txOne networks

Whitepaper June 2024

Securing Semiconductor Manufacturing:

Tackling Cybersecurity Challenges in SECS/GEM Protocols

Keep the Operation Running

Contents

Executive Summary

As the global demand for electronics continues to surge, the semiconductor market has reached unprecedented revenue levels. The industry's value chain is comprised of four critical stages: Integrated Circuit (IC) Design, Photomask Production, Wafer Fabrication, and Packaging. Each stage presents unique cybersecurity challenges that need careful consideration.

Building on previous research highlighted in "Potential Threats to Semiconductor Processes",[1] this paper delves deeper into the specific protocols used during the wafer fabrication and packaging phases. This research will examine the underlying processes of the semiconductor industry and scrutinize the design of the key protocols involved.

Key Risks Identified

Cybersecurity Design Flaws

Due to a lack of robust cybersecurity measures in some semiconductor protocols, there is a significant risk of Man-in-the-Middle (MITM) attacks. These attacks could impair process control or inhibit response functions, thereby disrupting operations.

Vulnerability in Interface Designs

The varying designs of interfaces in semiconductor controllers can be exploited through specially crafted files, potentially leading to Denial of Service (DoS) attacks. This risk emphasizes the need for stringent validation and security checks on file inputs across all interfaces.

Remote System Access

Protocols that allow remote access to systems outside of the physical factory increase the risk of unauthorized entry into regulated Operational Technology (OT) environments. Implementing secure remote access protocols and continuous monitoring is essential to mitigate these threats.

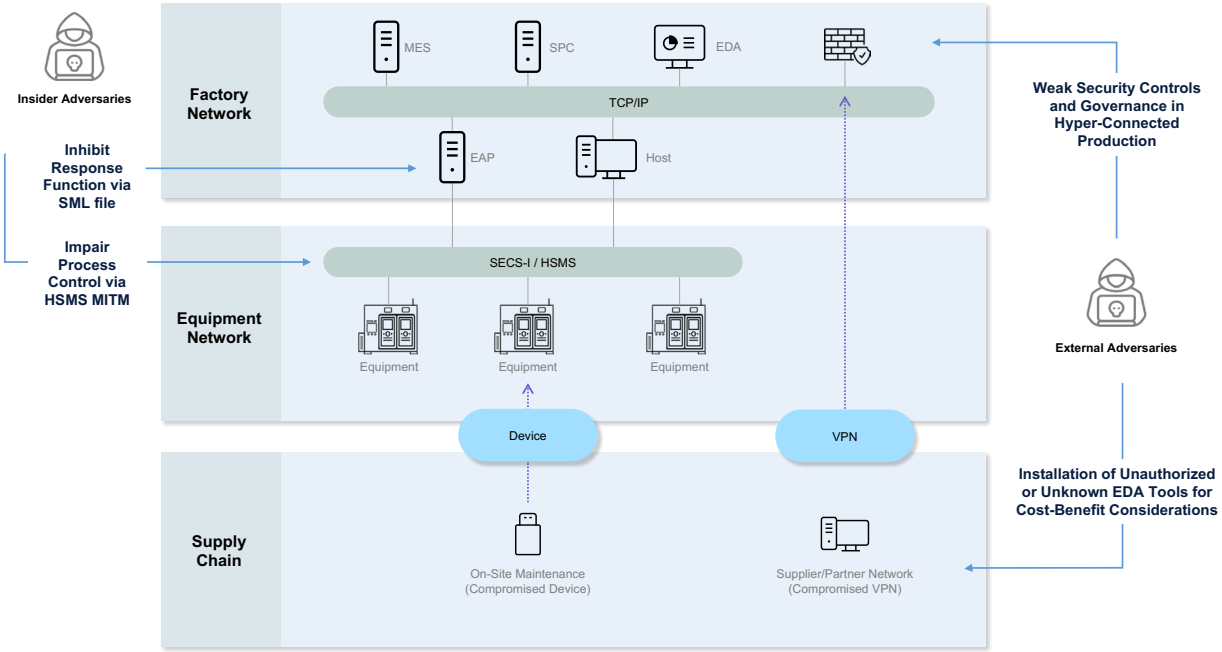

Figure 1. Cyberattack Pathways in the Semiconductor Industry

01 The Value Chain of the Semiconductor Industry

The main stages in the semiconductor industry consist of IC design, photomask production, wafer fabrication, and packaging. Each of these stages is crucial in the semiconductor workflow, and together they enable the production of devices used in a wide range of electronic applications. Due to stage specialization in semiconductors, many supply chains have developed to support the vertical's unique needs. A breakdown of these stages is as follows:

- **Integrated Circuit Design**

  This is the initial stage of semiconductor manufacturing. Engineers employ specialized software tools known as Electronic Design Automation (EDA) to design the layout and functionality of integrated circuits. IC design defines the arrangement of components, interconnections, and circuit functions on the chip.

- **Photomask Production**

  Photomask production is the transition stage between IC design and wafer fabrication. Advanced photolithography techniques are used to create these high-precision templates, transferring the circuit patterns onto a photomask using photoresist layers and exposure to light, and then etching or developing the pattern on the mask. The photomask then acts as a stencil for patterning the semiconductor materials on the wafer during the fabrication process.

- **Wafer Fabrication**

  Wafer fabrication is the process of transferring the designed patterns onto silicon wafers, using the photomasks to define the circuit patterns on the wafers. This includes transferring the lithographic patterns onto the wafer surface, depositing materials onto the wafer, etching, and other processing steps to form the structure of integrated circuits. This stage builds the IC using the photomasks as a template

- **Packaging**

  Packaging is the final stage in semiconductor production. After wafer fabrication is complete, individual chips need to be packaged to form complete integrated circuits (ICs). Packaging involves mounting the chips into packages, connecting leads, encapsulating chips to protect them, and providing electrical connections and physical protection for the ICs.

By understanding these stages and the technologies involved, this research gains insight into the complexity and vulnerabilities of the semiconductor supply chain, reaffirming the importance of cybersecurity and resilient production processes.

