Averting Espionage in Distributed Control System Environments through Partial Data Disclosure

Introduction

* Why your topic is important (convince us!)
* Where is it used? Applications
* What you will talk about / do
* Overview of the rest of your paper

Background and related work

* Any relevant and specific info, e.g. software / hardware statistics, equipment used
* What other people had to say on this topic
* What other people did on this topic
* Problems and shortcomings of their work
* How your work is different and better

Proposed methodology

* Your approach to the problem
* What you did
* Code / Algorithms
* What did / didn’t work
* Results – include graphs, equations, pictures, as appropriate

Conclusions

* What was accomplished / learned
* What you would have done differently
* Future work

References

Appendix

Technology Discussion

Project description should include: **“…of its scientific objectives, of the methodology that will be used, and of the significance and originality of this research project within the applicants’ field.”**

1. Description of field
   1. Overall systems look
   2. The age of cloud and data
   3. Data protection
2. Frameworks and questions to ask
   1. Many standards: IEEE, IEC, NIST; primary framework: OPC, PLC, TCP
   2. Main question: How to distribute/obfuscate/guard detailed data to protect against intellectual theft/espionage?
      1. Who wants access to the data?
      2. How does data security function in frameworks right now?
      3. What requirements are needed to preserve data security?
   3. Cloud aspect of DCS
      1. How does it work now?
      2. What does it offer in terms of data security?
      3. Is encryption enough?
3. Objective: Research ways of preserving data security in DCS to disclose data within requirements
   1. Possible routes
      1. Differential Privacy
      2. Disclosure of only relevant indexes of data required
      3. Cloud security

Problem Approach

What do business need to do to create detailed data privacy within compatible standards?

How do we protect the detailed data against intellectual theft/espionage?

Can we create a framework to protect data while meeting compatible standards?

If not what changes need to be implemented to current standards to protect detailed data?

* Logically centered control authority in networks control the control plane. When a switch or device wants information they ask the controller.
* SCADA operates the reverse way (assumpiton) where the decision was made at the forward data plane and now we tell the data historian.
* Need to protect data plane at both device level and historian level
* **Reading ISA99, ISA95.** Could only get old versions from 2005-ish for ISA95 (ISA 99 is IEC 62264, newest version in WIP), but least supported version in modern day is ISA95.2000. I read ISA95.2005 and skimmed ISA99's working document.
  + ISA95 details specific object relational models to abstract machines within a DCS, but more importantly, specifies the business structure to operate a DCS. Doug talked about this as the different levels of a DCS in relation to what the machines have to communicate to (like reporting to level 4, the business level, for stats on how much a product is produced in tons).
  + A lot of people reference these slides: [http://www.apsom.org/docs/T061\_isa95-04.pdf.](http://www.apsom.org/docs/T061_isa95-04.pdf) ISA95 details the MES ([Manufacturing Execution System](https://en.wikipedia.org/wiki/Manufacturing_execution_system)) and ERP ([Enterprise Resource Planning](https://en.wikipedia.org/wiki/Enterprise_resource_planning)) specs of DCS. These two fields must be satisfied in the development of any solution to gain better security. The first one's main goal is "How do we optimize this process?" while the second asks "How much will this process cost?". ERP focuses more on the purchasing and managing of material in order to produce both the product and profit. MES focuses on the actual handling or material to produce profit within bounds acceptable to the ERP. This narrows the questions, problems, and solutions available to consider for the paper, especially when considering business level logic.
* **Focusing on the worst case.** Specifically focusing on the implementation of DCS in the pharmacy space. Represented by this [source](https://pdfs.semanticscholar.org/f730/d187866e06b8a4a6d245f81f4f1aaec05757.pdf) where they simulate a pharmacy DCS and hardware with DeltaV by Emerson.
  + Looking at the most complex case to protect brings "emergent properties" from a specific situation that can be generalized. Previous questions of "What data is to be protected?" becomes "What data is important to the MES in order to provide error correction via PID controllers? What reported data is important to the ERP? Is the data on logic circuits of the PLC important to protect?" Seeing a case example gives a better overall view of the data involved, the problem statement to address, and the solution to protect the data.
  + The specific case of pharmacy DCS mentions ISA88 "Recipe" data where simple logic is used to create data blocks to abstract and execute complex machine processes (aka instead of implementing a new process to heat up some complex chemical for creating medicine, businesses hire someone like Doug to take care of the engineering specifics and create a recipe). Considering the main Toray case example, this Recipe data is a key target for attackers and a key aspect to protect for datashares in the cloud where MES, ERP, Audit, and Malicious agents can access.
  + While I can understand what data computer scientists will deal with, another question that arose was what data did engineers handle and how to protect it. This source gave me ability to look at what they manage and eliminate some questions that did not need to be considered (but nevertheless write in a tracked document).
* **Acknowledging the evolving field of privacy.** I learned about the [McCumber Cube](https://en.wikipedia.org/wiki/McCumber_cube). And really, this [video](https://www.youtube.com/watch?v=bwCae2V4kmY) gives a good summary (just need to watch the beginning up till round 4 minutes). She is the [Executive Director of CIAC](https://www.uwb.edu/ciac/community/ciac-staff) at the University of Washington summing up the field of protecting information assurance and its murky definition. [Source](http://faculty.washington.edu/marcjd/articles/Systems%20Thinking%20Pedagogical%20Design.pdf) of original paper.
  + It seems that many security, privacy, and general computer science concepts come from the field of information theory, which in itself is discussed among many disciplines and journals such as ACM, IEEE, and AMC. But never explicitly established, just like the professor in the video says. It seems we're just starting to deal with the consequences of large information in the modern day and it exists in the background of research. Most research into privacy protected applications are in vehicular network applications.
  + For a long time, the CIA concepts of security encompassed the subject of privacy when it seems they are in fact distinct. Security seems to bias in control structures and singular solutions (aka cryptography, one input, one output). Privacy seems to bias in data structures and multiple solutions (aka one input, many outputs). (The singular vs multiple is a weak argument, but just take as for example). Learning about the McCumber cube explicitly details the problem space in the Transmission, Storage, and Processing of data —a space in which to explore solutions for this paper.
  + As a result, I have added 2 new potential solutions for the paper. [Private Information Retrieval](https://en.wikipedia.org/wiki/Private_information_retrieval) as given by this [source](https://www.tau.ac.il/~bchor/PIR.pdf) and [Fuzzing](https://en.wikipedia.org/wiki/Fuzzing) specifically related to [Fuzzy control systems](https://en.wikipedia.org/wiki/Fuzzy_control_system) and specific situations like [Fuzzy extractors](https://en.wikipedia.org/wiki/Fuzzy_extractor#Privacy_guarantees) to gain privacy sensitive data like fingerprints and transform them to protected signatures that can link them to specific key for user authentication. In summary:
    - PIR gives the two or more probability distributions in which private information is not explicitly given to a data collector but instead inferred with very strong probability. "I really have a high chance of having this property." This is executed in data structures such as [bloom filters](https://en.wikipedia.org/wiki/Bloom_filter) and some one-way hash functions.
    - Fuzzy control systems can interpolate values between 0 and 1 (or perhaps -1 to 1) continuous to give varying degrees of state. Kind of like a "Hot and Cold" game where if you're getting warmer, you are getting closer to the specified value.
    - [Differential privacy](https://en.wikipedia.org/wiki/Differential_privacy) is the other mentioned potential solution. It provides a theorem in which give an infinite amount of queries to any data collector, an attacker can never get the whole picture of data across a system (ie probabilities of all queried values), even when knowing the system itself (strong relation to cryptosystems).