 shows the time, mini-batch sizes (number of records), scheduling time, processing and the total delay of several mini-batches that have the highest input data rate. It is clear that the computing platform is powerful in providing high computing capacity with a small time for scheduling and processing so that all the data samples were classified smoothly with low-total-latency. In summary, Open5GMEC is powerful enough to be deployed for the industrial networks



Spark 2.1.0 Spark Master at spark://master:7077

URL: spark://master:7077
REST URL: spark://master:6066 *(cluster mode)*
Alive Workers: 4
Cores in use: 32 Total, 0 Used
Memory in use: 31.3 GB Total, 0.0 B Used
Applications: 0 Running, 0 Completed
Drivers: 0 Running, 0 Completed
Status: ALIVE

Workers

Worker Id
worker-20170508070138-10.0.20.67-59322
worker-20170508070138-10.0.20.68-47510
worker-20170508070138-10.0.20.69-35286
worker-20170508070138-10.0.20.71-38686

Figure 18. Spark computing cluster

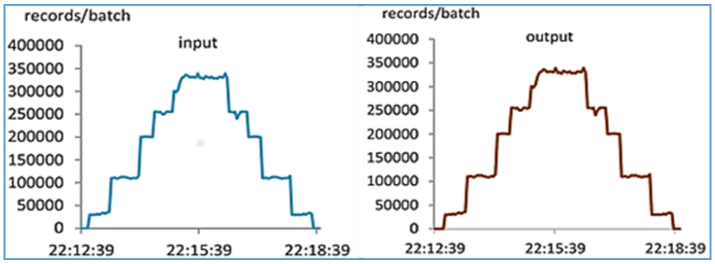
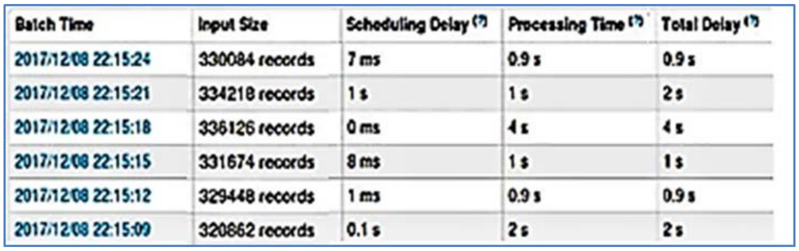


Figure 19. Input and output timelines



Batch Time	Input Size	Scheduling Delay (?)	Processing Time (?)	Total Delay (?)
2017/12/08 22:15:24	330084 records	7 ms	0.9 s	0.9 s
2017/12/08 22:15:21	334218 records	1 s	1 s	2 s
2017/12/08 22:15:18	336126 records	0 ms	4 s	4 s
2017/12/08 22:15:15	331674 records	8 ms	1 s	1 s
2017/12/08 22:15:12	329448 records	1 ms	0.9 s	0.9 s
2017/12/08 22:15:09	320862 records	0.1 s	2 s	2 s

Figure 20. Mini-batches computing timeline

7 Conclusion

The proposed Open5GMEC integrating state-of-the-art technologies, such as SDN/NFV, big data, and ML, is expected as a comprehensive solution for future 5G MEC and vehicular communications. It supports the full power of programmability and virtualization, robust and fast computation, automatic and intelligent optimization to make network components more transparent and coordinated. Furthermore, based on the platform, this study analyzed and implemented several significant applications for V2X communication that their results are pioneers for building more complex applications in V2X and 5G areas.

Our future work will focus on applying the framework to develop other MEC and CORD components and V2X applications such as SDN-based R-CORD and M-CORD.

ACKNOWLEDGMENT

This paper is particularly supported by "the Center for Open Intelligent Connectivity from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project" by the Ministry of Education (MOE) in Taiwan, and the Ministry of Science and Technology of Taiwan under Grants: MOST 107-2221-E-009-056.

REFERENCES

[1] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A Survey on Mobile Edge Computing: The Communication Perspective," vol. 19, no. 4, pp. 2322–2358, 2017.

[2] M. C. Computing, X. Chen, L. Jiao, and W. Li, "Efficient Multi-User Computation Offloading for Mobile Edge Cloud Computing," vol. 24, no. 5, pp. 2795–2808, 2016.

[3] S. A. A. Shah, E. Ahmed, M. Imran, and S. Zeadally, "5G for Vehicular Communications," IEEE Commun. Mag., vol. 56, no. 1, pp. 111–117, 2018.

[4] S. Chen et al., "Vehicle-to-Everything (v2x) Services Supported by LTE-Based Systems and 5G," IEEE Commun. Stand. Mag., vol. 1, no. 2, pp. 70–76, 2017.

[5] R. Molina-Masegosa and J. Gozalvez, "LTE-V for Sidelink 5G V2X Vehicular Communications: A New 5G Technology for Short-Range Vehicle-to-Everything Communications," IEEE Veh. Technol. Mag., vol. 12, no. 4, pp. 30–39, 2017.