01

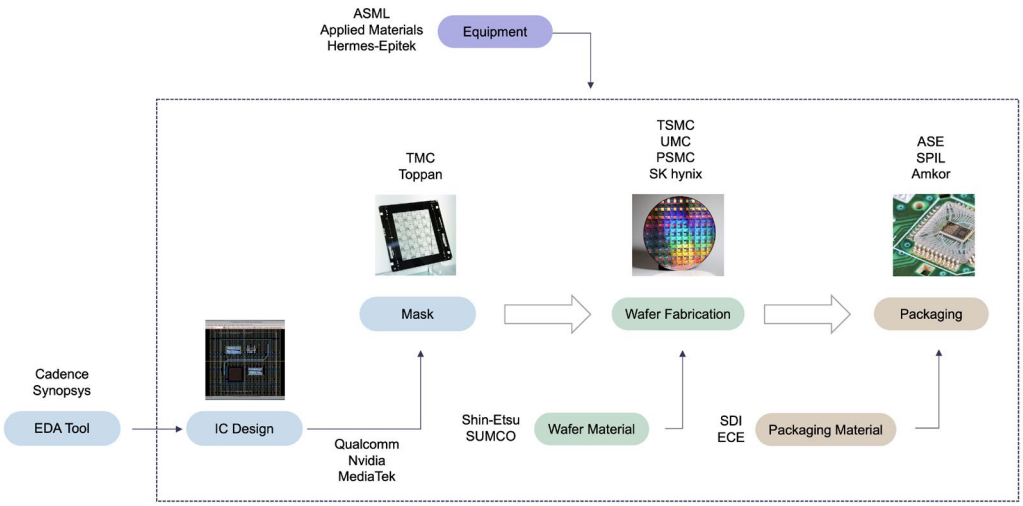

Figure 2. Core Processes in the Semiconductor Industry

More specifically, as can be seen in Figure 3, each stage also has its respective complex processes. In wafer fabrication, several processes are traditionally involved in the manufacturing of semiconductor wafers. These processes, including Oxidation, Diffusion, Deposition, Photolithography, and Etching are carried out sequentially in a cycle to create the necessary structures and patterns on the wafer surface. Since all these processes require specialized action, they have their own equipment, which may come from different suppliers. However, to exchange messages between equipment, independent manufacturers need a way to produce implementations that can connect and interoperate without requiring specific language of one another. HSMS is a common example of communication protocols used in semiconductor factories. Notably, during the wafer fabrication and packaging stages, most equipment optionally supports these protocols. This whitepaper will focus on explaining cyber threats to HSMS.

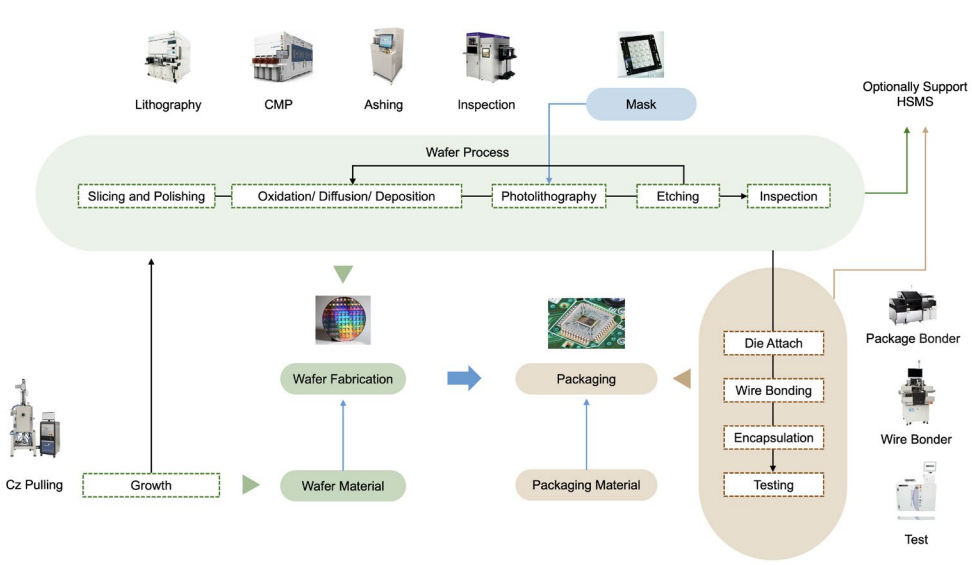

Figure 3. Key Stages of the Semiconductor Production Process

02

Protocols in the Semiconductor Industry: Critical Roles and Security Challenges

In the bygone era of traditional manufacturing, factories operated within regulated environments, rendering cybersecurity concerns secondary. However, the evolving landscape of Information Technology (IT) and Operational Technology (OT) convergence has expanded the attack surface, granting adversaries more opportunities for initial access breaches. This convergence illuminates a critical concern: successful penetration into an OT environment enables attackers to disrupt process control systems or compromise response functionalities via outdated protocols, posing significant risks to manufacturing operations.

The manufacturing sector relies on various common protocols for machine-to-machine communication, such as Modbus, Message Queuing Telemetry Transport (MQTT), Open Platform Communications – Unified Architecture (OPC-UA), Constrained Application Protocol (CoAP), Data Distribution Service (DDS), and Semiconductor Equipment Communications Standard/Generic Equipment Model (SECS/GEM), among others. TXOne Networks has conducted comprehensive analyses on the vulnerabilities inherent to these protocols, as detailed in their series of reports on OPC-UA, DDS, MQTT, and protective strategies for utilities.[2] [3] [4] [5]

The focus of this paper narrows to the SECS/GEM, a protocol predominantly utilized within the semiconductor industry by leading corporations including Intel, Samsung, TSMC, IBM, Qualcomm, Broadcom, UMC, SK Hynix, Micron, Texas Instruments, Toshiba and so on. Semiconductors, the cornerstone of modern electronics, are pivotal for economic growth, national security, and international competitiveness. This criticality renders the semiconductor sector a lucrative target for cybercriminal activities.

