Security Framework for Control System Data Classification and Protection

Bryan T. Richardson and John Michalski

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under Contract DE-AC04-94AL85000.

Approved for public release; further dissemination unlimited.
Security Framework for Control System Data Classification and Protection

Bryan T. Richardson and John Michalski
Information Assurance & Survivability
Sandia National Laboratories
P.O. Box 5800
Albuquerque, New Mexico 87185

Abstract

This document presents a data classification process that gives utility administrators, control engineers, and IT personnel a cohesive approach to deploying efficient and effective process control security.
Acknowledgements

The authors would like to acknowledge the work resulting in a framework to categorize and protect control system data, was funded by the U.S. Department of Energy/Office of Electricity Delivery and Energy Reliability (DOE/OE) as part of the National SCADA Test Bed (NSTB) Program.
Executive Summary

This document presents a data classification process that gives utility administrators, control engineers, and IT personnel a cohesive approach to deploying efficient and effective process control security. The fundamental goal is a clear delineation of control system data that will enable effective implementation of security techniques and technologies so the control system can function as required in the face of threats. Once created, the data classification security framework will help reduce the risk of energy disruptions due to control system failure by securing data critical to the operation of the control system.

Many new regulatory requirements and recommendations have been developed since 9/11 that focus on making critical infrastructure control systems less vulnerable to malicious attacks. A significant problem with these new requirements and recommendations is that system designers and administrators do not know what steps to take to meet them. Examples of such requirements and recommendations include data authentication and data exchange integrity\(^1\), network security and secure network management\(^2\), compartmentalizing communication\(^3\), and blocking access to resources and services\(^4\).

Effective and efficient protection of control system data, in terms of both operational complexity and cost, requires that the types of data used in the system be identified and classified according to their importance in operating the control system. This enables system designers to determine where and how to secure the system. Then a protection profile addressing the threats present in the operating environment is assigned to each data type. The profile must take into account the importance of the data to operations, the physical location of the data, and the traversal of the data across interface boundaries. Finally, practical implementation details are described that will provide the level of security specified by the protection profile. The data classification framework outlined in this document is generic in nature, so it can be used by all critical infrastructure sectors. It is intended to be flexible, making it possible to include sector-specific security requirements such as NERC CIP\(^5\).

This document is intended to familiarize the reader with the concept of a data classification framework for control systems. The basic descriptions of the four main components (data type identification, data classification, data protection profile, and implementation guide) given in this report require some additional development and refinement for application to real-world systems.

---

This page intentionally left blank
Table of Contents

1 Introduction ................................................................................................................................. 9
  1.1 Background .......................................................................................................................... 9
    1.1.1 Description ................................................................................................................ 9
    1.1.2 Historical Information ............................................................................................. 9
    1.1.3 Significance ............................................................................................................. 9
    1.1.4 Literature Review ................................................................................................... 9
  1.2 Purpose .................................................................................................................................. 10
    1.2.1 Reason for Investigation ....................................................................................... 10
    1.2.2 Roadmap challenges ............................................................................................ 10
    1.2.3 Audience ............................................................................................................... 10
    1.2.4 Desired Response ............................................................................................... 11
  1.3 Scope .................................................................................................................................... 11
    1.3.1 Extent and Limits of Investigation ....................................................................... 11
    1.3.2 Goals .................................................................................................................... 11
    1.3.3 Objectives ........................................................................................................... 11

2 Approach .................................................................................................................................... 13
  2.1 Methods ............................................................................................................................... 13
  2.2 Assumptions ....................................................................................................................... 13
  2.3 Procedures .......................................................................................................................... 13

3 Results and Discussion ............................................................................................................ 15
  3.1 Data Type Identification ................................................................................................. 15
  3.2 Data Classification ............................................................................................................. 17
  3.3 Data Protection Profile ..................................................................................................... 18
  3.4 Implementation Guide ...................................................................................................... 21
    3.4.1 Hypothetical Architecture Implementation Example .......................................... 21

4 Conclusions .............................................................................................................................. 23

5 Recommendations .................................................................................................................. 25

