This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

SAND2023-05186C

Proceedings of the 2023
30th International Conference on Nuclear Engineering
ICONE30
May 21-26, 2023, Kyoto, Japan

ICONE30-1708

USING THE INFORMATION HARM TRIANGLE TO MODEL SEQUENCES OF UNSAFE CONTROL ACTIONS IN INSTRUMENTATION AND CONTROL SYSTEMS

Lee T. Maccarone
Sandia National Laboratories
Albuquerque, NM

Andrew S. Hahn
Sandia National Laboratories
Albuquerque, NM

Michael T. Rowland
Sandia National Laboratories
Albuquerque, NM

ABSTRACT

The Information Harm Triangle (IHT) is an approach that seeks to simplify the defense-in-depth design of digital instrumentation and control (I&C) systems. The IHT provides a novel framework for understanding how cyber-attacks targeting digital I&C systems can harm the physical process. The utility of the IHT arises from the decomposition of cybersecurity analysis into two orthogonal vectors: data harm and physical information harm. Cyber-attacks on I&C systems can only directly cause data harm. Data harm is then transformed into physical information harm by unsafe control actions (UCAs) identified using Systems-Theoretic Process Analysis (STPA). Because data harm and physical information harm are orthogonal, defense-in-depth can be achieved by identifying control measures that independently limit data harm and physical information harm.

This paper furthers the development of the IHT by investigating the defense-in-depth design of cybersecurity measures for sequences of UCAs. The effects of the order and timing of UCAs are examined for several case studies to determine how to represent these sequences using the IHT. These considerations are important for the identification of data harm and physical information harm security measures, and they influence the selection of efficient measures to achieve defense-in-depth. This research enables the benefits of the IHT's simple approach to be realized for increasingly complex cyber-attack scenarios.

Keywords: Cybersecurity; Operational Technology; Instrumentation and Control; Industrial Control Systems; Systems-Theoretic Process Analysis; Risk

NOMENCLATURE

AIH	Apparent Information Harm
DH	Data Harm
IHT	Information Harm Triangle
PIH	Physical Information Harm
UCA	Unsafe Control Action

1. INTRODUCTION

The Information Harm Triangle (IHT) was initially proposed as a process that could merge cybersecurity and Systems Theoretic Process Analysis (STPA) analyses. The core of the IHT concept is a set of four postulates that rely upon the causal nature of Data Harm (DH) in initiating Unsafe Control Actions (UCAs) that transform DH to Physical Information Harm (PIH). The IHT can be applied to analyze the cybersecurity of operational technology (OT) systems, that is, systems that implement hardware and software to monitor and control a physical process. The Nuclear Power Plant (NPP) Pressurizer system was the initial example to demonstrate and challenge the validity of these postulates and develop the IHT concept.

This effort expands upon the initial works to investigate the order and timing of UCAs and its impact for both data and physical information harm. These analyses provided further analytic support of the IHT postulates as well as additional insights into the application of the IHT concept.

2. BACKGROUND

The following background information about the IHT is provided from [1, 2, 3]. The four key postulates of the IHT are quoted below from [3]:

1. Data can only be interpreted and understood by digital systems
2. Cyber-attacks can only directly cause DH
3. Physical consequences are directly caused by PIH
4. Cyber-attacks that result in physical consequences need an efficient transform function that converts DH to PIH

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

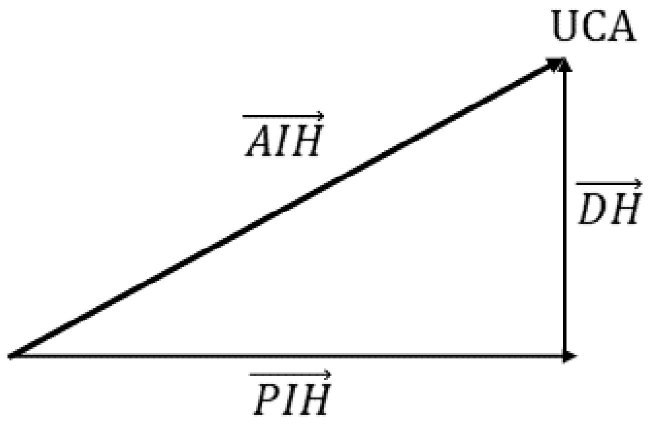

Figure 1: THE INFORMATION HARM TRIANGLE [1]

