IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 19, Issue 6, Ver. II (Nov.- Dec. 2017), PP 07-12
www.iosrjournals.org

Formulating a Simple Theoretical Framework for Data Flow in SCADA System

Alade A. A [1], Ajayi O. B [2] , Okolie S. O.[3], Alao D. O.[4]
1(Department of Computer Science, Babcock University, Nigeria)
[2](Department of Computer Science, Federal University of Agriculture, Abeokuta, Nigeria)
[3](Department of Computer Science, Babcock University, Nigeria)
[4](Department of Computer Science, Babcock University, Nigeria)

Abstract: *Studies in Supervisory Control and Data Acquisition (SCADA) System revolves around SCADA System security due to the impact of attacks on the critical infrastructure such as national electricity network, complex oil/gas network, power station, etc. which SCADA System is expected to control and manage. However, in the paper, attention is directed towards "formulation of theoretical framework for SCADA System". This brings to light the basic understanding of the protocol transmission procedures and rules in SCADA System that set it apart from other industrial Control System (ICS). As a consequence of SCADA System unbalanced transmission in which only the master station initiates the communication while the slaves only respond (Fig.2), a theoretical framework that is applicable in a network with peer to peer entities may not apply for SCADA System. Using SCADA System multidrop topology, we derived a master/slave graph equations and arrived at its dynamic behavior pattern applying Finite State Machine (FSM) and finally produced a probabilistic Graphical Model using Markov Chain.*

Keywords - *attacks, state machine, graphical, protocol, master, slaves, topology*

Date of Submission: 07-11-2017 Date of acceptance: 18-11-2017

I. Introduction

Communication philosophy generally used in computer networks can be one of the two options: a contention or a polled approach. In SCADA System where there is the MTU monitoring and controlling the field equipment such as the RTU, PLC and Intelligent Electronic Devices (IED), a polled (Master to Slave) approach usually applies. The approach is used in the four common SCADA topologies depicted in Fig. 1A (direct or one-on-one topology), 1B (multidrop topology), 1C (hierarchical topology and 1D (multiple master topology). In each of these topologies, the master is in full control of the information flow as it makes repetitive and regular requests for data which is to be transferred from and to several slaves terminal. The MTU is the master in respect of a SCADA System, while the RTUs and IEDs are the slaves.

A master/slaves communication approach is basically a half-duplex approach where the slave does not initiate any transaction, it only responds to the master's request. Depending upon the polling algorithms, the master node may retry non responding slave node up to three or more times before switching to another and it may later retry the non responding slave node. This is referred to as unbalanced transmission procedure [1].
Some of the advantages of this approach are fast detection of a failed link between a slave and the master, software reliability and simplicity and predictable data throughput as there is no possibility of collisions on the network [2].

II. Methodology

Several online journals and text books on this subject were examined to have in-depth knowledge of the subject and be able to determine the state of research in the area. Using search engines, additional information on the subject was obtained. The results of the exercise follow:

2.1 SCADA Systems Data Classification

In SCADA systems, transmitted data comprises both the real time Input/Output (I/O) and the static data such as configuration data and control programmes. The SCADA System data are classified into two based on: 1) Time Requirement and 2) data generation mechanism [3].

Time requirement: This may be:

a **Time-critical** data such as alarm signal, interlock signal and controller I/O signal are time rigorous. Delay time of even a millisecond is prohibited. Late arriving data are usually discarded as only the latest data are useful.

b **Non-critical** data (message) – these are also called static data, huge and seldom changes. Examples are nodes' initialization information, configuration and program data. As time requirement is less rigorous, residual bandwidth can be used for its transmission [4].

1. Data generation mechanism

a **Periodic data** – An example of the several periodic data in ICS is sensors acquisition data. They are characterized by cyclical events, fixed transmission information, predictability and high frequency of occurrence.

b **Statistical Data**: examples are numeric control program and database management messages. They are characterized by unpredictability, random communication, lowest priority and criticality. They are also known as aperiodic data [3].

