International Journal of Engineering Innovation & Research
Volume 2, Issue 2, ISSN: 2277 – 5668

New Genetic Algorithm Based Intrusion Detection System for SCADA

Aarcha Anoop
PG Scholar,
Department of ECE,
Younus College of Engineering and Technology,
Kollam - 691 005, Kerala, India.
Email: aarchamohan1@gmail.com

Sreeja M. S.
Assistant Professor,
Department of ECE,
Younus College of Engineering and Technology,
Kollam - 691 005, Kerala, India.
Email: sreejams2002@gmail.com

Abstract **- Securing SCADA systems is a critical aspect of industrial systems. Industrial systems have installations which actively using the public network in order to provide new features and services which make the system unsecured .By introducing a filtering system ,we can analyse the critical state of the system which can be monitored and secure SCADA network protocols. But in this approach, there is no mathematical method for calculating filter parameters for DDOS, R2L, U2R attacks. In this paper, we present a new genetic algorithm based approach for calculating those parameters to make the system more secure.**

Keywords **- SCADA, Intrusion Detection, Genetic Algorithm.**

I. INTRODUCTION

SCADA[1] is a type of industrial control system that are controlled by a computer which monitor and control large distances .These processes include industrial, infrastructure, and facility-based processes. Industrial processes include those of manufacturing, production, power generation, fabrication, and refining, and may run in continuous, batch, repetitive, or discrete modes. Industrial processes that exist in the physical world. SCADA systems[2] are different from other ICS systems by being large scale processes that can include multiple sites, Infrastructure processes may be public or private, and include water treatment and distribution, wastewater collection and treatment, oil and gas pipelines, electrical power transmission and distribution, wind farms, civil defence siren systems, and large communication systems.

Facility processes occur both in public facilities and private ones, including buildings, airports, ships, and space stations. They monitor and control heating, ventilation, and air conditioning systems (HVAC), access, and energy consumption.

The term SCADA usually refers to centralized systems which monitor and control entire sites, or complexes of systems spread out over large areas (anything from an industrial plant to a nation). Most control actions are performed automatically by RTUs or by PLCs. Host control functions are usually restricted to basic overriding or supervisory level intervention. For example, a PLC may control the flow of cooling water through part of an industrial process, but the SCADA system may allow operators to change the set points for the flow, and enable alarm conditions, such as loss of flow and high temperature, to be displayed and recorded. The feedback control loop passes through the RTU or PLC, while the SCADA system monitors the overall performance of the loop.

Data acquisition begins at the RTU or PLC level and includes meter readings and equipment status reports that are communicated to SCADA as required. Data is then compiled and formatted in such a way that a control room operator using the HMI can make supervisory decisions to adjust or override normal RTU (PLC) controls. Data may also be fed to an Historian, often built on a commodity Database Management System, to allow trending and other analytical auditing.

SCADA systems typically implement a distributed database, commonly referred to as a tag database, which contains data elements called tags or points [15][16]. A point represents a single input or output value monitored or controlled by the system. Points can be either "hard" or "soft". A hard point represents an actual input or output within the system, while a soft point results from logic and math operations applied to other points. (Most implementations conceptually remove the distinction by making every property a "soft" point expression, which may, in the simplest case, equal a single hard point.) Points are normally stored as value-time stamp pairs: a value and the timestamp when it was recorded or calculated. A series of value-timestamp pairs gives the history of that point. It is also common to store additional metadata with tags, such as the path to a field device or PLC register, design time comments, and alarm information.

SCADA systems are significantly important systems used in national infrastructures such as electric grids, water supplies and pipelines. However, SCADA systems may have security vulnerabilities, so the systems should be evaluated to identify risks and solutions implemented to mitigate those risks. The following are the security vulnerabilities [2] of SCADA systems:

- Unauthorized Command Execution
- SCADA DOS
- Man-in-the-Middle (MITM)
- Replay attacks

Intrusion detection [4] is the art of detecting inappropriate, inaccurate, or anomalous activity. Detecting intrusion is one of high priority and challenging task for network administrators and security professionals [6]. More sophisticated security tools means attackers come up with newer and more advanced penetration methods to

defeat the installed security system. Thus, there is a need to safeguard the networks from known vulnerabilities and at the same time take steps to detect new and unseen, but possible, system abuses by developing more reliable and efficient intrusion detection systems. Any intrusion detection system has some inherent requirements. Its prime purpose is to detect as many attacks as possible with minimum number of false alarm. However, an accurate system that cannot handle large amount of network traffic and is slow in decision making will not fulfil the purpose of intrusion detection system.