In response to the escalating threat landscape, Semiconductor Equipment and Materials International (SEMI), the global industry association representing the semiconductor sector, has spearheaded the development of robust cybersecurity frameworks encompassing Standards, Assessments, and a Cybersecurity Architecture. A landmark achievement in this initiative was the November 2023 unveiling of the “Cybersecurity Reference Architecture for Semiconductor Manufacturing Environment”.[6] This comprehensive framework provides a blueprint for securing semiconductor foundries, by illustrating the essential components and strategies for achieving resilient cybersecurity postures within the industry.

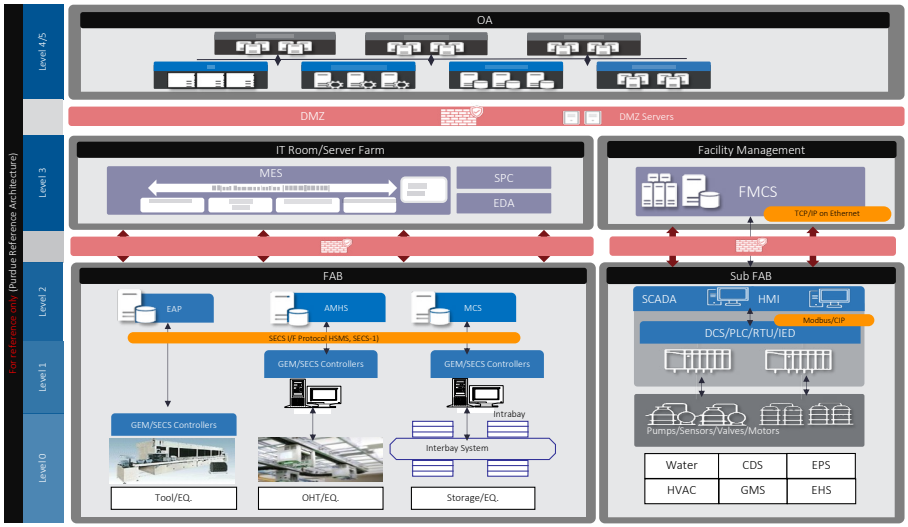

Figure 4. Reference Architecture of a Semiconductor Foundry

03 Understanding the SECS/GEM Protocol

Modern semiconductor factory automation epitomizes complex systems engineered to enhance efficiency throughout the manufacturing process. This integration of mechanical and electronic automation technologies encompasses an array of equipment—ranging from lithography, chemical vapor deposition, and plasma etching, to ion implantation machinery. These components work in concert to facilitate seamless communication, control, and monitoring of manufacturing operations.

Central to these operations is the Fabrication (FAB) area, a nexus of the production process that includes stages such as wafer manufacturing, chip processing, packaging, and testing. Within this domain, SECS/GEM protocol plays a crucial role in ensuring communication coherence between diverse equipment and factory systems. This protocol fosters interoperability, a pivotal feature for integrating equipment from various suppliers into a cohesive factory system. Given this critical role, understanding the cyber vulnerabilities associated with the SECS/GEM protocol is paramount.

SECS/GEM operates as a point-to-point protocol where equipment, typically set in a passive mode, connects with a single host at a time. This protocol encompasses several standards, including E4, E5, E30, and E37, each serving distinct functions within the communication framework. For clarity:

- E30 (GEM) outlines the application of SECS-II messages, monitoring equipment behavior during message exchanges with the host.
- E5 (SECS-II) facilitates information exchange between equipment and host using a structured format of streams and function messages.
- E4 (SECS-I) employs RS-232 cables for communication, a foundational layer in the protocol suite.
- E37 (HSMS) upgrades communication to TCP/IP for enhanced reliability and speed.

```
S1F1 W
<L[2]
  <F8[1] 0.900000>
  <B[5] 0x00 0x05 0x06 0x09 0xFF>
>
```

Figure 5. A Sample E5 (SECS-II) Message

To illustrate, consider a sample E5 (SECS-II) message: 'S' denotes Stream, and 'F' signifies Function, followed by <L[2]>, indicating a list containing two elements. Within this structure, <F8[1] 0.900000> represents a floating-point value, and <B[5] 0x00 0x05 0x06 0x09 0xFF> specifies binary data in a five-byte sequence.

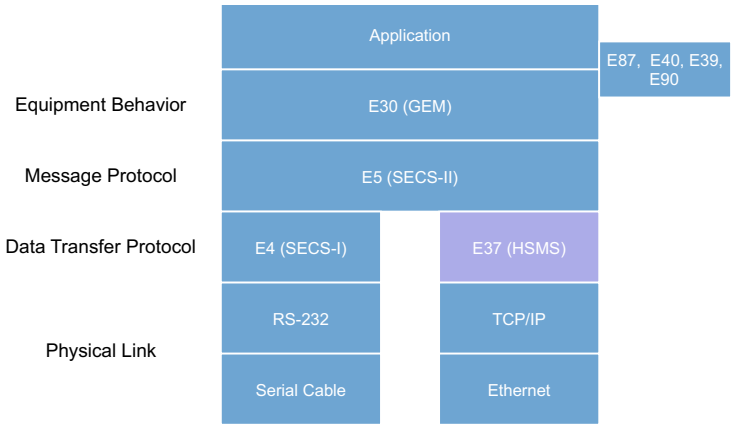

Figure 6. The SECS-GEM Relationship

In the realm of semiconductor manufacturing, the evolution of communication protocols has significantly impacted operational efficiency and cybersecurity. The E4 (SECS-I) protocol, characterized by its slow data transmission rate via RS-232, is ill-suited for applications requiring long-distance communication or robust noise immunity. Due to its limitations, this protocol, now considered legacy, is no longer incorporated into new factory equipment.

In contrast, the E37 High Speed SECS Message Services (HSMS) protocol, operating over TCP/IP, emerges as a superior alternative to the SEMI E4 (SECS-I). It caters to the needs of modern semiconductor plants where higher speed communication and complex network topologies surpass the capabilities of a simple point-to-point connection. This transition is part of a significant shift towards more sophisticated, efficient communication frameworks in contemporary manufacturing environments.