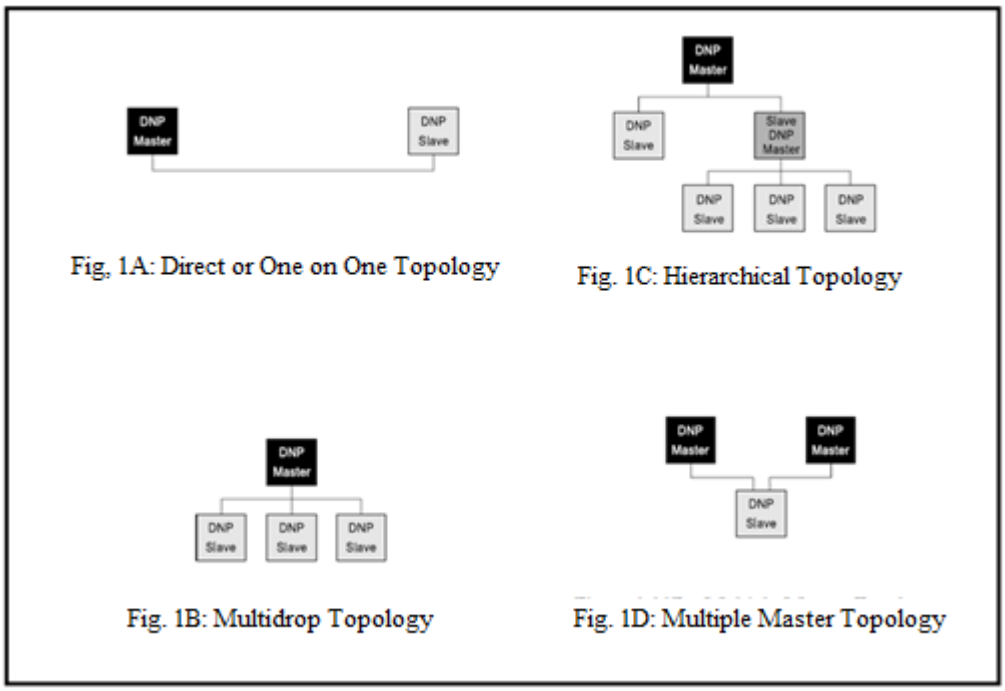

Fig. 1: SCADA System Topology
Source: Clarke, Reynders and Wright [2]

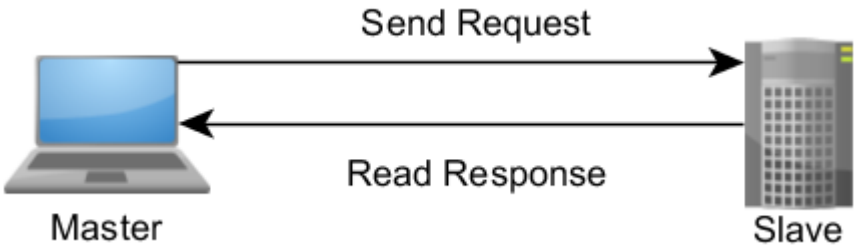

Fig.2: Master/Slave – Request/Response Relationship
Source: Hans-Petter [5]

2.2 SCADA Protocols Transmission Procedures and Rules

Transmission procedure may be one of the two in ICS protocols: balanced transmission where there is peer to peer communication between two nodes or unbalanced transmission as in SCADA System where only the master station initiates the communication while the slaves (outside stations) only responds. The master station is the primary node that controls the traffic of data flow by sequentially polling the outside stations. The slaves can only transmit after being polled. The transmission rules are summarised in the three services defined by the protocol as follows:

Send/No Rely Service: This applies in message broadcast to all the outside stations with no reply expected, delivery is not neither confirmed or guaranteed as theire is no error checking functions involved.