Intrusion detection systems are classified as network based, host based, or application based depending on their mode of deployment and data used for analysis. Additionally, intrusion detection system can also be classified as signature based or anomaly based depending upon the attack detection method[5]. The signature based systems are learned by extracting specific patterns from previously known attacks while the anomaly based systems learn from normal data collected when there is no anomalous activity. Another approach is to consider both the normal and known anomalous patterns for training a system and then performing classification on the test data. Such a system is known as the Hybrid System. Hybrid system incorporates the advantages of both the signature based and the anomaly based system. It can be very efficient and, subject to the classification method used and can be used to label unseen or new instances as they assign one of the known classes to every test instances. Intrusion detection functions include:

- Monitoring and analysing both user and system activities
- Analysing system configurations and vulnerabilities
- Assessing system and file integrity
- Ability to recognize patterns typical of attacks
- Analysis of abnormal activity patterns
- Tracking user policy violations

In order to overcome challenges occurred in SCADA security, we can use a Layer-based Intrusion Detection System[1]. The LIDS draws its motivation from Airport Security model, where a number of security checks are performed one after the other. This model is based on ensuring availability, confidentiality, and integrity of data or service over the network. The goal of using a layered model is to reduce computation and the overall time required to detect anomalous events [5]. The time required to detect an intrusive event is significant and can be reduced by eliminating the communication overhead among different of relevant features. Feature selection is significant for Layered Approach [4].

In order to make the layers independent, some features may be present in more than one layer. The layers essentially act as filters that block layers. This can be achieved by making the layers autonomous and self-sufficient to block an attack without the need of a central decision-maker. Every layer in the LIDS framework is trained separately and then deployed sequentially. We define four layers that correspond to them four attack groups mentioned in the data set. They are Probe layer,

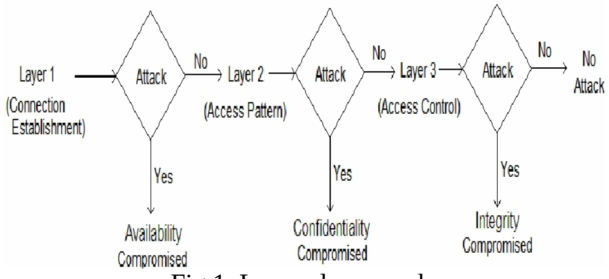

Fig.1. Layered approach

DOS layer, R2L layer, and U2R layer. Each layer is then separately trained with a small set any anomalous connection, thereby eliminating the need of further processing at subsequent layers enabling quick response to intrusion. The effect of such a sequence of layers is that the anomalous events are identified and blocked as soon as they are detected. In layered approach, there is no mathematical for calculating filter parameters for DOS, R2L, and U2R attack. In this paper, we present a genetic algorithm based approach for calculating those parameters. Section I gives the introduction of SCADA systems and Intrusion detection systems .Section II describes the Genetic Algorithm and section III explains the new genetic algorithm technique for SCADA. Section IV gives Testing and Results.

II. GENETIC ALGORITHM

Genetic Algorithms [25] (GA) are direct, parallel, stochastic method for global search and optimization, which imitates the evolution of the living beings, described by Charles Darwin. GA is part of the group of Evolutionary Algorithms (EA). The evolutionary algorithms use the three main principles of the natural evolution: reproduction, natural selection and diversity of the species, maintained by the differences of each generation with the previous. The selection principle is applied by using a criterion, giving an evaluation for the individual with respect to the desired solution. The best-suited individuals create the next generation. The large variety of problems in the engineering sphere, as well as in other fields, requires the usage of algorithms from different type, with different characteristics and settings. The following are the elements used in genetic algorithm.

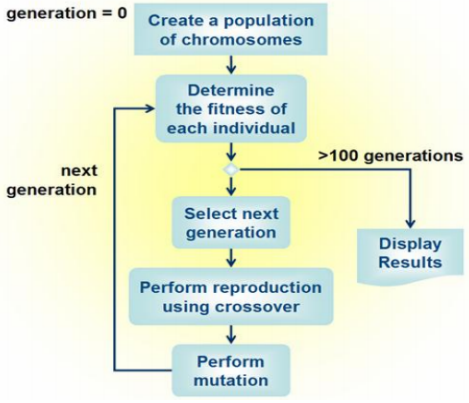

Fig.2. New Generation Algorithm

A. Chromosomes

During the division process of the human cells, the chromatin (contained in the nucleus and built from DNA (deoxyribonucleic acid), proteins and RNA (ribonucleic acid)) become shorter and thicker and forms spiral strings – chromosomes. Chromosomes are the genes that carry the inherited cell information. Every gene codes particular protein and is independent factor of the genetic information, which determines the appearance of different peculiarities. For the genetic algorithms, the chromosomes represent set of genes, which code the independent variables. Every chromosome represents a solution of the given problem. Individual and vector of variables will be used as other words for chromosomes. From other hand, the genes could be Boolean, integers, floating point or string variables, as well as any combination of the above.