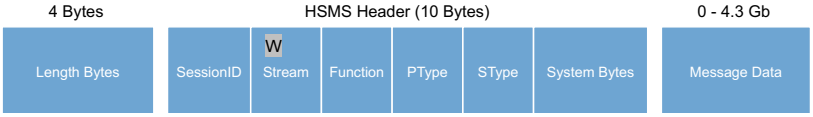

Figure 7. E37 (HSMS) Structure

03

E37 (HSMS) distinguishes itself from traditional TCP/IP protocols by introducing a nuanced approach to connection establishment, specifically delineating Active and Passive Modes for initiating communication (refer to Figure 8). This distinction ensures that only devices in Active Mode can make a request to establish a connection, fostering a secure and orderly communication landscape. Once a link is forged between the host and equipment, HSMS facilitates the transmission of binary encoded E5 (SECS-II) messages, enabling a wide array of functions such as recipe changes, monitoring, and reporting, which are crucial for the seamless operation of semiconductor manufacturing equipment. Several key aspects of HSMS operation merit attention:

- Given that connections within a semiconductor network often remain active for extended periods, it is essential to preemptively test the communication link with Linktest requests to ensure reliability.
- The protocol designates streams numbered 1-127, each capable of supporting functions numbered 1-255. While streams 1-63 (and their respective functions) are reserved for standard operations, the remaining identifiers allow for the implementation of custom, user-specific messages.
- The "W" flag in the message header signifies the necessity of a response, indicating a more interactive, responsive communication framework.
- The encapsulation of messages using a binary encoding process obfuscates the data, enhancing security.
- Typically, the established connection persists until deliberately terminated for reasons such as software upgrades, equipment removal, or maintenance. A "Separate" request can be issued by the host to formally conclude the communication.

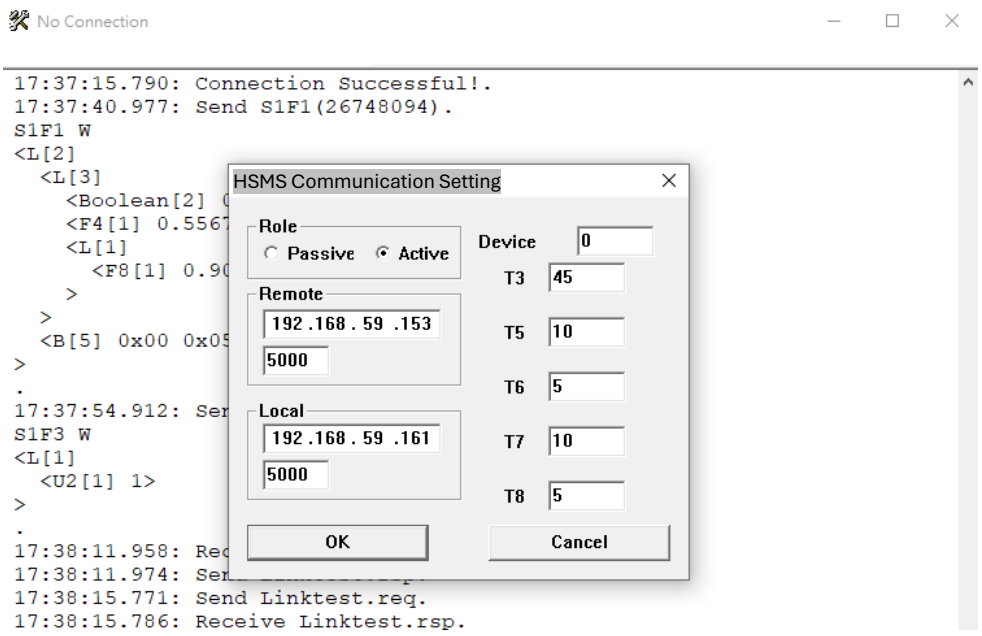

Figure 8. Active Mode and Passive Mode

In the context of semiconductor manufacturing, the SECS/GEM protocols, although foundational, are part of a legacy suite that includes E37 (HSMS), which lacks inherent security features. This deficiency becomes apparent in the transmission process, where payloads are sent in plaintext. Consequently, individuals with rudimentary understanding of binary encoding and the SECS/GEM framework can readily intercept and decipher these communications, exposing sensitive data to potential cyber threats.

Consider the scenario depicted in Figure 9: a SECS/GEM controller operates as the host in active mode, orchestrating communication with various equipment set in passive mode via a starting TCP port. The crucial flaw here lies in the exchange of SECS/GEM messages devoid of any form of integrity verification. Such a lapse in security protocols presents a glaring opportunity for cyber attackers. They could exploit this vulnerability to disrupt the standard operations and functionalities of the equipment.

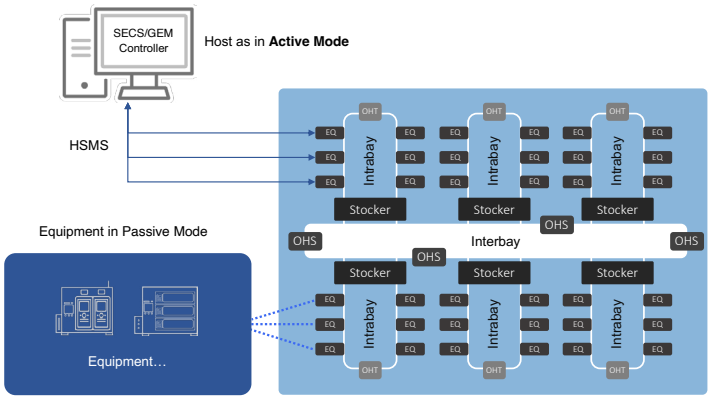

Figure 9. Interbay System in the FAB Production Area

Scenario 1: Compromising Process Control Through Limited Connection Protocols

The HSMS protocol, a cornerstone in semiconductor manufacturing automation, mandates that equipment must accommodate at least one active connection. However, it does not facilitate the ability for equipment to handle multiple concurrent hosts. In essence, once equipment establishes a connection with a host, any attempt by the same or a different host to initiate a secondary connection is instantly rebuffed, resulting in the termination of this subsequent connection attempt.