[6] X. Duan, Y. Liu, and X. Wang, "SDN enabled 5G-VANET: Adaptive vehicle clustering and beamformed transmission for aggregated traffic," IEEE Commun. Mag., vol. 55, no. 7, pp. 120–127, 2017.

[7] H. C. Chang et al., "Empirical Experience and Experimental Evaluation of Open5GCore over Hypervisor and Container," Wirel. Commun. Mob. Comput., vol. 2018, no. i, 2018.

[8] Le Luong Vy; Li-Ping Tung; Bao-Shuh Paul Lin, "Big data and machine learning driven handover management and forecasting," in IEEE Standards for Communications and Networking (CSCN), Helsinki, 2017, 2017, pp. 214–219.

[9] L. Le, D. Sinh, L. Tung, and B. P. Lin, "A practical model for traffic forecasting based on big data, machine-learning, and network KPIs," in 2018 15th IEEE Annual Consumer Communications & Networking Conference (CCNC), 2018, pp. 1–4.

[10] N. Cheng et al., "Big Data Driven Vehicular Networks," IEEE Netw., vol. PP, pp. 1–8, 2018.

[11] D. Sinh, L. Le, L. Tung, and B. P. Lin, "The Challenges of Applying SDN / NFV for 5G & IoT," 14th IEEE - VTS Asia Pacific Wirel. Commun. Symp. (APWCS), Incheon, Korea, Sep 2017.

[12] M. Boban, A. Kousaridas, K. Manolakis, J. Eichinger, and W. Xu, "Use Cases, Requirements, and Design Considerations for 5G V2X," Netw. Internet Archit., pp. 1–10, 2017.

[13] L. Le, D. Sinh, B. P. Lin, and L. Tung, "Applying Big Data, Machine Learning, and SDN/NFV to 5G Traffic Clustering, Forecasting, and Management," IEEE NetSoft, vol. 870, no. NetSoft, pp. 168–176, 2018.

[14] B. P. Lin, L. Tung, F. Tseng, I. Hsieh, Y. Wang, and S. Chou, "Performance Estimation of MAR for Outdoor Navigation Applications based on 5G Mobile Broadband by using Smart Mobile Devices," in Conference: IEEE VTS APWCS 2015, Singapore.

[15] B. P. Lin, F. J. Lin, and L. Tung, "The Roles of 5G Mobile Broadband in the Development of IoT, Big Data, Cloud and SDN," Commun. Netw., vol. 8, no. 1, pp. 9–21, 2016.

[16] I. C. Hsieh, L. P. Tung, and B. S. P. Lin, "On the classification of mobile broadband applications," IEEE Int. Work. Comput. Aided Model. Des. Commun. Links Networks, CAMAD, pp. 128–134, 2016.

[17] Luong-Vy Le; Bao-Shuh Lin; Do Sinh, "Applying Big Data, Machine Learning, and SDN/NFV for 5G Early-Stage Traffic Classification and Network QoS Control," Trans. Networks Commun., vol. 6, no. 2, pp. 36–50, 2018.

[18] G. X. and H. G. and L. Q. and B. H. and K. L. and J. Wang, "Vehicle Trajectory Prediction by Integrating Physics- and Maneuver-Based Approaches Using Interactive Multiple Models," IEEE Trans. Ind. Electron., vol. 65, no. 7, pp. 5999–6008, 2018.

[19] L. Liang, H. Ye, and G. Y. Li, "Towards Intelligent Vehicular Networks: A Machine Learning Framework," pp. 1–11, 2018.

[20] B. S. Lin et al., "The design of big data analytics for testing & measurement and traffic flow on an experimental 4G/LTE network," 2015 24th Wirel. Opt. Commun. Conf. WOCC 2015, pp. 40–44, 2015.

[21] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1137–1149, 2017.

[22] D. Oñoro-Rubio and R. J. López-Sastre, "Towards perspective-free object counting with deep learning," Comput. Vis. -- ECCV 2016, vol. 9911, pp. 615–629, 2016.

[23] T. Lin, C. L. Zitnick, and P. Doll, "Microsoft COCO : Common Objects in Context," Comput. Vis. -- ECCV 2014, pp. 740--755, 2014.

[24] H. T. Chen, C. L. Chou, T. S. Fu, S. Y. Lee, and B. S. P. Lin, "Recognizing tactic patterns in broadcast basketball video using player trajectory," J. Vis. Commun. Image Represent., vol. 23, no. 6, pp. 932–947, 2012.

[25] H. T. Chen, Y. Z. He, C. L. Chou, S. Y. Lee, B. S. P. Lin, and J. Y. Yu, "Computer-assisted self-training system for sports exercise using kinects," 2013 IEEE Int. Conf. Multimed. Expo Work., pp. 3–6, 2013.