A set of different chromosomes (individuals) forms a generation. By means of evolutionary operators, like selection, recombination and mutation an offspring population is created

B. Selection

In the nature, the selection of individuals is performed by survival of the fittest. The most available individual is adapted to the environment - the bigger are its chances to survive and create an offspring and thus transfer its genes to the next population

In EA, the selection of the best individuals is based on an evaluation of fitness function or fitness functions. Examples for such fitness function are the sum of the square error between the wanted system response and the real one; the distance of the poles of the closed-loop system to the desired poles, etc. If the optimization problem is a minimization one, than individuals with small value of the fitness function will have bigger chances for recombination and respectively for generating offspring.

C. Recombination

The first step in the reproduction process is the recombination (crossover). In it the genes of the parents are used to form an entirely new chromosome. The typical recombination for the GA is an operation requiring two parents, but schemes with more parents' area also possible. Two of the most widely used algorithms are Conventional (Scattered) Crossover and Blending (Intermediate) Crossover [24].

In this recombination type, the parents exchange the corresponding genes to form a child. The crossover can be single- or multipoint. For the recombination process, a bit mask is used. The equations describing the process are:

C1 = Mask1 & P1 + Mask2 & P2;
C2 = Mask2 & P1 + Mask1 & P2;
P1, P2 – parent's chromosomes;
C1,C2–children'schromosomes (offspring individuals);
Mask1, Mask2 – bit masks (Mask2 = NOT (Mask1)) ;
& - bit operation "AND".

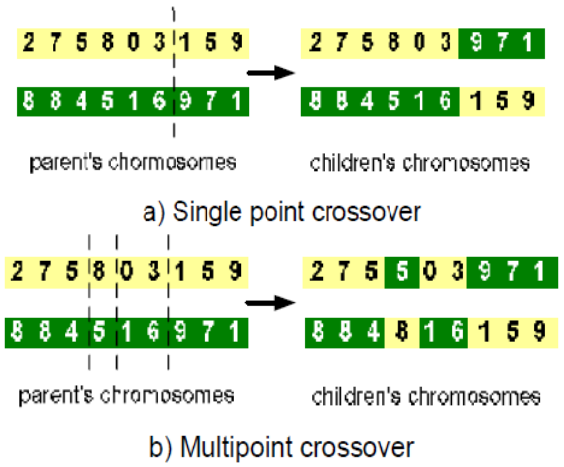

Fig.3. Crossover with bit mask

D. Mutation

The newly created genes by means of selection and crossover population can be further applied to mutation. Mutation means that some elements of the DNA are changed. The changes are caused mainly by mistakes during the copy.

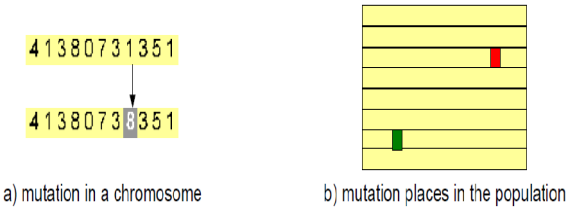

Fig.4. Mutation Process in NGA

In the terms of genetic algorithm, mutation means random change of the value of a gene in the population. Figure 4 shows mutation process in NGA .The chromosome, which gene will be changed and the gene itself are chosen by random as shown in above figure.

III. NGA FOR SACADA

Here input is the attack feedback and the output is optimal filter value for each attack. Here we consider three main factors:

- History Records(NH)
- Generation(G)
- No. of attacks(NH)

'N' chromosomes are randomly selected. Remaining chromosome are generated by the following processes. Selection Process is done by selecting two sequences randomly from input & copy best one (with highest fitness value) to the output. Recombination is done by selecting two sequences randomly from input & mixing them to produce a new one and then copy to the output. Mutation is done by selecting a sequence randomly from input and applies changes to its fitness value & copied to the output.

The Fitness Value of a chromosome is calculated by the Number of Feedback Records satisfy this chromosome minus Number of Feedback Records Reject this chromosome divided by the History Records.

IV. Testing and Results

Using dumping feature of Packet capturing library, dump packets towards a large network for 7days. A manually generated mining data supply to our software. Output shows following performance details. Then explain following diagrams and the system performance is calculated by using attack detection accuracy.