This operational limitation presents a tactical vulnerability. For an adversary aiming to disrupt communication between the host and the equipment, orchestrating a Man-in-the-Middle (MITM) attack becomes a go-to strategy, as depicted in Figure 10. By positioning themselves within the communication channel, attackers can exploit the protocol's single-connection limitation to intervene and potentially control the equipment's process.

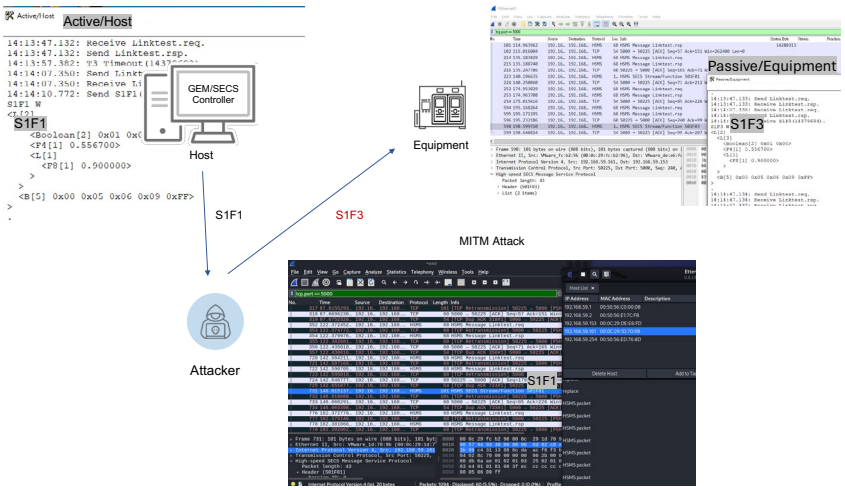

Figure 10. Man-in-the-Middle Attack Modifying Functionality

03

An attacker with the capability to modify communications between the host and the equipment poses a significant threat to semiconductor manufacturing processes. By intercepting and altering requests from the host, the attacker can issue unauthorized command messages to the control systems. Such commands may instruct devices to perform actions beyond their standard operational parameters, leading to potential disruptions in the manufacturing process. As illustrated in Figure 11, this breach grants attackers the opportunity to manipulate materials or alter processes, adversely affecting the wafer production line.

Notably, the SECS-II (E5) protocol accommodates user-defined streams and functions, which can vary significantly across different equipment. This variability introduces a spectrum of potential impacts, ranging from minor disruptions to severe operational failures. While the consequences may not mirror those of a compromised electrical substation—which could result in widespread power outages—affecting thousands of businesses and households, the stakes in semiconductor manufacturing are no less critical. Semiconductors are the backbone of numerous key industries, including communications, automation, artificial intelligence, healthcare, military applications, smart transportation, and clean energy solutions. A failure in the wafer process not only results in lost productivity and revenue but also carries major geopolitical implications.

- Stream 0 (Function 0) - Close Transaction
- Stream 1 - Equipment Status
- Stream 2 - Equipment Control and Diagnostic
- Stream 3 - Material Status
- Stream 4 - Material Control
- Stream 5 - Exception Handling
- Stream 6 - Data Collection
- Stream 7 - Process Program Management
- Stream 8 - Control Program Transfer
- Stream 9 - System Errors
- Stream 10 - Terminal Services
- Stream 11 - Host File Services (Deleted)
- Stream 12 - Wafer Mapping
- Stream 13 - Data Set Transfers
- Stream 14 - Object Services
- Stream 15 - Recipe Management
- Stream 16 - Processing Management
- Stream 17 - Equipment Control and Diagnostics
- Stream 18 - Subsystem Control and Data

User Definable
Streams 1-63, Functions 64-255
Streams 64-127, Functions 1-255

Furthermore, attackers can exploit this vulnerability to spoof reporting messages through MITM attacks. For instance, if equipment attempts to alert the host that a critical temperature threshold has been exceeded, the attacker could manipulate the message to falsely report normal operating conditions. This is achieved by altering the message's stream and function codes within the SECS-II framework, potentially leading to the omission of necessary corrective actions.

Scenario 2: Disrupting Communication with Man-in-the-Middle Attacks

Previously, this whitepaper highlighted that the HSMS protocol allows for a 'Separate' request, which effectively terminates the communication between the host and equipment. Exploiting this feature through a MITM attack, an adversary can sever the ongoing communication by issuing this 'Separate' request. Subsequently, the attacker can establish a new connection with the equipment, precluding any attempts by the legitimate host to reconnect. Given the protocol's design, which permits only a single active connection at any given time, the equipment will disregard attempts from the legitimate host to re-establish communication so long as the attacker maintains control.

This scenario, illustrated in Figure 12, underscores a critical vulnerability in the HSMS protocol's operation. By hijacking the communication channel, the attacker effectively isolates the equipment from its legitimate host. Such an attack not only highlights the need for enhanced security measures within the HSMS protocol but also the importance of vigilant monitoring to detect and mitigate MITM attacks swiftly.