Society for Science and Education
United Kingdom

TNC Transactions on Networks and Communications

Volume 6, No. 5
ISSN: 2054 -7420

Performance Improvement of OOFDM Systems Based On Advanced Logarithmic Companding Technique

[1]Ali N. Kareem, [2]Sinan M. Abdul Satar, [3]Mohammed A. Husein, [4]Liqaa A. Al-Hashime, [5]Ghaida A. Al-Suhail
[1,2,3]*Department of Electrical Engineering, University of Technology, Baghdad, Iraq;*
[4,5]*Department of Electrical Engineering, University of Basra, Basra, Iraq;*
alinkareem@yahoo.com; sinansma@yahoo.com; mfzay@yahoo.com; leqaa.abdulsattar@gmail.com; ghaida_alsuhail@yahoo.com;

ABSTRACT

High peak-to-average power ratio (PAPR) is a common problem in Orthogonal Frequency Division Multiplexing (OFDM) transmissions. In this paper, an advanced logarithmic companding technique is interested to reduce this factor due to their flexibility and low complexity. We evaluate the performance of the proposed advanced logarithmic technique comparing with typical logarithmic and un-companded schemes via simulations in Intensity Modulation/Direct Detection Optical Orthogonal Frequency Division Multiplexing (IM/DD OOFDM) system. The proposed advanced logarithmic Companding technique guarantees the improved performance in terms of Bit Error Rate (BER) and Quality Factor (QF) while reducing PAPR effectively and efficiently by modifying the amplitude of the transmitted signals. Our results confirm that the suggested scheme exhibits a good ability to reduce PAPR and a good BER performance based on the use the both of (k and y) factors to be chosen in relation to acceptable or desired PAPR, BER, and QF requirements. At the complementary cumulative distribution function (CCDF) of 10^{-3}, the PAPR value of our proposed scheme is about 6.4 dB lower than those of un-companded signal at best control factor used of y=0.1, accordingly the QF is 11.8 dBm and the BER is 4.9×10^{-5}.

Keywords: Peak-to-Average Power Ratio (PAPR), Orthogonal Frequency Division Multiplexing (OFDM), Intensity Modulation/Direct Detection Optical Orthogonal Frequency Division Multiplexing (IM/DD OOFDM), Quality Factor (QF), Bit Error Rate (BER), Complementary Cumulative Distribution Function (CCDF).

1 Introduction

An Orthogonal Frequency Division Multiplexing (OFDM) has been using in the optical communication systems because of the following advantages: immunity to Inter-symbol Interference (ISI), high spectral efficiency, co-channel interference and impulsive parasitic noise, lower implementation complexity in comparison with the single carrier solution [1, 2]. On the logical ground, the intensity Modulation/Direct Detection Optical OFDM (IM/DD OOFDM) system has been widely investigated in high-speed optical Communications, short-range as well as the cost-sensitive [3]. It has been illustrated that in addition to their many advantages, OOFDM systems have also several disadvantages such as ISI and Peak-to-Average

DOI: 10.14738/tnc.65.5065
Publication Date: 07th September, 2018
URL: http://dx.doi.org/10.14738/tnc.65.5065

Power Ratio (PAPR), which will be focused in this paper. Moreover, the high peaks to average power ratio set high power as an input to the Power Amplifier at the transmitter [2].

The high power amplifier (HPA) is working in the nonlinear section, therefore, In-band and out-of-band interferences are increased due to this nonlinearity. Reducing the PAPR will enhance HPA performance, reduce power consumption, reduce signal distortion by HPA and improve bit error ratio (BER) performance. The performance of the transmitter can be enhanced by using PAPR reduction techniques [4].

Multiple PAPR reduction techniques are offered for multi-carrier systems such as Clipping and pre-distortion [5-7], Partial Transmit Sequence, Selective Mapping [8-10], nonlinear companding [11-13], Tone Injection, Tone Reservation [14-16], etc. on all those PAPR reduction systems, the increasing in bandwidth and large memory with complexity in these techniques, so that the clipping and Companding technique is useful for PAPR to overcome this constraint [4, 5].

The clipping process is entirely the easiest system to use but this processing leads to distortions and causes an increase in the BER of the system [7]. Therefore, The other way is using the companding techniques that give better performance than clipping techniques due to the fact that the companding transformation is done at the transmitter to attenuate the high peaks, where as an increase low peaks and an inverse companding transform is done on the receiver end to reduce the distortion of signals and pick up the original signal, before transmission [6] [11].