Appendix A: References ............................................................................................................. 26
Appendix B: Acronyms, Symbols, Abbreviations .................................................................... 28
Appendix C: Glossary .................................................................................................................. 30
Appendix D: For More Information ......................................................................................... 33

Table of Figures

Figure 1. Security Profile Model ................................................................................................. 12
Figure 2. Control and Automation Reference Model ............................................................... 16
Figure 3. Data Classification Based on Organization Tiers .................................................... 17
Figure 4. OSI Communication Reference Model ..................................................................... 20
This page intentionally left blank
1 Introduction

1.1 Background

1.1.1 Description
Many new regulatory requirements and recommendations have been developed since 9/11 that aim to make control systems for critical infrastructure less vulnerable to malicious attacks. A significant problem with these new requirements and recommendations is that system designers and administrators do not know what steps to take to meet them.

1.1.2 Historical Information
Many of the control systems associated with the energy sector infrastructure of the United States were originally built upon proprietary implementations and protocols. Most of the networks that supported these control systems were isolated from open public networks. This afforded a sense of security for the data that traversed these networks. As many of these infrastructures move towards modern open architectures and standard protocols, the control systems are now at much broader risk due to exposure to malevolent cyber activity. To protect these control systems from adversary manipulation and compromise, it is important to limit the exposure of important resources associated with the operation of energy-related critical infrastructures.

1.1.3 Significance
Control system data varies in importance. Consider a data protection scheme in which all data is protected equally; in other words, a homogeneous scheme. Compared to an importance-based scheme, where data that’s more important is better protected, the homogeneous approach provides less protection than necessary for high-value data and more protection than necessary for low-value data. Resources are used over-protecting trivial data that should be used for important data.

It’s more effective and efficient to base the protection level on the importance of the data. The data classification framework described in this report is not only specific to control systems, it also enables importance-based data protection scheme. Assuming better protection costs more, this enables maximum risk reduction for a given data protection budget. This benefit of using the approach described here increases as the attacker becomes more sophisticated, because a sophisticated attacker understands control system function and will target the most important data.

1.1.4 Literature Review
Areas with difficult-to-meet requirements and recommendations new since 9/11 include data authentication and data exchange integrity [1], network security and secure network management [2], compartmentalizing communication [3], and blocking access to resources and services [4].
Data classification is currently used to determine how data will be secured, managed, retained, and disposed of in enterprise and government environments [5]. However, traditional security and risk management practices generally result in a data classification scheme oriented towards protecting privacy-related data [6]. Such a scheme would be irrelevant for control system data protection, since none of the data is privacy-related.

1.2 Purpose

1.2.1 Reason for Investigation
An important aspect of protection of any system is the identification of the system’s resources. A crucial process control system resource is the data that resides within the system; data control and manipulation is an important part of all control system operations. Since data is important, it needs to be protected. To provide the proper levels of data protection, there needs to be a method to identify the different types of data and their associated criticality to operations within control system architectures.

One of the pitfalls of providing security protection for resources associated with a control system is the tendency to protect all resources at the highest level of security. At face value this seems appropriate: a wall provides protection, and a higher wall offers more protection. This is undeniable, but, generally, the more imposed security, the more difficult the operator’s job. This is because higher levels of protection are typically accompanied by increasing operational complexity and cost, which can limit operational flexibility and/or performance. This would make incorporating data protection highly unattractive to an owner/operator. And, as discussed in Section 1.1.3, Significance, it’s an ineffective and inefficient use of resources to treat the data homogenously when system data resources differ in importance, as they do in a control system.

1.2.2 Roadmap challenges
This document responds to the following challenges in the Roadmap to Secure Control Systems in the Energy Sector [7]:

- Develop and Integrate Protective Measures: By using a process similar to the one described in this report, products and techniques currently available and in use today could be identified for use in securing critical infrastructure control system data at different levels of security as applicable and necessary.

- Sustain Security Improvements: Historically, cyber security for control systems has been a difficult business case because of the general “secure everything” train of thought. By using a process similar to the one described in this report, decision makers can determine exactly what needs to be secured and at what level, rather than securing everything at the highest level.