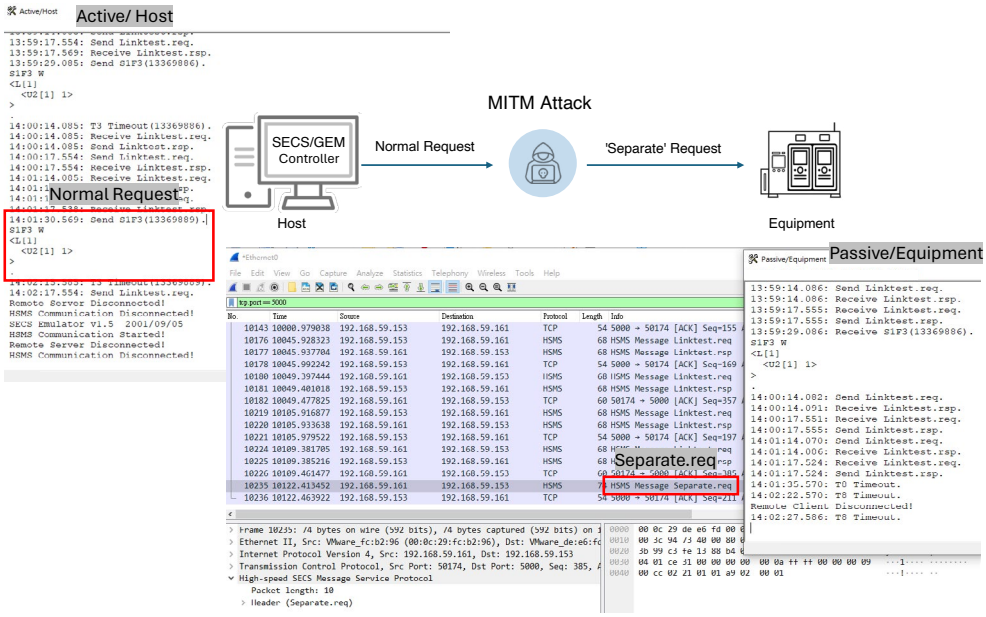

Figure 12. Man-in-the-Middle Attack Involving 'Separate' Requests

Scenario 3: Exploiting SECS Message Language File Vulnerabilities

On a different note, the SECS Message Language (SML) was introduced by GW Associates to simplify the interpretation of SECS messages for operators by converting the content into readable text. This language, integral to the semiconductor industry, is supported by many SECS/GEM controller interfaces and is recognized in various SEMI standards. Operators can easily replicate specific Streams and Functions by importing an SML file.

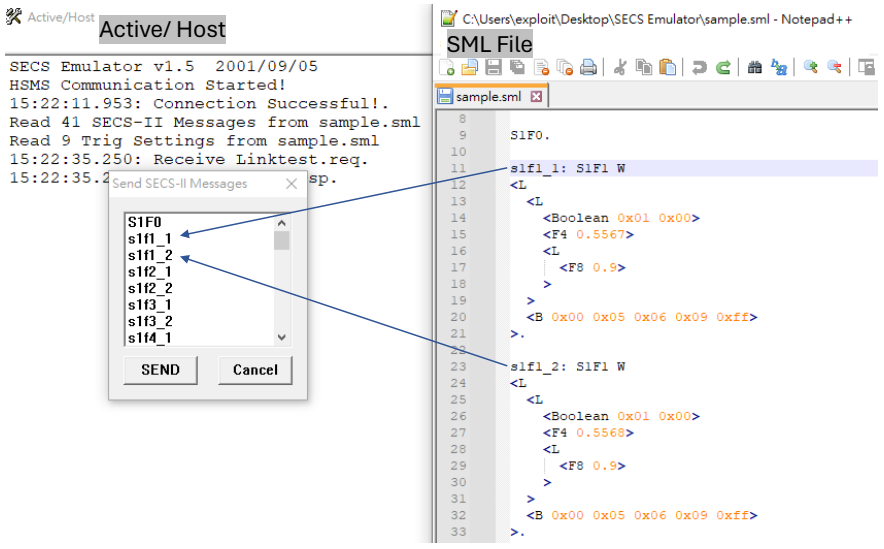

Figure 13. An Example of SML (SECS Message Language)

03

However, the capabilities of handling SML files varies across different interface designs. Some interfaces may fail to properly process the content of SML files, leading to system crashes. For instance, a string represented as <A[7] “example”> signifies a 7-byte ASCII string. Altering this to <A[-1] “example”> disrupts the interface, causing it to disconnect from the equipment. This vulnerability provides an avenue for attackers to masquerade as integrators or suppliers, sending malformed SML files to operators. Should an operator import such a file, the host's connection to the equipment would be severed, allowing the attacker to establish and maintain control over the equipment. This potential attack vector is illustrated in Figure 14, demonstrating the importance of validating SML files before importation so as to safeguard against unauthorized access and control.

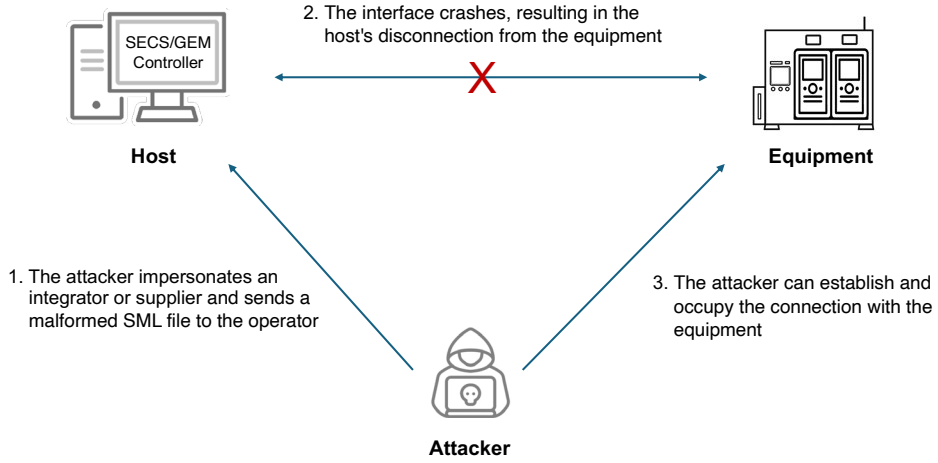

Figure 14. Crashing a Controller with a Malformed SML File

Scenario 4: Vulnerabilities in Non-Strict HSMS Protocol Implementations

The HSMS protocol is a critical component in semiconductor manufacturing, defining four principal states to manage communications between devices over TCP/IP. These states include:

1. TCP/IP Not Connected: The entity is poised to initiate or receive TCP/IP connections but has not established any yet, either because none have been initiated or because all previous connections have been terminated.
2. TCP/IP Connected: This state indicates an established TCP/IP connection, further subdivided into NOT SELECTED and SELECTED substates.
3. HSMS Not Selected: A substate where no HSMS session is currently active.
4. HSMS Selected: This substate indicates an active HSMS session, allowing for the normal exchange of data messages.