The OOFDM signal involves severally modulated subcarriers that may provide a large PAPR once value-added up coherently. An OOFDM signal features a massive PAPR that is terribly sensitive to the non-linearity of the high peak [3]. To illustrate this point, blocks of the symbol in the OFDM are selected with every symbol modulating one from a set of subcarriers and these subcarriers are recognized to be orthogonal [8]. The representation of complex OFDM signals is given as:

$$x(K) = \frac{1}{N}\sum_{n=0}^{N-1} X_n \ e^{\frac{j2\pi nk}{N}} \tag{1}$$

Where:
$K = 0,1,2,\dots,N-1$

2 The Peaks to Average Power Ratio (PAPR)

The PAPR for a signal x(K) is outlined as the ratio of maximum instantaneous power to the average power as illustrated in (2) [3]:

$$\text{PAPR}=10*\log\left(\frac{\max|x(K)|^{\,2}}{\text{Avg}|x(K)|^{\,2}}\right) \tag{2}$$

Furthermore, the reduction of PAPR simply shows the probabilities that the PAPR of data block exceeds a given threshold value. The expression of Complementary Cumulative Distribution Function (CCDF) is given in (3) (Akhtman et al.,2003)[2]:

$$\text{CCDF=Probability(PAPR>PAPRth)} \tag{3}$$

Where PAPRth represents the Threshold level.

3 Advanced Logarithmic Companding Technique

It has been noticed that Due to low complexness regardless of the number of subcarriers, the Companding technique has been using for PAPR reduction in OFDM systems [2]. Moreover, the logarithmic transform using at the transmitter as a compressor after Inverse Fast Fourier Transform (IFFT) that is described in equation (4) and an expander before Fast Fourier Transform (FFT) at the receiver that described in equation (5):

$$\mathrm{f(x)} = \log\big(1 + (\mathrm{K} * |\mathrm{x}|)\big) * \mathrm{sgn(x)} \tag{4}$$

$$\mathrm{f(x)}^{-1} = |(\exp\left(\frac{|\mathrm{x}|}{\mathrm{K}}\right) - 1)| * \mathrm{sgn(x)} \tag{5}$$

Where:
sgn: sign function
k: is a positive number for controlling the amount of companding.

Additionally, a new modification of the logarithmic technique has been proposed depending on the original equation by proposing a new factor (y) that's also controlling the amount of companding as well as maintain the input and output signals at the same average power level. The advanced companded function is given in equation (6) and equation (7):

$$\mathrm{f(x)} = \log(1 + (\mathrm{K} * |\mathrm{x}|)^{\mathrm{y}}) * \mathrm{sgn(x)} \tag{6}$$

$$\mathrm{f(x)}^{-1} = |(\exp\left(\frac{|\mathrm{x}|}{\mathrm{K}}\right) - 1)^{1/\mathrm{y}}| * \mathrm{sgn(x)} \tag{7}$$

Where:

y: positive number (1, 0.9, …, 0.1)

4 The Proposed Modeling OOFDM System

Firstly, the traditional configuration of the OOFDM system without using the companded technique is shown in figure 1.

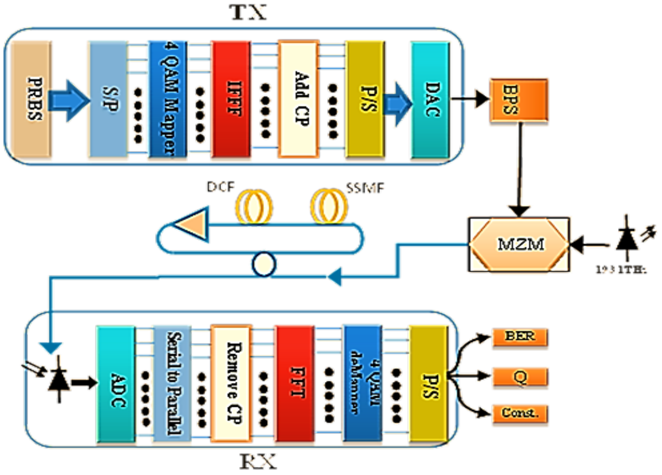


Figure 1. The traditional OOFDM system.

The proposed modeling system by using an advanced logarithmic companding technique is shown in figure 2 and compared with the traditional system, the logarithmic transform applied in transmitter and receiver as shown in figure 2 after Fourier in the transmitter and before Fourier in the receiver.