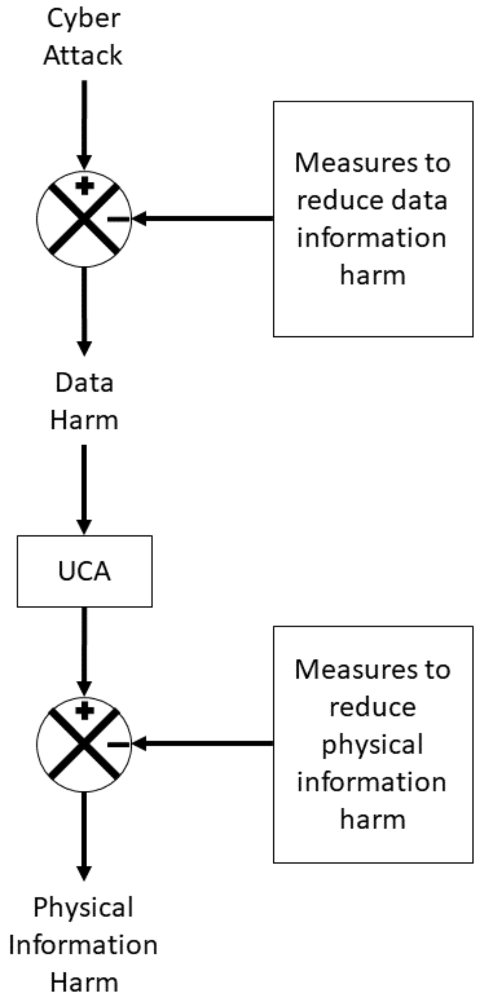
Cyber Attack
+
−
Measures to reduce data information harm
Data Harm
UCA
+
−
Measures to reduce physical information harm
Physical Information Harm

Figure 2: INFORMATION HARM TRIANGLE RELATIONSHIPS [1]

Since a cyberattack can only directly harm data and only OT systems can impact physical information, there is the potential for two types of orthogonal effects associated with a cyberattack. One type harms data, and the other harms physical information (i.e., information that exists independent of data in real time and space). Information harm is the deviation of information from its intended or true value. These effects are orthogonal because they exist in separate domains and cannot influence one another without a sufficient transform function.

When the magnitude of DH is sufficient to initiate a UCA, the UCA acts as a transform function, and PIH occurs. The risk of occurrence can be modified by measures that protect against DH or those that protect against PIH.

The IHT has the following three parts as shown in Figure 1.

1. PIH: Real plane (x-axis) representing harm that results in a physical hazard or loss, assuming there is a transform function (e.g., UCA) that can be initiated by a cyberattack
2. DH: Complex plane (y-axis) representing harm caused to data by a cyberattack
3. Apparent information harm (AIH): The sum of the orthogonal components that meet at a vertex representing the apparent harm generated by a UCA causing a harmful consequence. The UCA is represented as the upper vertex of the IHT.

A simplified sequence of harm effects from cyberattacks and the effects of protective measures are shown in Figure 2. The UCA transforms DH to PIH by causing changes to physical processes if DH thresholds are exceeded. Note that the simplified sequence does not show the potential for cyberattacks to harm measures.

The utility of the IHT has been demonstrated using several case studies. One example is the analysis of a pressurizer system [1, 2]. In this example, the cyberattack was altering a logic setpoint to turn off the heater while spoofing operator interfaces to hide the attack. This cyberattack resulted in the UCA of the pressurizer controller applying the energizing signal to the heaters after the pressure had reached the intended setpoint. The resulting PIH is the exceeding of the pressure boundary at the pressurizer instrument nozzles or heater sleeves. Two controls were implemented to reduce the PIH resulting from the cyberattack. The first control was a limit on DH imposed by restricting the capability to record the data needed to spoof the operator. The IHT for this scenario is shown in Figure 3. The second control was a limit on PIH imposed by implementing an overpressure relief valve. The IHT for this scenario is shown in Figure 4.