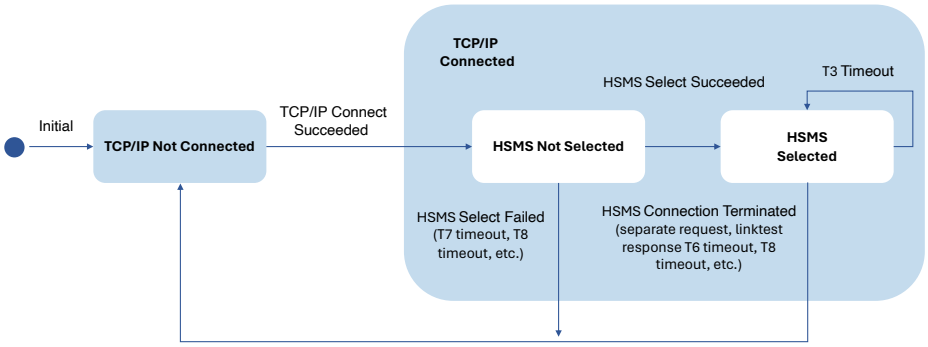

Figure 15. HSMS-SS State Machine Diagram

Utilizing HSMS-SS (a specific mode of HSMS) as an example, transitions between these states are guided by multiple timeouts and procedures, including T3 (reply timeout), T6 (control timeout), T7 (connection idle timeout), and T8 (network intercharacter timeout). Notably:

- The Select Procedure is permissible solely in the Not Selected state, requiring a specific SessionID and indicating readiness for communication. A failure in this procedure necessitates reverting to the Not Connected state by closing the TCP/IP connection.
- The Reject Procedure serves as an optional measure for handling unauthorized data messages, leading to an immediate termination of the TCP/IP connection in cases of non-compliance.
- The Separate Procedure applies exclusively within the Connected state and its substates, mandating the closure of the TCP/IP connection upon initiation or receipt of a 'Separate' request, transitioning back to the Not Connected state.

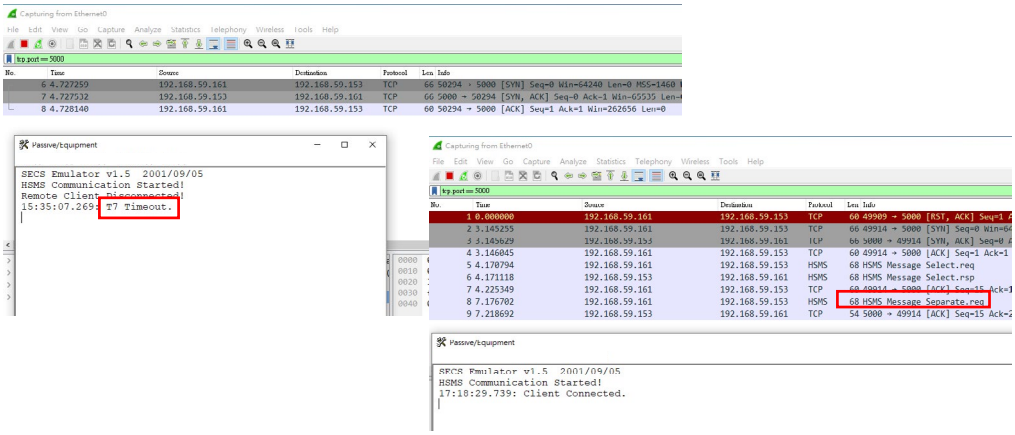

Figure 16. Non-Standard Compliance of Some Equipment

03

Despite these well-defined protocols, discrepancies arise in adherence across different equipment. For instance, as illustrated in Figure 16, certain devices fail to terminate the TCP/IP connection as mandated by the Select or Separate procedures (e.g., due to a T7 timeout or receipt of a 'Separate' request). This lapse in protocol adherence provides attackers with a loophole to maintain connections without needing a comprehensive understanding of the HSMS protocol. Consequently, legitimate operators may find themselves unable to re-establish connections with equipment after a disconnection, underscoring the necessity for stringent compliance with HSMS standards.

Scenario 5: Potential Threats in Specialized Semiconductor Applications

The HSMS protocol traditionally limits equipment to supporting only a single host connection at any given time. However, for specialized applications within the semiconductor industry, there are instances where equipment may need to accommodate multiple concurrent host connections. Such configurations are defined by the equipment itself, which manages the coordination of activities among various hosts. This flexibility, while advantageous, introduces potential security vulnerabilities. For example, certain open-source HSMS servers may not effectively handle unexpected requests by default. An attacker could exploit this by sending incorrect procedures or unexpected data, leading to server failures. As depicted in Figure 17, an unanticipated S1F1 request sent to a server—without a preceding 'Select' request—could cause the server to crash due to an unhandled error event, thereby preventing legitimate hosts from interacting with the server.

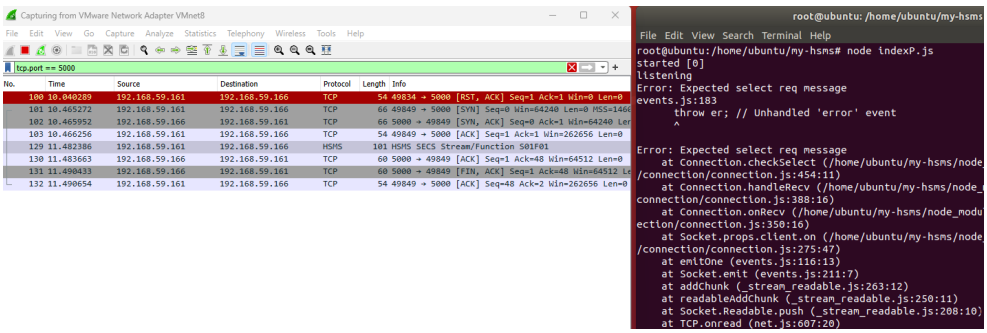

Figure 17. Attacker Sending an Incorrect Procedure

Though such an attack may initially appear to be a simple Denial of Service (DoS) tactic, its implications for the semiconductor industry are far-reaching. Given the sequential interconnection of machinery along the production line, an attacker could strategically disrupt communications across various equipment to elude detection. In scenarios where visibility within the FAB is limited, the repercussions of such targeted attacks could be profound, halting production and obscuring the failure's root cause. This scenario mirrors the significant operational and financial impacts observed when computer viruses infect semiconductor production lines, necessitating dayslong shutdowns and resulting in considerable revenue losses.

