# The Industrial Internet of Things Trustworthiness Framework Foundations

An Industrial Internet Consortium Foundational Document

Version V1.00 – 2021-07-15

Marcellus Buchheit (WIBU-Systems), Frederick Hirsch (Upham Security), Robert A. Martin (MITRE), Dr. Vincent Bemmel (Corlina), Antonio J Espinosa (Corlina), Bassam Zarkout (IGnPower), Charles F. Hart (Hitachi), Mitch Tseng (Tseng InfoServ).

## CONTENTS

## FIGURES

## PRINCIPLES

The Cambridge Dictionary defines *trustworthy*[1] as *deserving of trust, or able to be trusted*. In the context of an industrial system or a component used for an industrial system, *trustworthiness* means that a subject deserves trust or can be trusted. The dictionary says *trust*[2] is *to have confidence in something, or to believe in someone*, but this lacks technical guidelines, so the Industrial Internet Consortium (IIC) has refined the definition to apply the concept of trust and trustworthiness in the context of acquiring, deploying and operating industrial systems.

The Industrial Internet Reference Architecture (IIRA)[3] designated five key system characteristics to support a system's business purpose and to ensure that functions perform adequately without compromise. The five characteristics are: safety, security, reliability, resilience and privacy. A similar list of "dimensions" was created in 1999 by the *Committee on Information Systems Trustworthiness* to describe the trustworthiness of networked information systems.[4] In early 2016, IIC adapted the term to its own list of five characteristics, defining the core of trustworthiness. Contemporaneously, NIST also reintroduced the term based on the *same characteristics*.[5] In addition to the trustworthiness characteristics, IIC also specified four groups of threats that endanger a trustworthy system, which resulted in the following definition:

Definition

**Trustworthiness**
The degree of confidence one has that the system performs as expected. Characteristics include safety, security, privacy, reliability and resilience in the face of environmental disturbances, human errors, system faults and attacks. (IIC Vocabulary[6])

These five characteristics are *trustworthiness characteristics.* There are four classes of trustworthiness threats: environmental disturbances, attacks, human errors and system faults. The trustworthiness characteristics and threats are shown in Figure 0-1. These threats can result in hazards that can lead to loss, as described in section 2.3 related to risk management.

Trustworthiness, and confidence in that trustworthiness, are an essential aspect of industrial systems. Such systems are complex systems of systems that can exhibit emergent properties due to the interconnection and interactions of their subsystems. These subsystems can include information technology (IT) that focuses on data and on operational technologies (OT) that use data, sensors, and actuators to change the physical environment. The consequences of incorrect action can lead to loss of human life, impact on the environment, interruption of critical

[1] see [CD-TW][CD-TW]
[2] see [CD-TR]
[3] see [IIC-IIRA2019]
[4] see [Schneider1999]
[5] see [NIST-Gl-TW]
[6] see [IIC-Voc2020]

infrastructure, or other consequences such as disclosure of sensitive data, destruction of equipment, economic loss or damage to reputation. Other trustworthiness concerns for the business include compliance with regulations, avoiding potential liability and litigation and consideration of the potential benefits from a trusted reputation.

