

Report for CSAW ESC 2017

Chenglu Jin, Saeed Valizadeh, Mason Ginter and Marten van Dijk

University of Connecticut

Email: {chenglu.jin, saeed.val, mason.ginter, marten.van_dijk}@uconn.edu

Abstract—In this report, we briefly explain our defense strategy in order to secure Programmable Logic Controllers (PLCs).

I. INTRODUCTION

A host-based intrusion detection (HIDS) is commonly used for monitoring specific activities and characteristics of a single device by means of software or appliance-based components known as agents. In this regard, in this project, we implemented a very lightweight HIDS module on the OpenPLC framework in order to keep track of input/output operations on the device.

To locate signs of likely security related incidents, the HIDS's agent, which we call it the *snapshotter* agent, essentially, takes a snapshot of the input/output operations happening on each PLC and sends them *in a stealthy secure way* to a server periodically. Then, the server

To locate signs of likely security related incidents, this agent, which is installed on each plc, logs all the events on the controller (more specifically, input/output operations) and sends the logs to a server periodically for the purpose of analysis and intrusion detection.

On the other hand, the implemented server has a predefined set of acceptable input/output pairs for each device and checks the integrity of the log itself (in case, the attacker has already compromised the device) and the validity of the operations happening on the controller. If any of the previous incidents happen, i.e., whether the log's integrity check fails, or an operation is detected as invalid, a flag will be raised and an intrusion is indeed captured¹. The server takes proper actions consequently which could be terminating the controller, revoking it from the network, or even recovering it to a previous clean/safe state.

For the log integrity checking feature, the server actually

For the operation validation purposes, the server actually traces deviations from predefined normal profile activities/behaviors of the input/output operations. Such profile activities could be generated based on the specific applications that the PLC is being used for. (see the comments in tex file)

Our defense mechanism can be summarized in security related information gathering and logging, incident identification, and taking effective actions to foil such incidents.

¹Assuming no errors in the controller functionality itself. Note that, even if there is error in the PLC functionality, our proposed method can capture it as well with same procedure

II. INTRODUCTION

A host-based intrusion detection (HIDS) is commonly used for monitoring specific activities and characteristics of a single device by means of software or appliance-based components known as agents. In this regard, in this project, we implemented a very lightweight HIDS module on the OpenPLC framework in order to keep track of input/output operations on the device.

To locate signs of likely security related incidents, the HIDS's agent, which we call it the *snapshotter* agent, essentially, takes a snapshot of the input/output operations happening on each PLC and sends them *in a stealthy secure way* to a server periodically. Then, the server

To locate signs of likely security related incidents, this agent, which is installed on each plc, logs all the events on the controller (more specifically, input/output operations) and sends the logs to a server periodically for the purpose of analysis and intrusion detection.

On the other hand, the implemented server has a predefined set of acceptable input/output pairs for each device and checks the integrity of the log itself (in case, the attacker has already compromised the device) and the validity of the operations happening on the controller. If any of the previous incidents happen, i.e., whether the log's integrity check fails, or an operation is detected as invalid, a flag will be raised and an intrusion is indeed captured². The server takes proper actions consequently which could be terminating the controller, revoking it from the network, or even recovering it to a previous clean/safe state.

For the log integrity checking feature, the server actually

For the operation validation purposes, the server actually traces deviations from predefined normal profile activities/behaviors of the input/output operations. Such profile activities could be generated based on the specific applications that the PLC is being used for. (see the comments in tex file)

Our defense mechanism can be summarized in security related information gathering and logging, incident identification, and taking effective actions to foil such incidents.

III. PROPOSED DEFENSE SCHEME

A host-based intrusion detection (HIDS) is commonly used for monitoring specific activities and characteristics of a single device by means of software or appliance-based components

²Assuming no errors in the controller functionality itself. Note that, even if there is error in the PLC functionality, our proposed method can capture it as well with same procedure

known as agents. In this regard, in this project, we propose a very lightweight HIDS module on the OpenPLC framework in order to keep track of input/output operations on the device.

To locate signs of likely security related incidents, the HIDS's agent, which we call it the *snapshotter* agent, essentially, takes a snapshot of the input/output values together with the current time stamps on each PLC, encrypts the log, and sends them *in a stealthy secure way* to a server periodically. Forward security is achieved in the key management of the encryption key. Next key is updated to the hash of the current key after each encryption. Therefore, even the adversary compromised the device completely, meaning that the encryption key is also exposed to the adversary, he is still not able to tell from the previous encrypted logs that whether he is caught by the intrusion detection systems or not.