Send/Confirm Service: Send/confirm link procedures are used to transmit data (parameters, command, etc) in the control direction. Confirmation is expected and a resend/confirm message follows if the reply times out.
Request/respond service: In unbalanced transmission systems, Data acquisition is monitored by means of the request/respond link. Such data may be events, monitored information, command confirmation, etc. The master station (controlling station request for data and the slave station transmits the RESPOND frame if available else negative response follows [1].

2.3. Theoretical Framework for SCADA Systems

Four types of SCADA System topologies were identified (Fig. 1). Of these, the multidrop SCADA System Master/Slave topology is the most widely used. It has one Master station communicating directly with several Slave stations (Fig. 1B). This model of SCADA System Master/Slave topology, extended to cover many Slave stations is our reference in this section. In the abstraction of fig. 1B, a small fully filled circle denotes the master node, labeled M while several slave nodes are denoted with small unfilled circles.

Solid lines connect each of the Slave nodes to the Master nodes. The physical distance so presented can be as short as just one kilometer or as long as hundreds of kilometers, especially for a SCADA System that monitors and controls a nations electricity networks with several field stations which are geographically spread (Fig. 3).

We present SCADA System formally here as formal specification is often used to avoid ambiguity that may arise from mere description [6], [7]. Three approaches are adopted: Graph representation, inspection of the dynamic behavior of SCADA System Master/Slave Protocols using Finite State Machine Model and derivation of its Probabilistic Graphical Models.

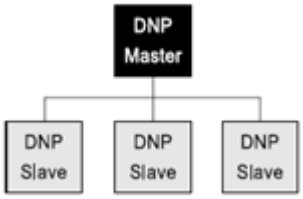

Fig. 1B: Multidrop Topology

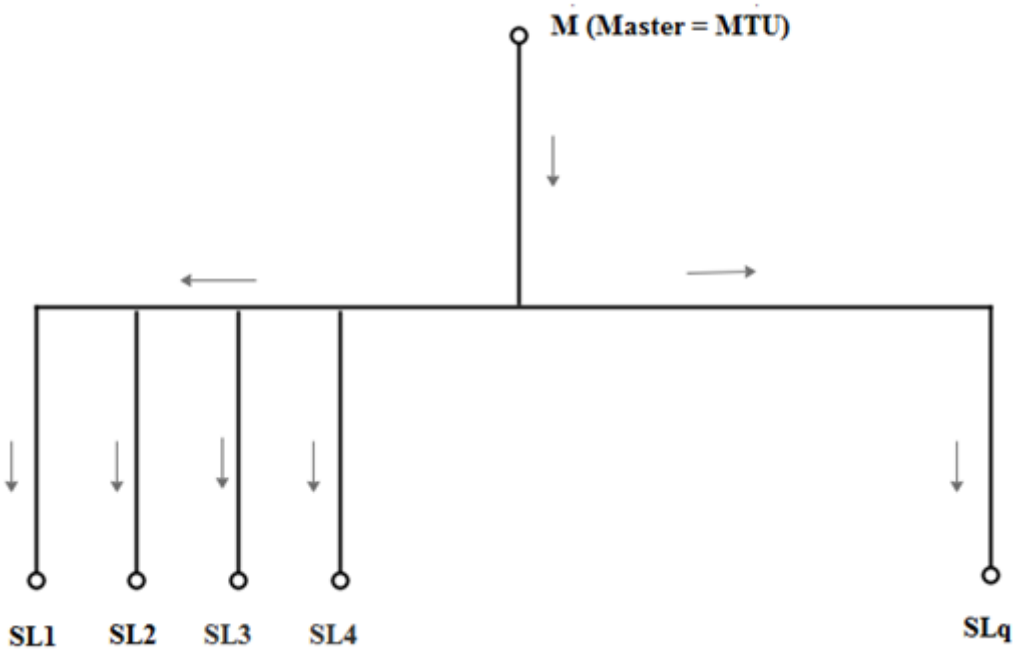

Fig. 3: SCADA System Multidrop Topology

2.3.1. Graph of SCADA System Master/Slave Topology

"A graph is a pair of set (V, E), where V is the set of vertices and E is the set of edges formed by the pair" [8], [9].