1.2.3 Audience
This document outlines a framework for use by control system asset owner/operators to identify, classify, and protect their data.
1.2.4 Desired Response
We would like to see critical infrastructure control system security processes specified in terms of a data classification and protection approach that control system asset owners and operators have arrived at by executing the approach described in this report.

1.3 Scope

1.3.1 Extent and Limits of Investigation
The approach we present in this document considers both the criticality of the data to the day-to-day operations of the control system and the unique characteristics of a control system. Thus, with this security framework we also take into account the requirements put forth by the control system architecture (availability requirements, latency issues, etc) in order to guarantee control will not be affected. This security framework helps to maximize the sharing of protected information: Within this framework data can be classified in such a way that it satisfies the required security constraints, but not to the extent that people and/or applications can’t get access to the data they need [8].

1.3.2 Goals
Our goal is to ensure appropriate levels of data protection while minimizing impact on control system operations.

1.3.3 Objectives
The objective is the description of a process for developing an efficient, effective protection scheme for process control data.

The first step is to identify the different types of data associated with the operations of a utility control system. The type identification process described here allows data types to be analyzed based on their roles in control system operations. Knowledge of these roles enables attributes such as location, transportation, manipulation, and storage to be associated with data elements. This higher-resolution classification scheme limits the need for a “one-size-fits-all” security approach and allows a more granular approach to data protection. Once the identification of the data is complete, the data is then classified, or “ranked”, to determine which data types are most important to the continuity of operations. Then each data type is assigned a data protection profile that takes into account the operational aspects of the data and the attributes that are most important to protect. Finally an implementation approach is identified to provide a practical means of protection that balances operational efficiency with data protection. Figure 1 shows the structured approach to data classification and protection.
Figure 1. Security Profile Model
2 Approach

The approach taken in the development of this document was to bring together well-received ideas and knowledge in the area of IT security, control systems security, and data security to create the basis of a data classification framework applicable to control systems. The idea is that this document will provide a solid base for generating an actual framework of this nature by providing key models, standards, and methodologies that already exist as possible building blocks.

2.1 Methods
We first chose to take a look at what problems a data classification framework might help solve. From there, we looked to seed how incorporating a more general case of the Reference Model for Control and Automation Systems in Electric Power [9] would help in solving some of these problems. We also looked into the use of existing information assurance elements and existing IT network standards to help identify potential methods of implementing data security. Lastly, we considered a use case for a framework of this nature.

2.2 Assumptions
Our only major assumption was that the control system reference model used here [9], while specific to power systems, does not differ from a general process control system case in any way that would limit the applicability of this document’s process to the framework outlined.

2.3 Procedures
We generated a data classification process description based on experience, lessons learned, existing models, standards, frameworks, etc., and logical extension. Searches were performed to see if previous work in the area of data classification specific to control systems has been done, of which none was found. We also spoke with people fluent in the CoBiT methodology to see if it specifically addresses any of these issues, and also spoke with IT networking experts to see how existing security standards might be applicable for this type of exercise.
This page intentionally left blank
3 Results and Discussion

The process of creating a security profile is decomposed into four sub-processes. These sub-processes provide an integrated solution for data security for each classification type. A detailed explanation of each of the four sub-processes follows in the next four sections of the report. The general process is described in the following paragraph.

First, data elements in the control system under scrutiny are named, described, and assigned a data type according to their function as described by the control system reference model. Then, the data is classified according to its criticality to control system operations. Next a protection profile that addresses the operationally encountered threats is assigned to each class of data. Finally, practical implementation guidance and tactics are given to enable the level of security required by the protection profile. The data classification framework outlined in this document is generic in nature and can be utilized by all critical infrastructure sectors. It is intended to be flexible so sector-specific security requirements such as the NERC CIP Standard [10] can be incorporated.

3.1 Data Type Identification

It is the intent of the research described in this report to enable control system data protection in both the electrical industry and the oil and gas industry and to a lesser extent water and industrial controls. This objective for broad applicability entails a data type identification methodology that can generalize data types in such a way that the terminology is relevant over a large subset of control systems.