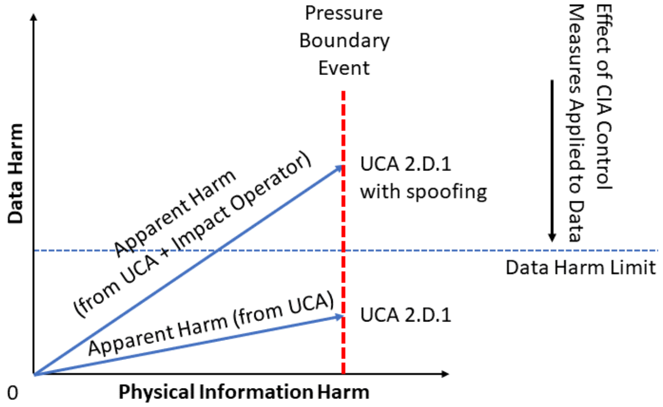

Figure 3. IHT DEMONSTRATING DH CONTROLS FOR PRESSURIZER EXAMPLE [1]

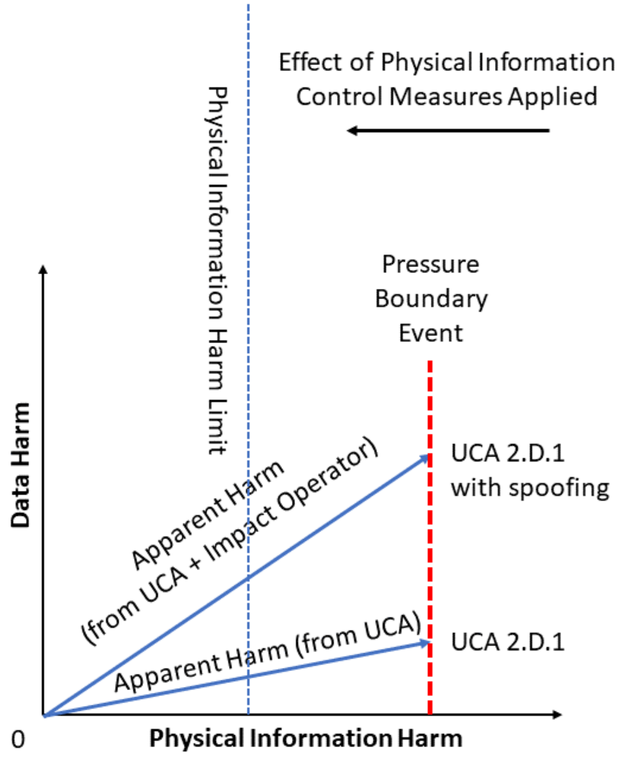

Figure 4. IHT DEMONSTRATING PIH CONTROLS FOR PRESSURIZER EXAMPLE [1]

The use case demonstrates that the IHT provides insights into defense in depth as DH and PIH are orthogonal to one another. This allows for the design and implementation of independent measures that have orthogonal effects and therefore will provide resilience from an attack that either harms data or physical information. Orthogonal security controls provide resilience because they limit PIH through independent means [1].

More importantly from this use case, the data (capability constraint) and physical information (overpressure relief valve) security controls provide for two independent measures that both guard against the UCA plus spoofing whereas UCA 2.D.1 is prevented by the overpressure relief valve [1].

3. METHODS

This section details the experimental methods used in this analysis. The following paragraphs regarding MiniMega, PHENIX, Asherah hypothetical Pressurized Water Reactor (PWR) simulator, and PLC emulation are quoted from [4].