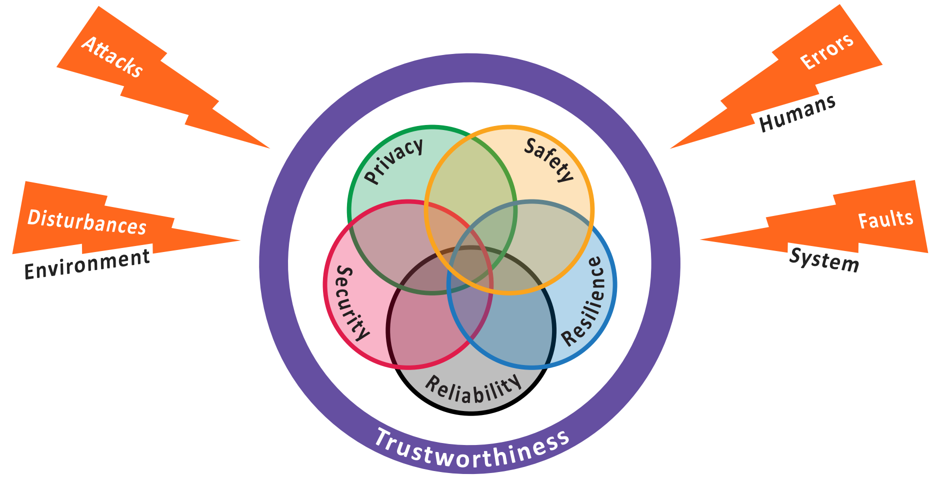


Figure 0-1: Trustworthiness characteristics and threats

This document defines and motivates trustworthiness, highlights the need to consider trustworthiness throughout the system lifecycle and raises awareness of technologies, processes and practices. We also highlight traceability and assurance of trustworthiness based on evidence. Trustworthiness applies to both new and existing systems. Trustworthiness characteristics and their interrelationships are important in all systems, but the relative priorities and the nature of the relationships depend on the context of the industry vertical and system under consideration.

The audience for this document includes owners, operators, system integrators, business decision makers, architects, engineers, buyers, and any stakeholder with interest in the security, safety, reliability, resilience and privacy of cyber-physical systems.

Business decision makers, architects, and engineers can use this document to guide the development of interoperable technologies and solutions while considering trustworthiness, balancing it with other stakeholder and operational requirements. Owner, operators, system integrators, and buyers can use it as a common starting point of system conception and design related to trustworthiness for their specific systems, making this applicable to a variety of verticals.

This work enhances the discussion of trustworthiness in early sections of the Industrial Internet of Things, Security Framework (IISF)[1], but it does not address the security, functional and implementation details in the later portions of the IISF. We include all of the trustworthiness characteristics defined by the IISF; what they are, why they are important, how they relate to

[1] see [IIC-IISF2016]

business decisions and how they interrelate. We focus on cross-cutting concepts and the implementation, operational and support considerations surrounding trustworthiness.

The IIRA[1] framed the architectural issues. This '*Industrial Internet of Things, Trustworthiness Framework Foundations'* provides a foundational view of trustworthiness.

We anticipate updates to the IIRA to include discussion of trustworthiness in the reference architecture, and the IISF to make these three publications consistent.

Trustworthiness is intimately related to the maturity of an organization, meaning its understanding of the need and appropriateness for controls shaping the behavior of their systems and capabilities, including the acquisition or creation of the systems and the maintenance across their useful life. The IIC IoT Security Maturity Model (SMM) Practitioner's Guide[2] is directly relevant and may also be extended to be more specific to trustworthiness. Some preliminary explorations on this topic have been published in the IIC Journal of Innovation, September 2018[3].

[1] see [IIC-IIRA2019]

[2] see [IIC-SMPG2020]

[3] see [IIC-JOI2018]

# 1 TRUST AND TRUSTWORTHINESS

## 1.1 WHAT IS TRUST AND TRUSTWORTHINESS: TRUSTWORTHINESS CHARACTERISTICS

Trust of an individual or organization within a specific context means that one has the expectation that they will act in a trustworthy manner.[1] Trust and trustworthiness both refer to achieving confidence in another party, within a context. Trust can be achieved in a long-term relationship through repeated confirmation of behavior. In the digital world, trust and trustworthiness are achieved by the trustworthiness characteristics reaching appropriate levels for the context and having evidence to support that level is actually being reached. These characteristics are defined in the IIC Vocabulary[2].

*Safety* ensures that a system operates without unacceptable risk of physical injury or damage to the health of people and indirectly on damage to property or the environment. Nearly any damaging environmental event (e.g., pollution of soil, air or water) presents a risk to human health. Safety implementations should reduce those risks. This requires an analysis of risks, determining ranges of safe operation and designing the system to operate within constraints defined for safe operation. When safety accounts for security risks, additional measures may be needed, such as ensuring a supervisory system is separated from the system being controlled.

Definition

> **Safety**
> The condition of the system operating without causing unacceptable risk of physical injury or damage to the health of people, either directly, or indirectly as a result of damage to property or to the environment.
> (ISO/IEC Guide 51:2014[3] as referenced in the IIC Vocabulary).