Fig. 3 is redrawn as in fig. 4 to reflect the actual physical connection where it is evident that the master (Master Terminal Unit - MTU) has direct link with each of the slaves (Remote Terminal Units – RTUs). As discussed earlier, the link may be bounded medium such as optic fibre or unbounded medium such as radio or satellite.

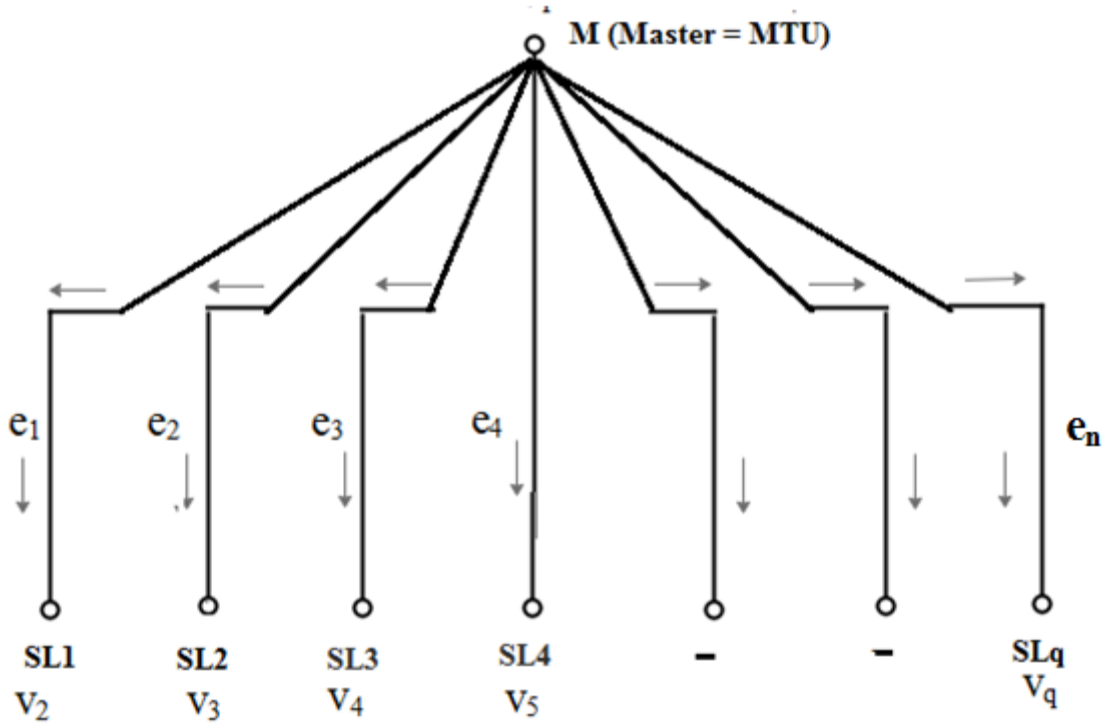

Fig. 4: Equivalent SCADA System Multidrop Topology

Consider the graph, G = (V, E)
The vertices, V in fig. 4 is a set {m, sl_1, sl_2, sl_3, sl_4, … sl_q}
This is generally written as

$$V = \{v_1, v_2, v_3, v_4, \ldots v_q\}$$
$$E = \{e_1, e_2, e_3, e_4, \ldots e_n\}$$

In this configuration, q= n+1, hence it is a star topology.

2.3.2. Dynamic Behaviour of SCADA System Master/Slave Protocols Using Finite State Machine Model

With a model of Finite State Machine (FSM), the dynamic behavior of Systems can be easily described[10]. The behavior of a master/Slave SCADA System using the Enhanced Protocol Architecture (EPA) as earlier described, having 3 layers mainly (Application, Link and Physical layers) can be modeled with Finite State Machine.