As number of rounds increase weight also increase, and finally this will tends to one. But this will increase time complexity.

No.of rounds	Probability of Failure
50	0.678
70	0.698
80	0.72
90	0.72
100	0.74
150	0.76
200	0.78
250	0.79
300	0.81

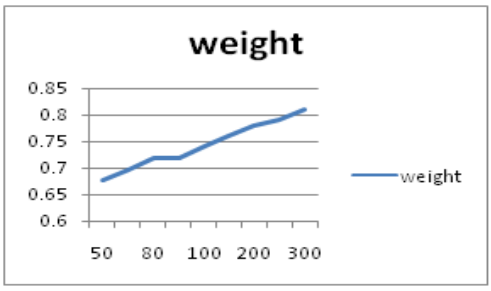

Fig.5 Number of rounds against probability

As number of rounds increases, accuracy also increases, but vary for different attack. DOS accuracy is small compare to others. This is because DOS contain different types of attack. If mining data contain all types of attack then accuracy increases.

Rounds	DOS	R2L	U2R
50	5	25	27
70	7	35	37
80	20	40	42
90	20	40	43
100	25	55	47
150	40	75	68
200	55	90	70
250	60	100	72
300	65	100	80

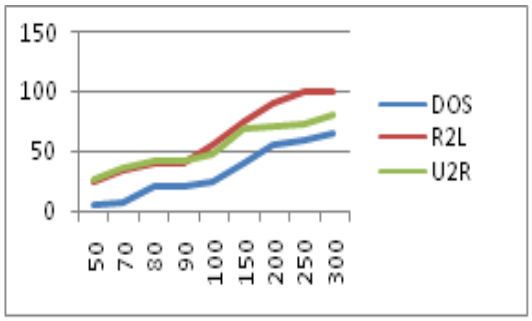

Fig.6. Rounds against different attacks

Size of Mining data	Overall Accuracy
5000	20
7000	21
10000	37
20000	45
30000	50
40000	55
50000	65
100000	80

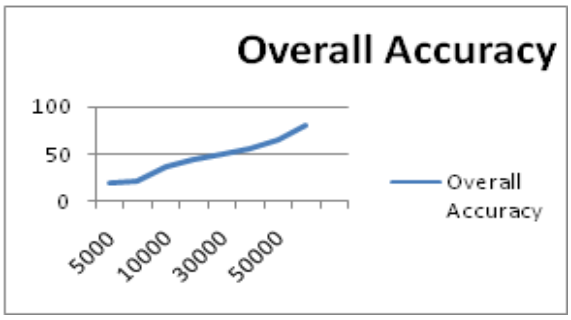

Fig.7. Overall accuracy of attacks

V. Conclusion

In this work, a layered approach for intrusion detection is proposed which is based on genetic algorithm. In this framework, the four layers correspond to four attack groups. They are probe, DOS, R2L and U2R attacks. But in this approach, there is no mathematical method for calculating filter parameters for DOS, R2L, U2R attacks. So, we present a new genetic algorithm based approach for calculating those parameters to make the system more secure .This efficiently detect R2L attack and 90 % accuracy detected.

Acknowledgment

The authors would like to thank the management, and Faculty Members, of Department of Electronics and Communication Engineering, Younus College of Engineering and Technology, for many insightful discussions and the facilities extended to us for completing the task.

REFERENCES

[1] Kapil Kumar Gupta, Baikunth Nath, and Ramamohanarao Kotagiri "Layered Approach Using Conditional Random Fields for Intrusion Detection" IEEE Transactions on dependable and secure computing, vol. 5, no. 4, October-December 2010.

[2] K.K. Gupta, B. Nath, and R. Kotagiri, "Network Security Framework," Int'l J. Computer Science and Network Security, vol. 6, no. 7B, pp. 151-157, 2006.

[3] P. Garcia Teodoro, J. Diaz Verdejo, G. Maica Fernandez, E. Vazquez, "Anomaly Based Network Intrusion Detection", Second International Conference on Communication Software and Networks, 2008.

[4] K.K. Gupta, B. Nath, and R. Kotagiri, "Conditional Random Fields for Intrusion Detection," Proc. 21st Int'l Conf. Advanced Information Networking and Applications Workshops (AINAW '07), pp. 203-208, 2007.

[5] K.K. Gupta, B. Nath, R. Kotagiri, and A. Kazi, "Attacking Confidentiality: An Agent Based Approach," Proc. IEEE Int'l Conf. Intelligence and Security Informatics (ISI '06), vol. 3975, pp. 285-296, 2006.