These experiments were performed using a simulation and evaluation system developed at Sandia National Laboratories. The research platform was created using a Sandia-developed emulation suite (e.g., MiniMega, PHENIX) to recreate the control network of a nuclear power plant, coupled with the Asherah PWR simulator and Siemens virtual programmable logic controllers (PLCs) [5, 6]. This platform enables the simulation of multi-node tiered networks by emulating network components (switches, firewalls, access points) and the connections between them. The virtual PLCs (vPLCs) allow implementation of dynamic control logic that responds as physical conditions change in the plant. This dynamic virtual analysis can be scaled to mimic an entire plant control structure and in addition hardware PLCs can be added in the loop to validate real PLC performance for commissioning [4].

The Sandia-developed environment allows the simulation of large-scale networks and their components [7] and is built upon another Sandia developed Virtual Machine (VM) deployment suite, MiniMega [8]. MiniMega is an open-source tool that allows the generation of virtual networks and the deployment of VMs on that virtual network. The environment expands on the capabilities of MiniMega and provides more refined environment development controls and better physics integration support. PHENIX, a key component of this environment, is now openly available to the public (i.e., open source) [9]. It deploys VMs that are stored on large file servers and then ported to operate on high performance computing servers. The virtual networks are built by reading from storage and then populating into a predesigned virtual network. These designs constitute the network hardware such as switches and routers as well as the controllers, sensors, actuators, and in our case, virtualized PLC's and their control logic, and physics simulators [4].

The physics are simulated by the Asherah PWR Simulator [5]. Sensor signals and control signals are routed via the management network to VMs of smart sensors and actuators. The Asherah PWR simulator provides a high-fidelity physics simulation of a nuclear power plant for the network and controllers to interact with one another. The controllers within the Simulink model provide a template of the vPLCs to be later created on our network. As vPLCs are programmed, the controllers in the Simulink model are commented out and the emulated PLCs take over [4].

PLCs are somewhat difficult to fully emulate currently. Few manufactures of hardware that would be found in a real plant have dedicated the effort to develop emulations of their products. Emulation of hardware that is dependent on the accuracy of time measurement is a difficult task on modern computers. Siemens offers the PLCSIM Advanced emulation software which allows Siemens S7-1500 PLCs to be virtually emulated [10]. High fidelity emulations of PLCs can be run from Windows 10 VMs and provides accurate virtual representations of real PLC hardware.

This simulation platform was used to simulate the effects of two UCAs on plant physics. These UCAs were identified by evaluating the steam generator pressure and reactor coolant pump controllers using Systems-Theoretic Process Analysis (STPA) [11]. The UCAs are shown in Figure 5 and listed below:

1. UCA 1: Send reactor coolant pump decrease speed command when decrease in speed is not needed. This UCA is implemented in the simulation by setting the input value to the PID controller as 11,000 kg/s when the desired setpoint is 8,800 kg/s. This causes the PID controller to decrease the pump flow rate to try to reach the setpoint [4].

2. UCA 2: Freeze steam generator throttle position during transient. This UCA is implemented in the simulation by setting the output value to the turbine steam isolation as the previous value (i.e., a constant value). This causes the actuation valve to freeze and initiate rapid depressurization [4].

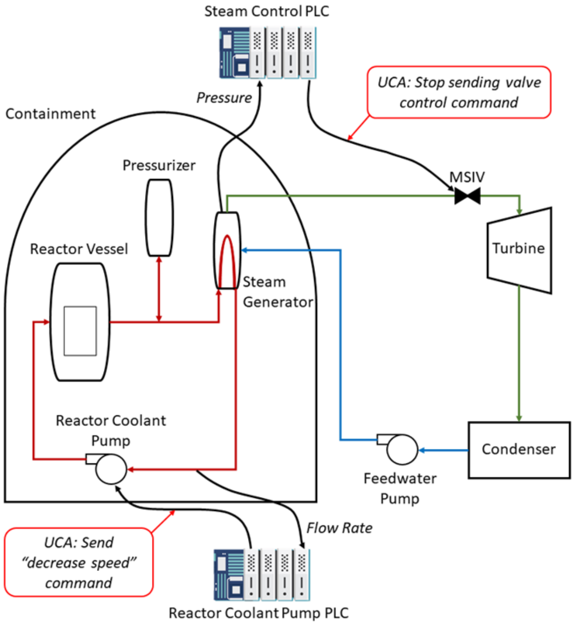

Figure 5. STEAM GENERATOR PRESSURE AND REACTOR COOLANT PUMP CONTROLLERS

These UCA were simulated at varying time intervals. The effects of the UCAs were analyzed in terms of the pressure in the steam generator (P_{SG}). The baseline P_{SG} is 6.4 MPa. The following section presents the analysis of four simulation cases using the IHT.