A Finite State Machine (FSM) is described with 5 – tuple ($\sum$, S, S_0, δ, τ)
where

$\sum$ - a finite set of inputs (events)
S - a finite set of states
S_0 - an initial state which is a member of S
δ – the state transition function, $\delta : \sum X S \rightarrow S$
τ – a non empty set of transition

The link layer object cycle has six states:

Disconnect	S_0 (initial State)
Remote Link	S_1
Link reset	S_2
Link free	S_3
Request/response	S_4
Send/confirm	S_5

FU, Chen and Kui [1].

δ – The state transition function is in this wise the protocol implementation program at the Link layer of EPA that decides transition from one state to another.

τ - a non empty set of transmission in this case is six. $\tau = \{t_1, t_2, t_4, t_5, t_6\}$. This is so because the starting state is also the end state.

Using the states notation, the six stations are labeled as in the states chart diagram (Fig. 5).

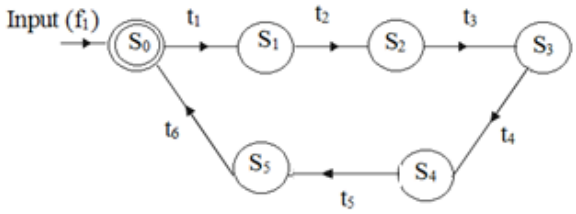

Fig. 5: The SCADA System Protocol States

2.3.3. Probabilistic Graphical Models

Probabilistic Graphic Model (PGM) employs graph theory as foundation for encoding a complex distribution compactly. Markov Chain is one of them being used extensively [11]. Markov Chain probability network for SCADA System protocol's state is derived here.

Given the states of the MTU when the input is frame f_i at time t_i we predict using probability model, states of the frame f_{i+1} at time t_{i+1} applying the first order Markov Chain.

Let $\acute{S}$ represent the state of the SCADA System. The following probability can be computed:

$$P(s'_n | s'_{n-1}, s'_{n-2}, s'_{n-3}, s'_{n-4}, s'_{n-5}, s'_{n-6}, \ldots, s'_1)$$

Applying Markov Assumption

$$P(s'_n | s'_{n-1}, s'_{n-2}, s'_{n-3}, s'_{n-4}, s'_{n-5}, s'_{n-6}, \ldots, s'_1) \approx P(s'_n | s_{n-1})$$

i.e. $$P(s'_1, s'_2, s'_3, s'_4, s'_5, s'_6, \ldots, s'_n) \approx \prod_1^n P(\acute{s}_i | \acute{s}_{i-1})$$

This means that the knowledge of the MTU's states presently with frame f_i can enable us predict the next states of frame f_{i+1} applying the knowledge of f_{i-1}, f_{i-2}, etc[12], [13].

For our case, states $s' = s_0, s_1, s_2, s_3, s_4, s_5$ (6 states), Transition matrix P[6 x 6] is given in table 1.

Table 1: Transition Matrix for 6 states MTU communication

		Frame f_{i+1}					
Frame f_i		s_0	s_1	s_2	s_3	s_4	s_5
	s_0	p_{00}	p_{01}	p_{02}	p_{03}	p_{04}	p_{05}
	s_1	p_{10}	p_{11}	p_{12}	p_{13}	P_{14}	P_{15}
	s_2	p_{20}	p_{21}	p_{22}	p_{23}	p_{24}	p_{25}
	s_3	p_{30}	p_{31}	p_{32}	p_{33}	p_{34}	P_{35}
	s_4	p_{40}	p_{41}	p_{42}	p_{43}	p_{44}	P_{45}
	s_5	P_{50}	p_{51}	P_{52}	P_{53}	P_{54}	P_5