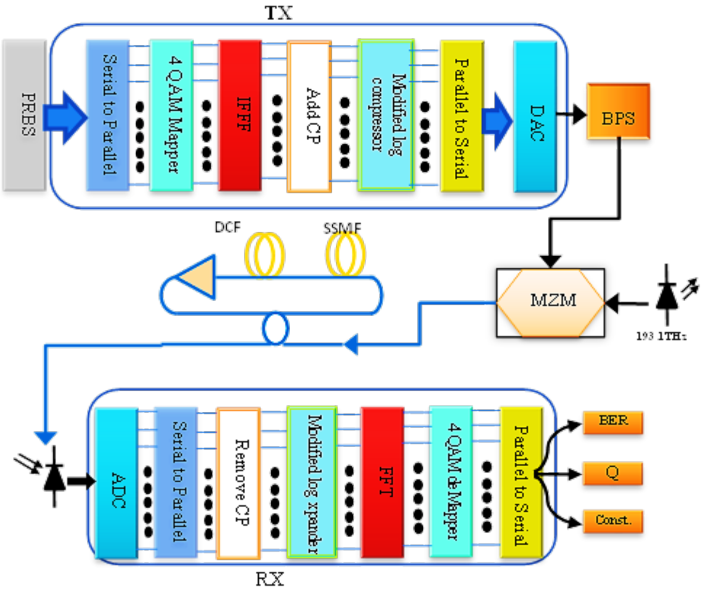


Figure 2. The proposed OOFDM System with advanced logarithmic transform.

As far as it is concerned, The transmitted digital bit streams are originated in VPI Transmission Maker software package from the Pseudorandom binary sequence (PRBS) at (2^{13} -1) and then OFDM data are decoded with and without logarithmic transform by MATLAB software package. VPI Transmission Maker does the optical modulation and optical up-down conversion and transmission link. The companded signal transmitted through the optical link with length equal to 900 Km (60 Km × 15 loops). The Fiber link contains a Standard Single Mode Fiber (SSMF), a Dispersion Compensation Fiber (DCF), and an optical amplifier. In all as feds, the SSMF is 50 km length with 0.2 dB.km^{-1} attenuation coefficient and 16 ps.nm^{-1}.km^{-1} of Chromatic Dispersion (CD) coefficient. For compensating the fibers losses, the optical amplifier has gain and noise figure of 12 dBm and 4 dBm respectively. Finally, in the receiver, the reverse operation occurs after detection to restore to the original signal.

5 Results and Discussion

The coded OFDM signal (real and imaginary part of companded OFDM) is up-converted at 7.5 GHz an Intermediate Frequency (IF) as shown in figure 3.

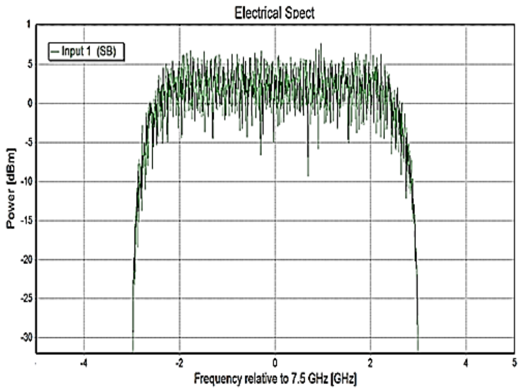


Figure 3. Electrical Spectrum of proposed logarithmic OOFDM at IF=7.5 GHz.

The output spectrum power from MZM contains an optical carrier that is targeted at 193.1 THz in this case and two side-bands targeted at 7.5 GHz of the optical OFDM signal as shown in figure 4.

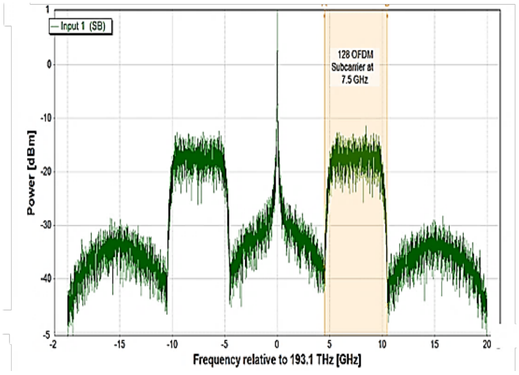


Figure 4. The spectrum of the proposed advanced logarithmic OOFDM.

Likewise, an optical filter is then responsible for suppressing the lower sideband by 18 GHz bandpass optical filter, the filtered advanced OOFDM single shows in figure 5.

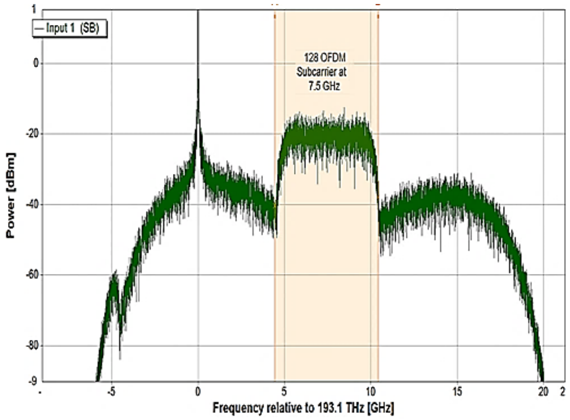


Figure 5. The spectrum of the proposed advanced logarithmic OOFDM after Filtering.