4. ANALYSIS AND RESULTS

Four simulations were examined using the IHT. The first case is the simulation of only UCA 1 and the second case is the simulation of only UCA 2. The results from these simulations are shown in Figure 6 and Figure 7, respectively. In both cases, the individual UCAs were initiated at a time of 500 seconds. The third case was the simulation of simultaneous UCAs as shown in Figure 8. In this case, both UCAs were initiated at a time of 500 seconds. The fourth case was the simulation of sequential UCAs as shown in Figure 9. In this case, UCA 1 was initiated at 500 seconds and UCA 2 was initiated at 1400 seconds.

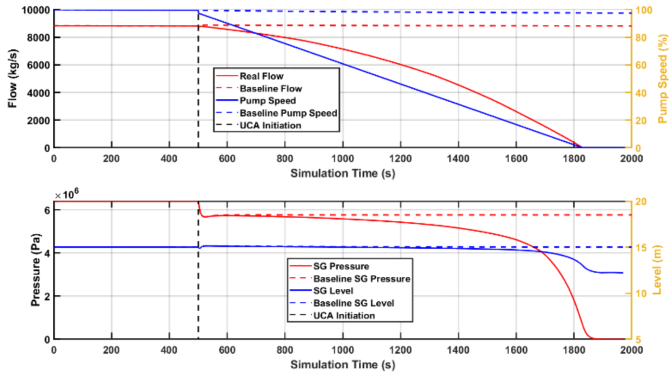

Figure 6. SIMULATION OF UCA 1

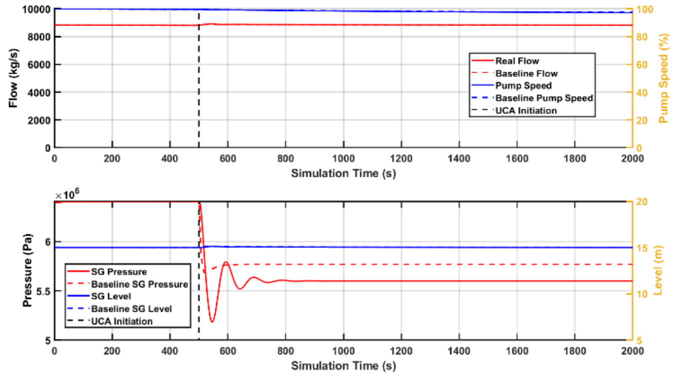

Figure 7. SIMULATION OF UCA 2

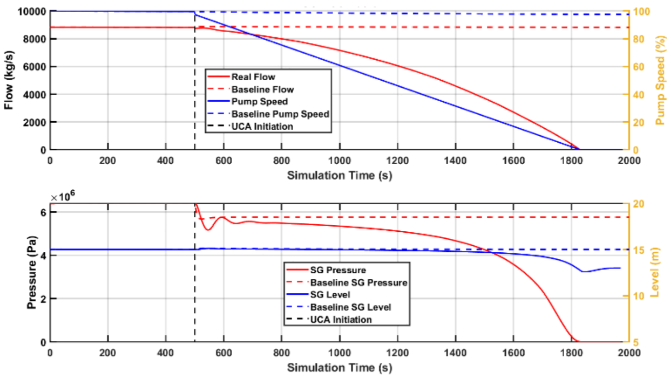

Figure 8. SIMULATION OF SIMULTANEOUS UCAS [4]