To determine the different types of data associated with control systems, a generic reference model for data types needs to be identified or developed. This data reference model can then be used to help the utility owner identify the different types of data resident on the control system under analysis. The generic reference model must be able to identify types of data based on functionally or purpose along with its association to the whole of operations. At this juncture in our analysis, the Reference Model for Control and Automation Systems in Electric Power [9] is used for data type identification. Figure 2 shows the reference model; see [9] for a detailed explanation.

Once the data type identification has been completed, the classification of the data types needs to be determined based on their importance to maintaining operations. This classification process is necessary to be able to determine not only the criticality of the data in its role in operations, but also some of the common characteristics that govern its interaction.
Figure 2. Control and Automation Reference Model
3.2 Data Classification

A data classification technique should be able find a balance between over-refinement, which can lead to an overly complex protection scheme, and a one-size-fits-all outcome that will almost certainly overprotect some data elements while under-protecting others. The approach doesn’t have to be very complex, but how granular the classification method becomes will impact the way in which protection mechanisms are identified for its implementation.

In order for a classification technique to be manageable and flexible, the technique must be able to allow changes in operational requirements and infrastructure. This requires an iterative approach to classification requirements that aligns itself to service offerings on the control system network that may take multiple reviews, with the first review creating a baseline and subsequent reviews or interactions refining the classification results. The criticality of the data should be determined by the service it supports and the service’s value to the control system. Figure 3 displays this concept.

Creating system priority tiers requires gathering significant descriptive information about the process control system for use in developing the system’s data structure. This activity must include all the groups that use the data, including maintenance technicians responsible for trouble shooting and repairing equipment, master control personnel monitoring and analyzing process activity, IT staff responsible for the underlying network, and the business units that process data on the business LAN. To accurately create the tiers, the total operations picture must be taken into account to include all applications responsible for the proper operation of the control system.

![Data Classification Based on Organization Tiers](image)

Figure 3. Data Classification Based on Organization Tiers
Data classification must be associated with requirements to perform operations. The association of operations requirements to data classification can provide a structural overview for classifiers, reducing overall level of effort and providing a ranking of data or applications based on operational needs. This prevents the case where all data is important and requires the most expensive approach to its protection. Data classification must be able to align the classification of data with a standardized infrastructure that provides enough granularity to identify data storage management, user’s interaction, interface identification, and transmission.

Key performance constraints must also be identified for each data type to assist in the data classification process. For example, some of the data that is resident on a control system has a temporal constraint that must be met to be able to maintain the validity of the data (for example, usefulness of data to proper process operation). Other types of data are more persistent or static and do not have a time critical constraint associated with their processing. A complete understanding of operational requirements can help identify the need for multiple data storage classifications. The ability to map applications to logical mappings of the user environment with respect to the infrastructure can also facilitate the classification task.

Data can also be associated with a service that is provided within the control system. Binning data according to the service it provides may also shed some light on its criticality. Looking at a service as it applies to operations is another valuable means of qualifying a classification. In fact, after services are cataloged, a service level agreement can be constructed based on metrics for each service implementation. This catalog can provide all the technical details needed to provide the level or quality of service on the network or infrastructure required to maintain its operation. This catalogue can be seen as a dynamic document that can be updated as technology, infrastructure, or service applications change over time.

Once a classification structure has been created and user requirements of the data identified, the initial classification can then be used to develop a protection profile.

### 3.3 Data Protection Profile

During the data protection profile stage, each identified data classification must be assigned an appropriate protection profile. To be able to assign a data protection profile to a data class, the needed protection level for that class must be determined. The protection profile will provide a statement of the security requirements that are required to address generalized threats that may exist in the operating environment against a specific class of data. The process of assigning a profile must take into account the overall importance of the data to operations, the physical location of the data with reference to the overall control system architecture, and the crossing of data at interface boundaries within the architecture, which may lead to protocol translation, performance metrics, and storage/retrieval activities.

The approach that will be pursued for this activity is the use of information assurance elements to characterize each profile. Information assurance can be understood as a system’s ability to protect data by ensuring its availability, confidentiality, integrity,
reliability, authenticity, and non-repudiation. This includes providing for restoration of information systems by incorporating protection, detection, and reaction capabilities.