Figure 6 summarizes the performance of various degree (y) with the PAPR of OFDM companded signals. The PAPR values for degrees larger than 0.2 are almost the same as that for less than 0.15. Moreover, the minimum PAPR must be obtained at this range where 0.2 ≤ y ≤ 0.15 due to these values add more compression at the transmitter's side system.

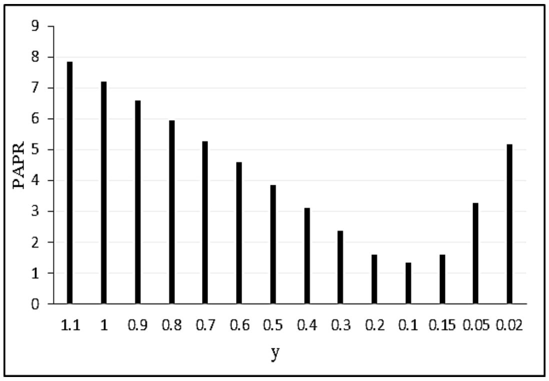


Figure 6. PAPR reduction vs control degree y.

The PAPR performance of the OOFDM transmitter based on original logarithmic technique is significantly improved. The PAPR with typical logarithmic technique is reduced while the controlling factor (k) is increased as shown in figure 7.

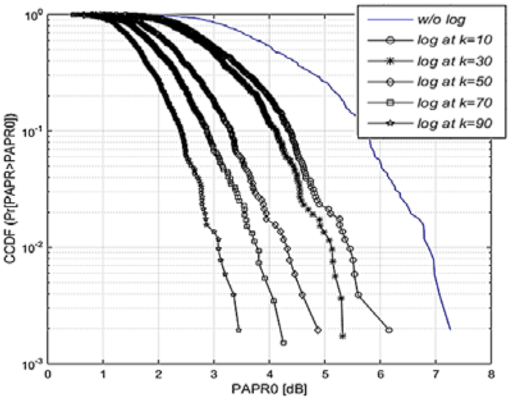
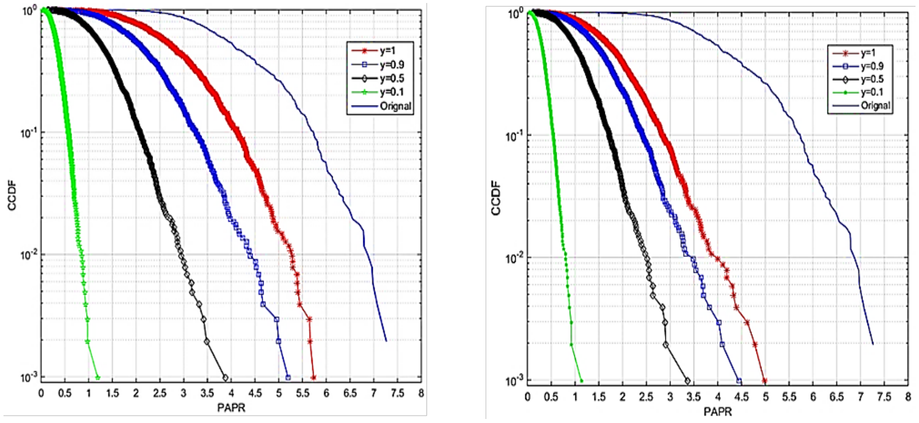


Figure 7. CCDF plot of typical logarithmic PAPR reduction technique

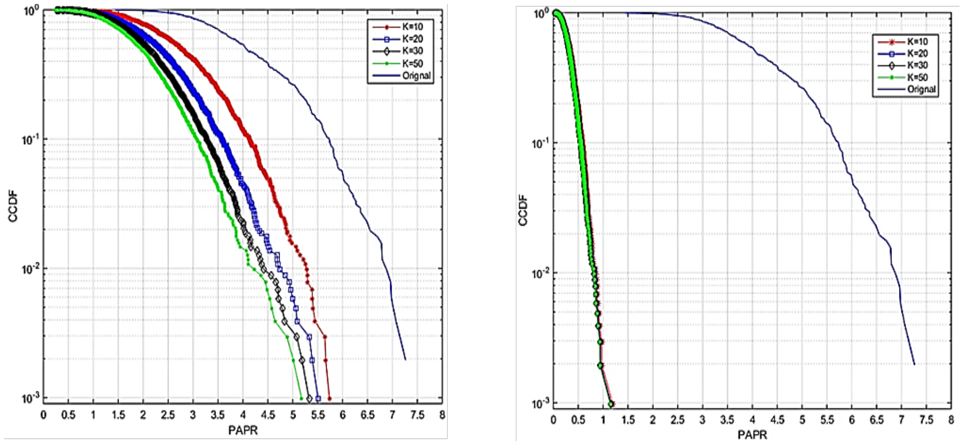
The PAPR reduction based on an advanced logarithmic technique depends on the two controls k and y as shown in figure 8 and figure 9. The CCDF is often used to display the PAPR probability distributions of OFDM transmissions. CCDF allow the probability of PAPR exceeding any given value of PAPR to be determined. From figure 8(a) and 8(b) it is seen that the PAPR of the companded OFDM signal is much lower than the un-companded signal. Companding the 128-subcarrier signal with k=30 and 50, gives the PAPR as 1.5 dB for k=30, which increases to 2.5 dB for k=50 where y = 1. It may be noted that the reduction in PAPR is smaller for higher values of k.