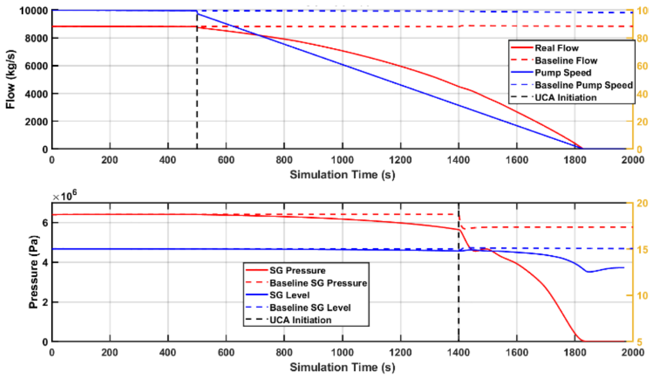

Figure 9. SIMULATION OF SEQUENTIAL UCAS [4]

The figures of simulation data each include two sets of axes plots. The first set of axes in each figure corresponds to the flow rate of the first reactor coolant pump and the second set of axes corresponds to the pressure of the first steam generator. The dashed vertical lines indicate the initiation of a UCA. Baseline data is given by the dashed plots and corresponds to the expected behavior. Scenario data is given by the solid plots and corresponds to the data obtained from the combined PLC emulation and physics simulation. The scenario data will be used in this work. The relevant P_{SG} data is summarized in Table I and Table II.

Table I. SIMULATION DATA FOR INDIVIDUAL UCAS

Data	**UCA 1**	**UCA 2**
Final P_{SG}	0.0 MPa	5.6 MPa
Baseline P_{SG} – Final P_{SG}	6.4 MPa	0.8 MPa
Time from UCA to Final P_{SG}	1,360 s	266.7 s
Average Rate of Change of P_{SG} from UCA to Final P_{SG}	4.74E-03 MPa/s	3.00E-03 MPa/s

Table II. SIMULATION DATA FOR MULTIPLE UCAS [4]

Data	**Simultaneous UCAs**	**Sequential UCAs**
Final P_{SG}	0.0 MPa	0.0 MPa
Baseline P_{SG} – Final P_{SG}	6.4 MPa	6.4 MPa
Time from UCA 1 to Final P_{SG}	1,412 s	1,360 s
Time from UCA 2 to Final P_{SG}	1,412 s	420 s
Average Rate of Change of P_{SG} from UCA 1 to UCA 2	N/A	9.64E-04 MPa/s
Average Rate of Change of P_{SG} from UCA 2 to Final P_{SG}	4.57E-03 MPa/s	1.21E-02 MPa/s
Average Rate of Change of P_{SG} from UCA 1 to Final P_{SG}	4.57E-03 MPa/s	4.74E-03 MPA/s

The DH occurring in these simulations can be calculated in terms of the DH to change a setpoint, DH_{SP}, the operating frequency of the PLC, f, and the duration of time over which the setpoint is changed, Δt, using the following equation.

$$DH = DH_{SP} f \Delta t \quad (1)$$

Assuming DH_{SP} and f are the same for the pump and valve PLCs, the DH is given in Table III and Table IV.

Table III. CALCULATED DH DATA FOR INDIVIDUAL UCAS

Data	**UCA 1**	**UCA 2**
Total DH(UCA)	$1360DH_{SP}f$	$267DH_{SP}f$
Rate of DH From UCA to Final P_{SG}	$DH_{SP}f$	$DH_{SP}f$

Table IV. CALCULATED DH DATA FOR MULTIPLE UCAS

Data	**Simultaneous UCAs**	**Sequential UCAs**
DH(UCA 1) Until UCA 2	N/A	$900DH_{SP}f$
Total DH(UCA 1)	$1412DH_{SP}f$	$1360DH_{SP}f$
Total DH(UCA 2)	$1412DH_{SP}f$	$420DH_{SP}f$
Total DH(UCA 1 & UCA 2)	$2824DH_{SP}f$	$1780DH_{SP}f$
Rate of DH from UCA 1 to UCA 2	N/A	$DH_{SP}f$
Rate of DH From UCA 2 to Minimum P_{SG}	$2DH_{SP}f$	$2DH_{SP}f$
Average Rate of DH from UCA 1 to Minimum P_{SG}	$2DH_{SP}f$	$1.31DH_{SP}f$