Upon receiving the reports from PLCs, the server can use the received input and known PLC logic to simulate the expected output of the PLC. If the received PLC output does not match with the expected (simulated) output, then the server will conclude that this PLC has been compromised, so it should be restarted and reprogrammed by a correct logic program. Of course, for a specific application, the server can have a predefined set of acceptable input/output pairs for each device. This will reduce the computation of the server side. Besides the input/output behaviors, the integrity of the log can be checked at the server, in case that the attacker has already compromised the device.

Our defense mechanism can be summarized in security related information gathering and logging, incident identification, and taking effective actions to foil such incidents.

This solution has five features that make our proposed scheme very suitable to the intrusion detection for PLCs:

- 1) We can guarantee the integrity of the log generated by each PLCs, because we are able to verify its integrity at the server side, and we can detect any dropped package in the communication.
- 2) The log is sent in a stealthy way, such that the attacker is not able to tell whether he gets detected or not. This gives more advantages to the defenders to record more behaviors of the attackers, which can be used for further investigations.
- 3) This approach requires no redundant PLCs for detecting an intrusion, so it saves the cost of extra PLCs for implementing any intrusion detection schemes, e.g. comparing the outputs of two redundant PLCs.
- 4) Our approach can be applied on any legacy PLCs, because we only record the inputs and outputs of each PLCs in the log. No internal states of the PLCs are required.
- 5) Since the logic running on the PLCs is usually very simple, a powerful server can easily simulate a large number of PLCs in parallel. This makes our solution easily scalable to monitor a large number of PLCs.
- 6) In the scenario of an industrial control system, all the devices are connected using cables, so the connectivity of each PLC device is very well established. This will

reduce the false positive of our system caused by the network delay.

A. Information Gathering

In order to achieve the stealth logging, we suggest to use the logging mechanism in [1]. The log has two requirements:

- 1) All the logs are encrypted, and the encryption key is updated in a forward secure key.
- 2) The logs are buffered in the device itself, and is sent periodically to the server.

The first requirement prevents the adversary, who eavesdrops the communication between the PLCs and the server, from figuring out whether the current attack is being detected or not. A forward secure key management scheme guarantees that even when the current key is compromised by the adversary, the adversary is still not able to decrypt all the encrypted logs sent before.

The second requirement prevents the attackers from dropping all the encrypted logs that possibly record some traces of the adversary, because if the report of one device is not sent to the server on time, the server will conclude that this device has compromised by attackers.

B. Incident Identification

If any of the incidents happen, i.e., whether the log's integrity check fails, or an operation is detected as incorrect, a flag will be raised and an intrusion is indeed captured³. The server takes proper actions consequently which could be terminating the controller, revoking it from the network, or even recovering it to a previous clean/safe state.

C. Mitigation

Once an incident has been identified, one can restart the compromised PLCs, and reprogram it to a safe/correct state/logic. If the monitoring server is not able to reset the PLC to a clean state, then a technician must be sent to physically approach the PLC and fix it. In the meanwhile, another uncompromised PLC can be deployed to replace the compromised PLC to continue the industrial process.

IV. IMPLEMENTATION

A. Algorithms

We implemented the stealthy logging mechanism in OpenPLC framework on a Raspberry Pi. The operations in one scan cycle can be described in Algorithm 1. In OpenPLC framework, **Read_Input** function and **Write_Output** function has been modified to return 1 if any value changes in the current scan cycle. Therefore, only when the input or output values change, this event will be recorded in the log. A single event is encrypted and its key is updated to its hash value, such that the old key for encrypting this event will be overwritten right after each encryption.

³Assuming no errors in the controller functionality itself. Note that, even if there is error in the PLC functionality, our proposed method can capture it as well with same procedure

Algorithm 1 PLC Logging

```
1: procedure LOGGING(Period)
2:    $EL = \{\}$ 
3:    $T = 0$ 
4:    $Event\_ID = 0$ 
5:    $Update = 0$ 
6:   while True do                                ▷ Scan Cycle
7:      $Update = 0$ 
8:      $T = T + 1$ 
9:      $Update = Read\_Input()$ 
10:     $PLC\_Logic()$ 
11:     $Update = Write\_Output()$ 
12:    if  $Update == 1 || T \bmod Period == 1$  then
13:       $TempLog = \{T, Input, Output, Event\_ID\}$ 
14:       $EL \leftarrow (ENCRYPT(TempLog, Key))$ 
15:       $TempLog = \emptyset$                                 ▷ Delete the plaintext
16:       $Key = HASH(Key)$ 
17:       $Event\_ID = Event\_ID + 1$ 
18:    end if
19:    if  $T \bmod Period == 0$  then
20:       $SEND(EL)$ 
21:       $EL = \emptyset$                                 ▷ Empty the buffer
22:       $Event\_ID = 0$ 
23:    end if
24:  end while
25: end procedure
```

Since the input or output of PLC does not change very often, only reporting changes of input/output values in the log, we can significantly reduce the size of the log. For an application that requires the inputs/outputs of PLCs to be updated very frequently, we suggest to reduce reporting period or ignore some small fluctuations in the analog output.