*Security* protects a system from unintended or unauthorized access, change or destruction. It ensures confidentiality, integrity, and availability of data at rest, in motion or in use and protects control systems, applications, and services from inappropriate use or behavior. The priority of concerns depends on the context. In operational technology security, availability and then integrity may take precedence over data confidentiality, which is often primary in information technology security, for example.

Security is achieved by protecting endpoint devices, networking equipment, machine identities, virtual machines, containers, and applications and data. In industrial systems, the control data used to execute physical operations has a potential of physical damage and requires advanced protection. Security is defined by the IIC as below, and aligns closely with IT security principles:

[1] see [Hardin2002]
[2] see [IIC-Voc2020]
[3] see [ISO-Guide-51]

Definition

**Security**
The property of being protected from unintended or unauthorized access, change or destruction ensuring availability, integrity and confidentiality.
(IIC Vocabulary)

*Reliability* describes the ability of a system or component to perform its required functions under stated conditions for a specified period of time. This includes any considerations for physical degradation, expired software versions, and well-known potential malfunctions that result in frequent maintenance, replacement of end-of-life components or software updates. Reliability enables uninterrupted operation of the system: an essential element in assuring that the system will meet expectations over time.

Definition

**Reliability**
Ability of a system or component to perform its required functions under stated conditions for a specified period of time.
(ISO/IEC 27040:2015[1] referenced in the IIC Vocabulary)

*Resilience* describes the ability of a system or component to maintain an acceptable level of service in the face of disruption. In contrast to reliability, resilience addresses unexpected and unplanned system statuses that can result, for example, from human errors in operation or an environmental event (loss of power, earthquake, etc.). The main purpose of resilience is to prevent or reduce serious impact of a disruption to the system by damage or loss of operation.

Definition

**Resilience**
Ability of a system or component to maintain an acceptable level of service in the face of disruption.
(IIC Vocabulary)

*Privacy* protects the right of individuals to control what information related to them may be collected and stored by whom and to whom that information may be disclosed, and for how long.

Definition

**Privacy**
The right of individuals to control or influence what information related to them may be collected and stored and by whom and to whom that information may be disclosed.
(ISO/TS 17574:2009[2] as referenced in the IIC Vocabulary).

These characteristics interact together in a variety of ways. Some examples:

[1] see [ISO-27040]
[2] see [ISO-17574]

- Reliability addresses the correct functionality of the system under specified conditions, while resilience addresses the functionality of the system under non-planned conditions.
- Privacy protects only human-related data but does not address business or operational data, which is instead covered by security.

Safety is concerned with protecting people, while security and reliability are responsible for the protection of the system and its outputs. Resilience describes the ability of the system to function when the normal, reliability-controlled condition is lost.

Trustworthiness characteristics interact dynamically and can enhance or inhibit each other. System design needs to consider the interactions of the characteristics and their tradeoffs. Evidence about the tradeoffs and the rationale for the choices can provide assurance that the system is trustworthy for a specific industrial system. Concerns in a factory are different from a hospital operating room. For each system, the designer must understand the trustworthiness considerations involved in the implementation, assembly, operation and maintenance of the system and maintain that composition to continue the system's trustworthiness.

Trustworthiness relates to a system. This means that the system of interest must be well defined, since the definition of the system will determine the relevant requirements. For example, defining the trustworthy system of interest as a set of cameras in a chemical factory is different from considering trustworthiness of the entire factory. Risks related to the cameras may include privacy or trade secret loss, prevention of safety or security monitoring, or possible security attack pivot points. When the full factory is the system of interest, it includes all the issues for the cameras and many more, such as a potential chemical explosion that could cause loss of life. The system of interest must be defined so appropriate concerns and tradeoffs are considered.

Figure 1-1 shows how the trustworthiness characteristics we have defined relate to other important attributes that are not part of the definition such as accountability, quality, sustainability and suitability. (Note: Each system has its own attributes depending upon its design and purpose. Those listed are examples only and are not intended to be definitive.)