The IHTs for all of the simulations are given in Figure 10. In this figure, the DH is expressed in terms of total DH and PIH is expressed in terms of the maximum deviation of P_{SG} from the baseline value of 6.4 MPa. The IHTs for the individual UCAs are given by the black dashed vectors. UCA 1 causes a much greater amount of PIH than UCA 2, but also requires more total DH to achieve the maximum PIH. The IHT for the simultaneous UCAs is given by the solid blue vector. This case causes the same amount of PIH as UCA 1 but requires much more DH because two UCAs are active throughout the entire attack. It can be seen from Figure 10 that if the adversary were to use both UCAs, the sequential UCA case would be more desirable for an adversary than the simultaneous case. The PIH is identical for the simultaneous and sequential cases because P_{SG} reaches zero for both cases, but the sequential case requires less DH. One challenge for a potential adversary in this scenario is that sequencing UCAs requires an element of coordination that is not required when arbitrarily initiating UCAs (although this coordination would also be required if the adversary sought to initiate the UCAs perfectly simultaneously).

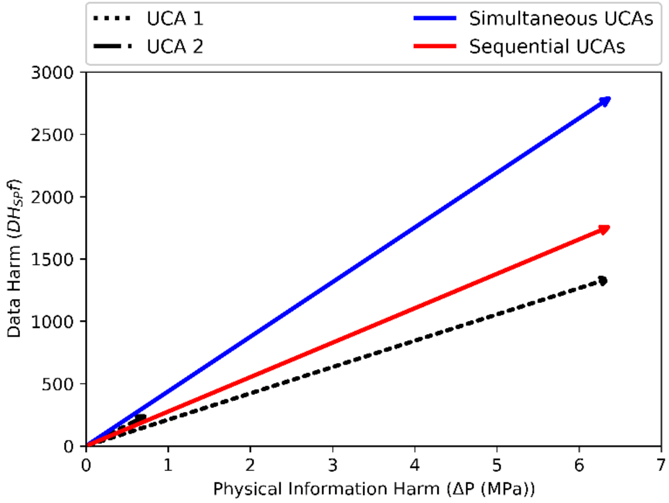

Figure 10. IHT FOR ALL UCA SIMULATION CASES USING CALCULATED TOTAL DH AS DH AND CHANGE IN P_{SG} AS PIH

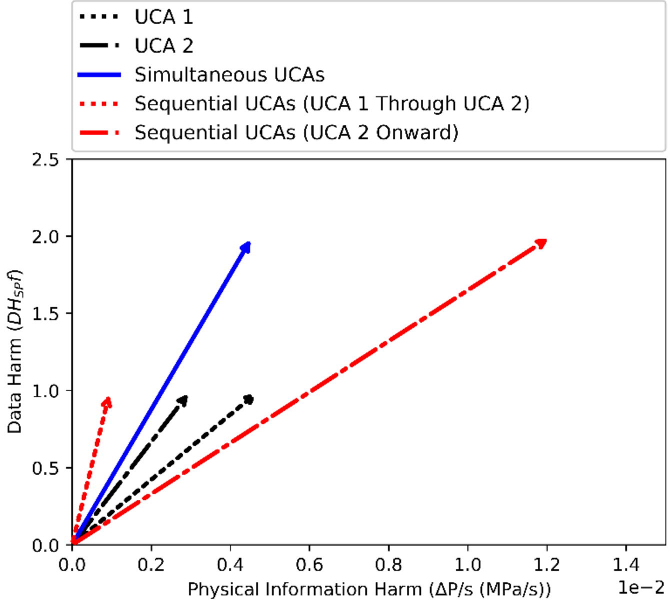

Figure 11. IHT FOR ALL UCA SIMULATION CASES USING CALCULATED RATE OF DH AS DH AND RATE OF CHANGE OF P_{SG} AS PIH

One potential method for considering cyber-attack "efficiency" is to calculate the ratio of DH to PIH. On the IHT, this metric is the tangent of the angle between AIH and PIH, or the slope of the vectors in Figure 10. It can be seen that the simultaneous UCA case is the least efficient attack by this metric because it has the greatest slope. UCA 2 and the sequential UCA case have nearly identical efficiencies and the UCA 1 has the greatest efficiency (i.e., smallest slope).