[6] A.E. Tombini, H. Debar, L. Me, and M. Ducasse, "A Serial Combination of Anomaly and Misuse IDSes Applied to HTTP Traffic," Proc. 20th Ann. Computer Security Applications Conf. (ACSAC '04), pp. 428-437, 2004.

[7] L. Ertoz, A. Lazarevic, E. Eilertson, P.-N. Tan, P. Dokas, V. Kumar, and J. Srivastava, "Protecting against Cyber Threats in Networked Information Systems," Proc. SPIE Battlespace Digitization and Network Centric Systems III, pp. 51-56, 2003.

[8] C. Sutton and A. McCallum, "An Introduction to Conditional Random Fields for Relational Learning," Introduction to Statistical Relational Learning, 2006.

[9] Denning D.E, "An Intrusion-Detection Model," IEEE Transactions on Software Engineering" 13(2), 222–232 (1987)

[10] G. Dondossola, M. Masera, I. Nai Fovino, and J. Szanto, "Effects of intentional threats to power substation control systems," *Proc. IJCIS*,vol. 4, no. 1/2, pp. 129–143, 2008.

[11] I. Nai Fovino, M. Masera, and R. Leszczyna, "ICT security assessment of a power plant, a case study," in *Proc. 2nd Int. Conf. Critical Infrastructure Protect.*, Arlington, VA, Mar. 2008.

[12] A. Carcano, I. Nai Fovino, M. Masera, and A. Trombetta, "Scada Malware, a proof of concept," in Proc. 3rd Int. Workshop Critical Inform. *Infrastructures Security*, Rome, Italy, Oct. 2008.

[13] T. Novak and A. Gerstinger, "Safety-and security-critical services in building automation and control systems," *IEEE Trans. Ind. Electron.*, vol. 57, no. 11, pp. 3614–3621, Nov. 2010.

[14] W. Granzer, F. Praus, and W. Kastner, "Security in building automation systems," *IEEE Trans. Ind. Electron.*, vol. 57, no. 11, pp. 3622–3630, Nov. 2010.

[15] A. A. Creery and E. J. Byres, "Industrial cybersecurity for power system and SCADA networks," *IEEE Ind. Appl. Mag.*, vol. 13, no. 4, pp. 49–55, Jul./Aug. 2007.

[16] R. Chandia, J. Gonzalez, T. Kilpatrick, M. Papa, and S. Shenoi, "Security strategies for Scada networks," in Proc. 1st Int. Conf. Crit. Infrastructure *Protection*, Hanover, NH, Mar. 19–21, 2007.

[17] M. K. Mahmood and F. M. Al-Naima, "Developing a multi-layer strategy for securing control systems of oil refineries," *Wireless Sens. Netw.*, vol. 2, pp. 520–527, Jul. 2010.

[18] I. H. Lim, S. Hong, M. S. Choi, S. J. Lee, T. W. Kim, S. W. Lee, and B. N. Ha, "Security protocols against cyber attacks in the distribution automation system," *IEEE Trans. Power Del.*, vol. 25, no. 1, pp. 448–455, Jan. 2010.

[19] T. Mander, F. Nabhani, L. Wang, and R. Cheung, "Data object based security for DNP3 over TCP/IP for increased utility commercial aspects security," in *Proc. Power Eng. Soc. Gen. Meeting*, Tampa, FL, Jun. 24–28, 2007, pp. 1–8.

[20] M. Roesch, "Snort-lightweight intrusion detection for networks," in *Proc.13th Syst. Admin. Conf. LISA*, Seattle, WA, 1999, pp. 229–238.

[21] Last access 9/04/2009. [Online]. Available: http://www.digitalbond.com/index.php/research/ids signatures/modbus-tcp-ids-signatures/

[22] P. Gross, J. Parekh, and G. Kaiser, "Secure selecticast for collaborative intrusion detection systems," in *Proc. Int. Workshops DEBS*, 2004, pp. 50–54.

[24] J. N. Amaral, **K.** Tumer, and J. Ghosh, "Designing genetic algorithms for the state assignment problem," *IEEE Trans. Syst., Man, Cybern.*, vol. 25, no. 4, Apr. 1995.

[25] E. J. Anderson and M. C. Ferris, "A genetic algorithm for the assembly line balancing problem," Commt. Sci. Dept., Univ. Wisconsin-Madison, Tech. Rep. R 926, 1990.

AUTHOR'S PROFILE

Aarcha Anoop
is a research scholar (M.Tech) at Younus College of Engineering and Technology, Kerala with Specialization in Applied Electronics And Instrumentation .She received the B-Tech degree in Electronics & Communication Engineering in 2010 from Cochin University Of Science And Technology, Kochi, Kerala.