After a predefined reporting period is elapsed, the PLC needs to send all the buffered encrypted log to the server and empty the buffer for the next period.

The algorithm on the server side works as Algorithm 2. The server waits for the incoming packet from the PLC. If the packet is not received on time, the server concludes that the PLC is compromised, we need to restart the PLC. If the server gets the packet on time, then we need to use the associated key to decode this packet and update the key stored at the server. After that, we need to check whether the decrypted data has the correct data format or not. If not, then it means the integrity of the packet has been compromised, we need to restart the PLC. If the integrity is also valid, then we need to extract the input change in this period with its time stamp. This input change and time information is applied to the PLC logic simulator, which will generate the expect output change with its time stamp. Then this expected output is compared with the output received from the PLCs. If they do not match, then the server knows that the logic running on the PLC has been maliciously modified. Therefore, it is also required to restart the PLC in this case.

Algorithm 2 Server Monitoring

```
1: procedure MONITORING(Period)
2:    $Alarm = False$ 
3:    $Last\_Rec\_T = Current\_T$ 
4:   while True do
5:     while True do
6:       if  $Current\_T - Last\_Rec\_T > Period$  then
7:          $Alarm = True$ 
8:         Break                                ▷ Missing one Packet
9:       end if
10:      if  $EL \leftarrow REC\_PACKET()$  then
11:         $Last\_Rec\_T = Current\_T$ 
12:        Break                                ▷ Receive one Packet
13:      end if
14:    end while
15:     $(L, Key) \leftarrow DECRYPT\_PACKET(EL, Key)$ 
16:     $Alarm \leftarrow VERIFY\_FORMAT(L)$ 
17:     $(T', Input', Output', Event\_ID') \leftarrow L$ 
18:     $(T'', Output'') \leftarrow PLC\_SIMULATE(T', Input')$ 
19:     $Alarm \leftarrow COMPARE(T', Output', T'', Output'')$ 
20:    if  $Alarm$  then
21:      Restart PLC
22:    end if
23:  end while
24: end procedure
```

B. Implementation Details

As cryptographic primitives, we use AES-128 and SHA-256 as the encryption algorithm and hash function respectively. The updated key is the first 16 bytes of the hash value of the previous key. The reporting period is set to 10 seconds in the prototype implementation, but it can be easily adapted to a different value if it is required by the application.

To minimize the size of the log we sent, we design one data format to record one event (input or output change) in less than 128 bits, which can fit in one encryption block of AES. The data format is depicted in Figure 1. The first byte is used as an indicator of the start of one event; we set it as 0xFF. The second and third bytes are used to store the event ID in the current time period. Since the PLC is required to run on 100Hz, and we set the reporting period to be 10 seconds, the maximum of events can happen in one period is 1000. Therefore, two bytes are more than enough to represent the maximum number. The forth and fifth bytes are reserved for device ID. In total, 65536 devices are allowed to be managed by one server. The following 4 bytes are used to store the time stamp, because in OpenPLC framework, it is a 4 byte variable. The next six bytes are divided for storing digital inputs, digital outputs and analog output value respectively. Each of them takes two bytes. Since the number of digital inputs and digital outputs on a Raspberry Pi is 14 and 11. Two bytes are enough to store all the input/output pins. Also, only one analog pin can be used on a Raspberry Pi, and this value can be represented by a 16-bit value stored in 2 bytes. In the end, another byte

of 0xFF is appended to indicate the end of one event log.

V. EVALUATION

In this section, we will evaluate our solution and implementation in terms of security and performance.

A. Security Analysis

Our adversarial model is that we assume the adversary has the capability to upload malicious logic to the PLC that will generate erroneous outputs or further compromise the entire PLC. After the entire PLC is compromised, we assume the attacker can even get access to the encryption key, which will be used to encrypt the next event, together with the encrypted logs that have not been sent to the server yet. This is a very strong adversarial model for a remote attacker who is connected over the network. Only when our solution is applied to a modern PLC or OpenPLC, meaning that the PLC logic and the When our solution is applied to the legacy PLCs, the logging mechanism is implemented outside of the legacy PLCs as another hardware layer, so a remote attacker has no way to get the key in it.

In addition, we assume that the logging mechanism is working properly until a certain point after the compromise initiates. In other words, we assume there exist a time window that the attacker has initiated the intrusion, but the logging is still working properly. The overall timeline can be illustrated as Figure. 2.