An IHT can also be constructed to show the rates of DH and PIH for P_{SG}. This IHT is shown in Figure 11. Note that there is one IHT corresponding to the simultaneous UCA case and there are two IHTs corresponding to the sequential UCA case. This is because there are two rates of change for the sequential case: one rate when only UCA 1 is active, and one rate when both UCA 1 and UCA 2 are active. It is also noteworthy that the UCA 1 IHT and sequential UCAs (UCA 1 through UCA 2) IHTs are different because the rates are calculated as average rates of change over the specified time. The UCA 1 duration is greater than the sequential UCA (UCA 1 through UCA 2) duration because the UCA 1 simulation continued until the final value of P_{SG} was reached, while the duration of UCA 1 within the sequential case was limited by the initiation of UCA 2. This IHT demonstrates that changes in P_{SG} are much more rapid in the sequential case after UCA 2 is initiated than in the simultaneous case, even though the rate of DH is identical.

5. CONCLUSION

This analysis demonstrated how the IHT can be used to represent scenarios where the timing of UCAs impacts the PIH resulting from a cyberattack. IHT representations of the attack were created for the magnitude of harm and rates of harm to demonstrate the insights that can be gained from both forms of the IHT. Future work will examine additional methods for measuring DH and PIH, and examine the effects of controls on DH and PIH to enable defense-in-depth security for combinatory UCAs.

ACKNOWLEDGEMENTS

This article has been authored by an employee of National Technology & Engineering Solutions of Sandia, LLC under Contract No. DE-NA0003525 with the U.S. Department of Energy (DOE). The employee owns all right, title and interest in and to the article and is solely responsible for its contents. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this article or allow others to do so, for United States Government purposes. The DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan https://www.energy.gov/downloads/doe-public-access-plan.

REFERENCES

[1] M. T. Rowland, L. T. Maccarone and A. J. Clark, "Using the Information Harm Triangle to Identify Risk-Informed Cybersecurity Strategies for Instrumentation and Control Systems," *Nuclear Technology,* 2022.

[2] M. T. Rowland and A. J. Clark, "Application of the Information Harm Triangle to Inform Defensive Strategies for the Protection of NPP I&C Systems," in *ANS NPIC&HMIT*, Virtual Online Meeting, 2021.

[3] M. T. Rowland, "Investigation of Data Harm and its Relevance to Unsafe Control Actions of Control Systems through Application of the Information Harm Triangle," University of London, London, UK, 2022.

[4] A. Hahn, D. R. Sandoval, R. E. Fasano and C. C. Lamb, "Automated Cyber Security Testing Platform for Industrial Control Systems," in *ANS NPIC&HMIT*, Virtual Online Meeting, 2021.

[5] B. E. Silva, R.A, R. Shirvan, J. Piqueira and R. Marques, "Development of the Asherah Nuclear Power Plant Simulator for Cyber Security Assessment," in *IAEA International Conference on Nuclear Security*, Vienna, Austria, 2020.

[6] Sandia National Laboratories, "SCEPTRE," Albuquerque, NM, 2016.

[7] N. Jacobs and J. Johnson, "SCEPTRE: Power System and Networking Co-Simulation Environment," Sandia National Laboratories, Albuquerque, NM, 2018.

[8] J. Crussell, J. Erickson, D. Fritz and J. Floren, "minimega v3.0," Sandia National Laboratories, Albuquerque, NM, 2015.

[9] Sandia National Laboratories, "PHENIX Documentation," Albuquerque, NM, 2021.

[10] Siemens, "PLCSIM Advanced V3.0," 2021.

[11] N. G. Leveson and J. P. Thomas, "STPA Handbook," 2018.