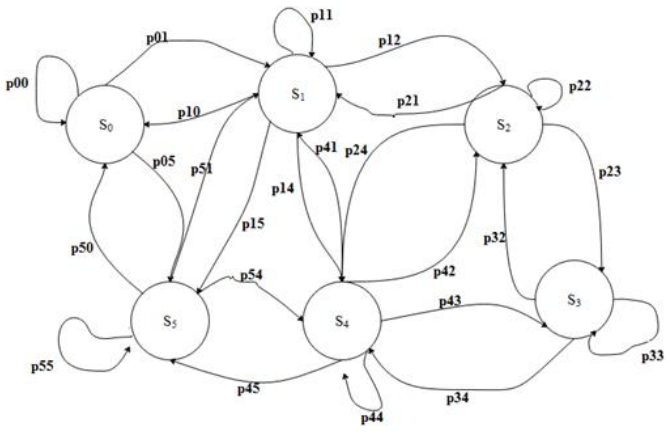

Figure 6: Markov Chain Probability Network for SCADA System Protocol States

III. Discussions

Theoretical framework for SCADA System protocols is presented using three approaches: graph, Finite State Machine (FSM) and Probabilistic Graphic Model (PGM) using first order Markov Chain. The graph equations obtained can apply to any positive number of vertices and edges. The graph arrows depict only the direction of data flow from the master during polling while the converse direction during response is assumed. Depending on the adopted number of states the dynamic Markov chain derived in this work may vary.

IV. Conclusion

The theoretical framework presented in this paper will aid objective decision and inference on issue of the Supervisory Control and Data Acquisition network behavior. In this work, a widely applied Master/Slaves multidrop SCADA System topology has been considered. Other open area of researches are the multiple master topology where there are at least two master serving several slaves and the hierarchical model where a master serves the sub-master and submaster in turn serves other slaves.

Acknowledgements

The authors wish to thank the Management of the Computer science department, Babcock University, Nigeria for their supports in the course if this research.

References

[1] Q. FU, J. Chen and Z. KIU, Implementation of Telecontrol Protocols in SCADA systems based on FSM, *Proceedings of the 8th International Power Engineering Conference*, China, 2007.
[2] G. Clarke and D. Reynders, *Practical modern scada protocols: DNP3, 60870.5 and related systems (* Burlington, MA: Newnes, Elsevier, 2004).
[3] H. Gao and W. Tong, Analysis and evaluation of fieldbus communication and protocol static characteristic, *Association for Computing Machinery* 978-1-60558-088-3, 2008.
[4] S.H. Hong, Bandwidth allocation scheme for cyclic-service fieldbus networks, *IEEE/ASME Transactions on Mechatronics*, *6(2), 2001*, 197-204.
[5] H. Hans-Petter, Modbus Overview, 2013. Available at http://home.hit.no/~hansha/documents/industrial_it/resources/resources /modbus/Modbus%20Overview.pdf.
[6] F. M¨uffke, *A better way to design communication protocols,* doctoral diss., University of Bristol, Bristol., 2004. Available at http://www.cs.bris.ac.uk/Publications/pub_master.jsp?type=116.
[7] B.F. Adiego, *Bringing Automated Formal Verification to PLC Development*, doctoral diss., University of Oviendo, Spain, 2014.
[8] M. Sipser, *Introduction to the theory of computation (*Boston, Massachusetts: Thomas Learning, Inc, 2006).
[9] K. Ruohonen, *Graph Theory, 2003.* Available at http://math.tut.fi/~ruohonen/GT_English.pdf,.
[10] G. Booch, J. Rumbaugh and I. Jacobson, *Unified modeling language user guide (*Massachusetts, USA: Addison-Wesley Longman Inc, 2001).
[11] D. Koller and N. Friedman, *Probabilistic graphical models principles and techniques (* Massachusetts, USA: MIT Press, 2009).
12] E. Fosler-Lussier, Markov models and hidden Markov models: A brief tutorial", 1998. Available at http://di.ub i.pt/~jpaulo/competence/tutorials/hmm-tutorial-1.pdf/
[13] S.M. Ross, *Introduction to Probability Models (* Burlington, MA : Academic Press, 2010).

Alade A. A Formulating a Simple Theoretical Framework for Data Flow in SCADA System." IOSR Journal of Computer Engineering (IOSR-JCE) , vol. 19, no. 6, 2017, pp. 07-12.