(a) (b)

Figure 8. CCDF plot vs PAPR based on advanced logarithmic technique for different control degree y where (a) k= 30 (b) k=50

Figure 9 shows the CCDF performance with companding degree y = 1 and 0.1, it can be noted that y = 0.1 provides a slightly better performance than y = 1 for all values of k.

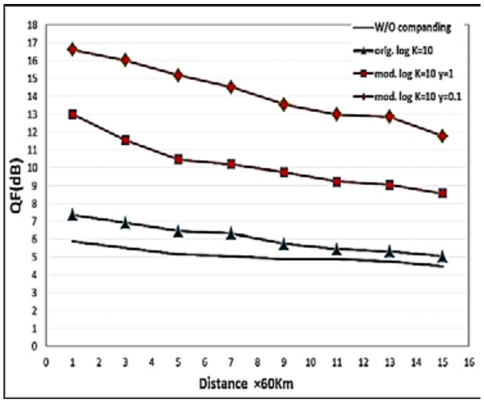


(a) (b)

Figure 9. CCDF plot vs PAPR based on advanced logarithmic for different control factor k where (a) y = 1 (b) y = 0.1

After the fiber length at 900 Km (60 Km × 15) with 4 QAM modulation formats The Quality Factor (QF), BER and constellation diagram at the receiver end is shown in figure 10, figure 11 and figure 12 respectively. Moreover, the degradation observed in the original system without companded technique used is shown in figure 10, figure 11 plotted without a marker, the measured QF is 4.8 dBm, BER is 1.58×10^{-3} and non-uniform constellation diagram in figure 12(a). This degradation is considered the maximum because of the increase in the chromatic dispersion when increasing the transmission length.

Mostly, the proposed advanced logarithmic transform that plotted with marker red color clearly improves the OOFDM system performance that is described as QF, BER, and constellation. From figure 10, figure 11 and figure 12, two values of y at 1 and 0.1 are proposed, the value of y = 0.1 is totally enhancing the all system in terms of QF at 12 dBm and BER at 4.8×10^{-5} than that is using y = 1 which QF = 8.8 dBm and BER = 4.96×10^{-4} at distance link = 900 km



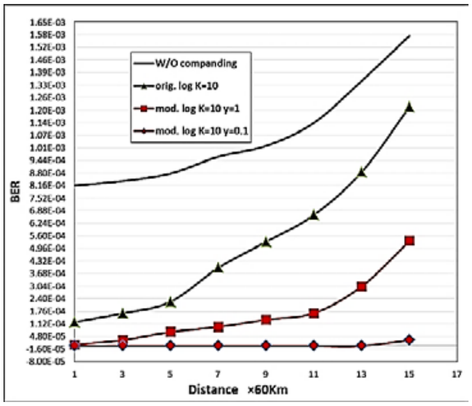
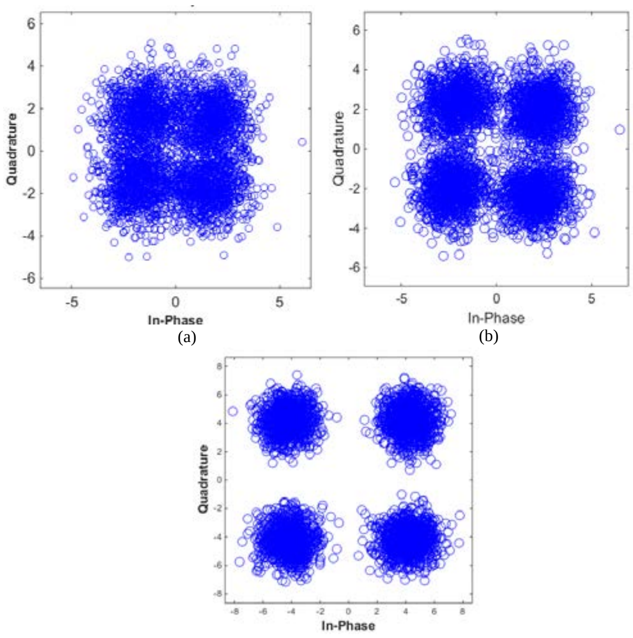


Figure 10. QF vs Transmission length

Figure 11. BER vs Transmission Length

Clearly as shown in figure 12, companding with y = 1 provides a degraded constellation diagram, however, with y = 0.1 a significant improvement in constellation performance over without companded or companded at y = 1.