Assurance in the context of data or information is a measure of confidence that the security features and architecture can mediate and enforce an approved security policy. Assuming a security policy has been created, the measured confidence is derived from analysis, verification, and testing.

Availability of the data refers to the idea that the data is accessible to all authorized users at all times. Its unavailability may be induced in a physical way, as the failure of critical network components, power disruptions, and physical plant disruptions, either malicious or natural. Availability can also be impacted in a logical way, in the form of improper addressing or routing, and through the use of denial-of-service attacks, which is the deliberate insertion of unwanted data into the network. This is often associated with address spoofing, which associates the introduction of unwanted data with a trusted end node.

Confidentiality of information refers to the protection of data that allows only the intended recipient to be able to read the information. Data should have a “need-to-know” quality associated with its handling, which can prevent unauthorized exposure. Confidentiality can be provided in multiple ways, such as in a physical approach with a physical building security enclave where data is processed creating physically protected networks, or virtual confidentiality, which relies on some form of encryption to prevent the disclosure of the information while the data is “in-flight” to its destination.

Integrity of information refers to the ability of a system or mechanism to detect changes or modifications to an original message. Modern techniques implement integrity across a packet header and/or data field by creating a hash across the contents of the packet. This hash is based on a one-way function, and can detect any modifications to the original contents of the packet. A systems integrity approach will review the architecture of a system and its implementation. How a system is designed and maintained is also an important aspect of system integrity, which includes contingency planning for power failures and disaster recovery.

Reliability within the context of data communications refers to the ability of a communication system to provide consistent intended service over a large percentage of the time. The reliability of the data transmitted over this network is subject to the interconnected network components of the system and the protocols that are used to provide the end host to end host communications. Communication protocols can provide a “reliability” facet to the data communications process. For example, a somewhat noisy network link creating bit errors within a packet does not by itself prevent communication between two communicating end nodes if the communications protocol is able to detect and retransmit the offended packets. The reliability of the packet communication process can still remain high in spite of occasional bit errors injected by the network link.

Authenticity of data refers to its original conception and the binding of its author. Maintaining this relationship of data and associated conception in modern network communications today is done with the use of public key encryption and a process called a digital signature. To create a digital signature, a hash is created across the data. This hash is sometimes referred to as a message digest. This hash creates a one-way
cryptographically strong series of bits that represent the original contents of the message. These bits are then encrypted with the private key of the originator author and sent along with the original message.

The purpose of *non-repudiation* is to assign attribution to an originated message that any third party could verify and be confident that it cannot be disputed. It can also prevent a recipient of a message from denying a message was received.

This data classification security framework outline requires the development of a data surety matrix to aide in the assignment of surety elements to data classes. The data surety matrix will help define what assurance elements are important for each type of user requirement defined in the classification of the data.

To be able to draw upon all available relevant protection mechanisms, it is necessary to represent the end-to-end transmission and reception structure that will be used for data delivery. This is best done using an accepted logical reference model because results are stated in canonical terms. One such model is the Open System Interconnect (OSI) communications model illustrated in figure 4 (see, for example, [11]). The OSI model represents the communication process in seven layers, each providing a logically distinct data transmission service. Once the logical interaction of the layers that make up a system of interest has been established, the association of logical protection and data surety elements to the logical elements of the communication process can be discussed in canonical terms.

![Figure 4. OSI Communication Reference Model](image)

This data classification security framework outline also requires the development of a list of protection mechanisms available for each OSI layer that provide element(s) of data surety. This list will help identify what protection mechanisms are available to provide the assurance elements required to protect the data, if the protection mechanisms have to be implemented using hardware or software, etc.
3.4 Implementation Guide

Once data is properly classified and a data protection profile has been assigned, the final step is to produce an implementation guide that details a practical implementation of the protection profile. This implementation guide tells implementers how to realize a physical implementation based on the logical description. The required protection areas include the identified means of data storage, retrieval, transmission and reception, whether the data lies within a physical boundary or not. A threat identification or risk evaluation can be useful, but the approach should show how unauthorized access or disclosure of data is prevented based on data criticality against a generic threat profile. This generic threat profile should be guided by a company security policy that, among other things, characterizes secure operation of the control system.