In response to these challenges and in an effort to enhance equipment efficiency and minimize the cost of equipment failures, the international SEMATECH manufacturing initiative proposed the concept of the Equipment Engineering System (EES). The EES aims to collect and analyze equipment parameters to predict the health status of the equipment and the quality of manufacturing outputs. Illustrated in Figure 18, the EES framework consists of three interfaces: A,

B, and C. Interface A defines the communication protocols and specifications for engineering data collection, Interface B facilitates communication between the Manufacturing Execution System (MES) and the EES, and Interface C outlines specifications for secure remote system access outside the factory network. Despite minimal development activities for Interface C,[7] the increasing convergence of IT and OT necessitates heightened vigilance against potential cyber threats, especially in FAB factories connected to the public internet. The lack of built-in security features in protocols like E37 (HSMS) exposes the industry to various attack vectors, including those that could impair process control or inhibit response functions.

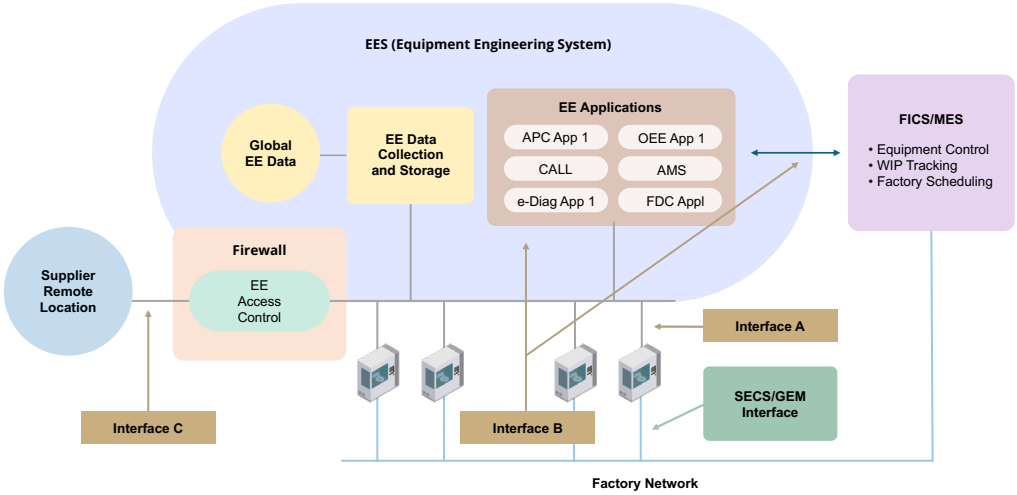

Figure 18. The EES Architecture[8]

Conclusion: Potential Attack Surface and Mitigation Strategies

Given the inherent cybersecurity limitations within the HSMS protocol, the semiconductor industry faces significant vulnerabilities. Attackers can exploit these gaps to impair process control and inhibit response functions, particularly through MITM attacks. The protocol's design, including both its procedural mechanics and the security posture of related software, makes it susceptible to Denial of Service (DoS) attacks. This vulnerability is compounded by HSMS's inability to accommodate multiple concurrent host connections, allowing attackers who establish a prior connection with equipment to block legitimate hosts from reconnecting.

Specialized applications where equipment may accept more than one host connection are at increased risk of direct compromise by attackers. This risk underscores the urgent need for robust cybersecurity measures that can safeguard against unauthorized access and control.

In the contemporary semiconductor manufacturing landscape, the implementation of Interface A, B, and C protocols—particularly for the collection and analysis of equipment engineering data—introduces new cybersecurity considerations. Specifically, Interface C, which facilitates remote system access outside of the factory, presents additional opportunities for attackers to breach regulated OT environments.

The flexibility of E5 (SECS-II) streams, which allows for the collection of equipment data and control over equipment actions, further complicates the security landscape. The ability of vendors to define streams means attackers could potentially inject malicious code into equipment, especially in scenarios where equipment integrates with other protocols.

In response to these challenges, TXOne Networks offers a comprehensive solution with its sustainable asset lifecycle defense framework tailored for the semiconductor industry. By conducting thorough analyses of SECS/GEM protocols and the behavioral patterns of equipment and cyber-physical systems within factory settings, TXOne enables the establishment and continuous monitoring of secure operational baselines. Adopting a "never trust, always verify" philosophy ensures that any deviation from normal interactions is subject to immediate investigation, enabling the prevention of unauthorized modifications, the detection of threats, and a swift response to any detected anomalies.

Reference

[1] TXOne Networks, "Potential Threats to Semiconductor Processes", TXOne Networks Blog, July 08, 2022.
[2] TXOne Networks, "OPC UA Protocol Cyber Threats," TXOne Networks Blog, March 29, 2023.
[3] TXOne Networks, "New Report Analyzing the Data Distribution Service Protocols", TXOne Networks Blog, April 23, 2022.
[4] TXOne Networks, "MQTT Series 1: Usage of MQTT in our IoT/IIoT World", TXOne Networks Blog, November 25, 2019.
[5] TXOne Networks, "Protecting Utilities: Strategies for Mitigating Common Attack Vectors", TXOne Networks Blog, May 31, 2023.
[6] TXOne Networks, "Unveiling SEMI Innovative Cybersecurity Architecture", TXOne Networks Blog, December 01, 2023.
[7] PEER Group, "Interfaces A, B, and C", PEER Group, Accessed April 16, 2024.
[8] Harvey Wohlwend, “e-Diagnostics and EEC Guidance", SEMATECH, March 18, 2002.