(a)

(b)

(c)

Figure 12. Constellation diagram of advanced OOFDM system at the receiver end after 900 Km at (a) original without companded (b) Proposed companded at y=1 and k=10 (c) Proposed companded at y=0.1 and k=10.

6 Conclusion

The proposed IM/DD OOFDM transmitter system based on an advanced logarithmic companding technique has been implemented in order to reduce the PAPR of OOFDM systems implemented based on the sequence of operation shown in figure 2. Moreover, the results show the y = 0.1 is a best one to reduce the PAPR system performance instead of any value of k. All results of QF, BER and constellation have been presenting by combination VPI Transmission Maker with MATLAB software package. In future, the performance of transceiver systems can be completely more improved by proposing a new value for k and y with combination another technique.

REFERENCES

[1] R. v. Nee and R. Prasad, *OFDM for Wireless Multimedia Communications*: Artech House, Inc., 2000.

[2] J. Akhtman, B. Z. Bobrovsky, and L. Hanzo, "Peak-to-average power ratio reduction for OFDM modems," in *The 57th IEEE Semiannual Vehicular Technology Conference, 2003. VTC 2003-Spring.*, vol. 2, pp. 1188-1192 vol.2, 2003.

[3] X. Zhang, P. Liu, J. Liu, and S. Liu, "Advanced A-law employing nonlinear distortion reduction in DCO-OFDM systems," in *2015 IEEE/CIC International Conference on Communications in China - Workshops (CIC/ICCC)*, pp. 184-188, 2015.

[4] J. Zhou and Y. Qiao, "Low-PAPR Asymmetrically Clipped Optical OFDM for Intensity-Modulation/Direct-Detection Systems," *IEEE Photonics Journal,* vol. 7, no. 3, pp. 1-8, 2015.

[5] V. Cuteanu and A. Isar, "PAPR reduction of OFDM signals using hybrid clipping-companding scheme with sigmoid functions," in *2011 International Conference on Applied Electronics*, pp. 1-4, 2011.

[6] A. N. D. Andrea, V. Lottici, and R. Reggiannini, "Nonlinear predistortion of OFDM signals over frequency-selective fading channels," *IEEE Transactions on Communications,* vol. 49, no. 5, pp. 837-843, 2001.

[7] J. Armstrong, "Peak-to-average power reduction for OFDM by repeated clipping and frequency domain filtering," *Electronics Letters,* vol. 38, no. 5, pp. 246-247, 2002.

[8] T. Jiang and Y. Wu, *An Overview: Peak-to-Average Power Ratio Reduction Techniques for OFDM Signals* vol. 54, 2008.

[9] L. Wang and J. Liu, "PAPR Reduction of OFDM Signals by PTS With Grouping and Recursive Phase Weighting Methods," *IEEE Transactions on Broadcasting,* vol. 57, no. 2, pp. 299-306, 2011.

[10] S. Y. L. Goff, B. K. Khoo, C. C. Tsimenidis, and B. S. Sharif, "A novel selected mapping technique for PAPR reduction in OFDM systems," *IEEE Transactions on Communications,* vol. 56, no. 11, pp. 1775-1779, 2008.

[11] Y. Wang, J. Ge, L. Wang, J. Li, and B. Ai, "Nonlinear Companding Transform for Reduction of Peak-to-Average Power Ratio in OFDM Systems," *IEEE Transactions on Broadcasting,* vol. 59, no. 2, pp. 369-375, 2013.

[12] K. Anoh, B. Adebisi, K. M. Rabie, and C. Tanriover, "Root-Based Nonlinear Companding Technique for Reducing PAPR of Precoded OFDM Signals," *IEEE Access,* vol. 6, pp. 4618-4629, 2018.

[13] S. Azou, S. Bejan, P. Morel, and A. Sharaiha, "A comparative study of nonlinear companding schemes for CO-OFDM transmissions," in *2014 13th International Conference on Optical Communications and Networks (ICOCN)*, pp. 1-4, 2014.

[14] N. Jacklin and Z. Ding, "A Linear Programming Based Tone Injection Algorithm for PAPR Reduction of OFDM and Linearly Precoded Systems," *IEEE Transactions on Circuits and Systems I: Regular Papers,* vol. 60, no. 7, pp. 1937-1945, 2013.

[15] A. Ivanov and D. Lakontsev, "Selective tone reservation for PAPR reduction in wireless communication systems," in *2017 IEEE International Workshop on Signal Processing Systems (SiPS)*, pp. 1-6, 2017.

[16] W. Wang, M. Hu, Y. Li, and H. Zhang, "A Low-Complexity Tone Injection Scheme Based on Distortion Signals for PAPR Reduction in OFDM Systems," *IEEE Transactions on Broadcasting,* vol. 62, no. 4, pp. 948-956, 2016.