Part of the implementation process will be the identification of device requirements described at a granular enough level that allows for the recognition of security attributes. All identified devices can then be associated with the appropriate subsystem as described in the Reference Model for Control and Automation Systems in Electric Power [9] and seen in figure 2. These subsystems can be reviewed independently, but if there are demarcation points between subsystems or identified inter-dependencies, the final level of security must fit into the overall integrated system.

Device identification and assignment should be general in nature with examples that can provide actual product insertion. Each identified device may fulfill both an operational and security need, and should be associated with an overall security policy. The implementation process will be guided by the previous layers in the security profile model shown in figure 1, which can be summarized with the following questions:

1. What type of data is my device/software processing?
2. How important is my data to operations?
3. What type of protection must be provided for the data being processed or manipulated?
4. What is the important performance metrics associated with the data?
5. Are there any dependencies associated with the data (service tier)?

Below is a hypothetical example of how a utility might use the previous data classification security framework steps, the Reference Model for Control and Automation Systems in Electric Power [9], and their own security policy to form and guide the architecture implementation step of the framework.

3.4.1 Hypothetical Architecture Implementation Example

Presumably, the utility’s security policy would be updated to include specific requirements for devices in their control system, categorized according to the reference model mentioned above. These requirements might include such things as efficiency of the transport of data, latency, security, management, etc., which may be specified at multiple levels. As security personnel evaluate new devices and validate that they meet the specified requirements at some level, the devices could then be added to the security policy as well. Then, after someone has gone through the first three steps of the data classification security framework, they could use the utility’s security policy to determine
which device(s) that have been evaluated and validated by the utility’s security personnel would provide them with the protection they need at the level they need, as specified by the data protection profile.
4 Conclusions

We have described a security framework to identify and properly classify each data type to provide the most applicable security solution to each system device associated with the storage, manipulation, transmission, and reception of data. The purpose of this data classification approach is to provide the utility administrator, control engineers, and IT personnel a cohesive approach to understanding the necessary protection, prospect, and limitations of deploying security methods based on data classification in process control environments.

The proposed security solution considers both operational requirements, such as performance constraints and services provided, and security requirements, such as availability, reliability, and authenticity.

Using this control-system-specific approach, data that’s more important will be better protected. This enables optimal use of security resources.
This page intentionally left blank
5 Recommendations

We recommend the approach described in this paper to provide a process control data description that includes protection profiles directly traceable to system functions. The resulting information is the foundation of an efficient, effective information security system.
Appendix A: References


http://csrp.inl.gov/documents/MitigationsForVulnerabilitiesCSNetsISA.pdf


http://www.nerc.com/~filez/standards/Reliability_Standards.html

http://searchnetworking.techtarget.com/sDefinition/0,,sid7_gci212725,00.html
This page intentionally left blank
Appendix B: Acronyms, Symbols, Abbreviations

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIP</td>
<td>Critical Infrastructure Protection</td>
</tr>
<tr>
<td>IT</td>
<td>Information Technology</td>
</tr>
<tr>
<td>LAN</td>
<td>Local Area Network</td>
</tr>
<tr>
<td>NERC</td>
<td>North American Electric Reliability Corporation</td>
</tr>
<tr>
<td>OSI</td>
<td>Open System Interconnect</td>
</tr>
</tbody>
</table>
## Appendix C: Glossary

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>9/11</td>
<td>September 11(^{\text{th}}), 2001 terrorist attacks on the United States of America.</td>
</tr>
<tr>
<td>Need-to-Know</td>
<td>Determination made by authorized holder of information that a recipient requires access to information in order to perform a function.</td>
</tr>
<tr>
<td>Tier</td>
<td>One of two or more layers, one atop another, that depend on or provide to other tiers.</td>
</tr>
</tbody>
</table>
Appendix D: For More Information

Jennifer DePoy, Manager  (505) 844-0891, jdepoy@sandia.gov
Bryan T. Richardson     (505) 845-2386, btricha@sandia.gov
John Michalski          (505) 844-3122, jtmicha@sandia.gov