The time window between the start of a compromise and the moment that the logging failed is very critical. Essentially the logs generated in this time window tells the server that an intrusion is happening. The attacker has four different approaches to deal with this log, and all of them will lead to detection. We list all the four possible approaches below:

- 1) **Let the encrypted logs send to the server.** Since this log contains the information about intrusion, the adversary will be detected.
- 2) **Try to decrypt the logs.** Notice that, it is possible that the adversary compromises the device completely (i.e. knowing the current encryption key) before the log has been sent. But the previous log has already been encrypted and buffered, and the encryption key is updated in a forward secure way. The attacker has no chance to inverse the hash function and gets to know the previous encryption key. Therefore, the attacker is not able to decrypt the log, and has to send this log as it is.
- 3) **Tamper with the encrypted logs.** The attacker can either send something random or replay a previous log about a clean state. Since the encryption key is updated after each event, the server with an updated key will not be able to decrypt this tampered log. Therefore, the server will also conclude that an intrusion is happening on this PLC.
- 4) **Block the communication between the PLC and the server.** If the communication between the compromised PLC and the server is blocked, no log will arrive the server on time. Since the PLC is required to send the

report periodically, any PLC which fails to send the log on time will be considered as compromised.

To summarize, there is no way for the attackers to avoid detecting by our intrusion detection scheme.

B. Sensitivity of Detection

Since our solution requires the logs from PLCs to be sent periodically, it is possible for the server to have a detection latency no larger than a reporting period. Hence, when we select the parameter of reporting period, we need to consider this detection latency as well.

C. Performance Overhead

VI. FUTURE WORKS

In the future, we can implement this scheme to a complete industrial control network, which has a real powerful server connected to. This can help us study how much this scheme can be scale in practice, due to the constraints on the network bandwidth or a certain application.

Also, we need to try to find a more efficient mitigation strategy, that affects the real industrial process as little as possible.

VII. CONCLUSION

In this project, we implemented stealthy logging mechanism on the PLCs, and the server also runs the monitoring program.

ACKNOWLEDGMENT

The authors would like to thank Thiago Alves for the helpful discussion.

REFERENCES

- [1] Bowers, Kevin D., et al. *Pillarbox: Combating next-generation malware with fast forward-secure logging.*, 3rd ed. International Workshop on Recent Advances in Intrusion Detection. Springer, Cham, 2014.

#Byte	1 Byte	2 Bytes	2 Bytes	4 Bytes	2 Bytes	2 Bytes	2 Bytes	1 Byte
	Start	Event ID	Device ID	Time	Digital Inputs	Digital Outputs	Analog Outputs	End
Example	0xFF	0x0002	0x1234	0x00000010	0xC000	0x8000	0x7832	0xFF



16 Bytes in total

Fig. 1. Data format of one event in the log.

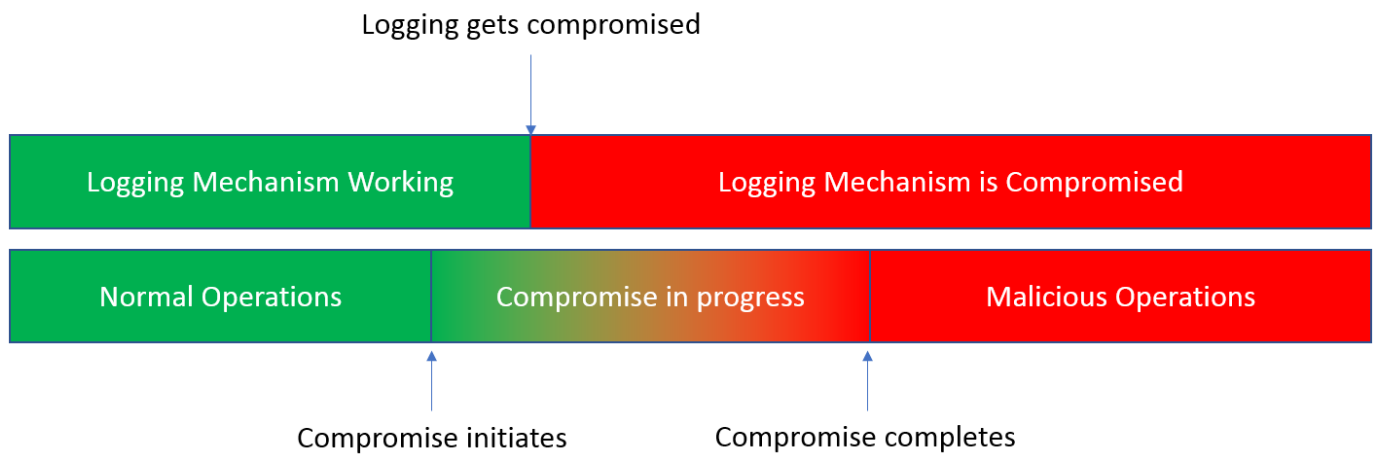


Fig. 2. Time line of one compromise.