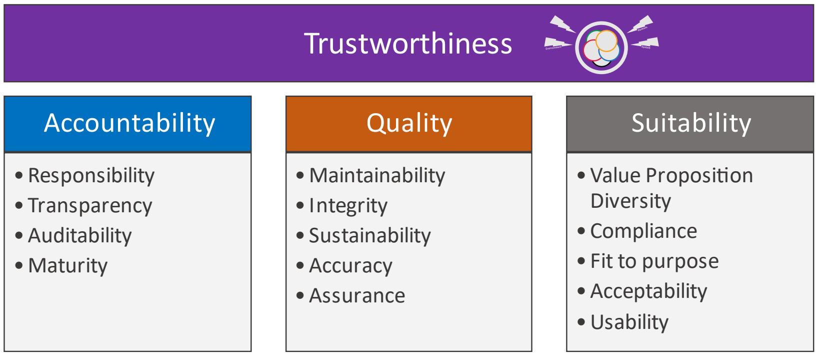


Figure 1-1: Trustworthiness in context

As an example of how these various attributes relate to trustworthiness, addressing safety concerns may mean putting design constraints on a system design and policies on operations.

People need to be responsible and should be held accountable for following the design and operational activities they can control. This is reflected in the following definition of accountability, originally noted in the context of health care but more widely applicable:

Definition

**Accountability**
Obligation of an individual or organization to account for its activities, for completion of a deliverable or task, accept responsibility for those activities, deliverables or tasks, and to disclose the results in a transparent manner. (ISO/TS 21089:2018 [1])

Assurance needs to be provided to stakeholders, in the form of evidence that a system can be trusted. We define assurance as follows:

Definition

**Assurance**
Grounds for justified confidence that a claim has been or will be achieved. (*ISO/IEC 15026-1:2013* [2] as referenced in IIC Vocabulary)

Understanding context is essential to support trustworthiness and make the needed tradeoffs.

## 1.2 Context in Trustworthiness Design

Understanding the context is essential to making decisions and tradeoffs since everything (measurements taken, tools used, influences, threats, hazards, errors, disturbances, and faults and consequences of trustworthiness failures) may differ depending on context, see Figure 1-2.

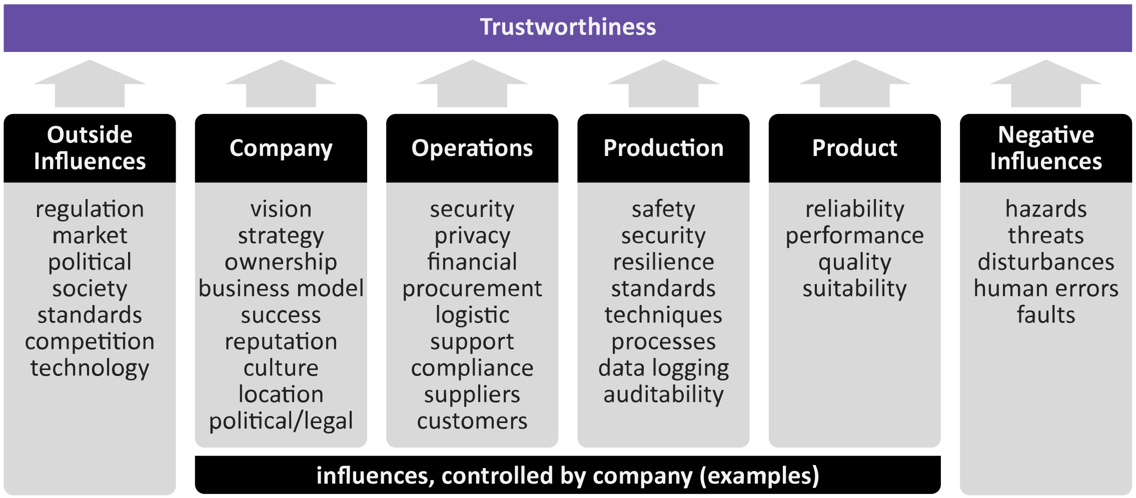


Figure 1-2: Trustworthiness context

[1] see [ISO-21089]
[2] see [ISO-15026-1]

The context of trustworthiness is important because monitoring of any adverse event and the understanding of its consequences are integral to designing and operating the system to achieve trustworthiness. Indirect influences may be harder to detect and control, for example, if they are in the supply chain for the components of the system. Trustworthiness characteristics must be considered together, in the context of the system, stakeholder concerns and potential consequences. A system must be considered as a whole, and this includes considering indirect effects of design elements or changes and how these indirect effects can affect trustworthiness.

Critical considerations include:

*Consequences:* Negative influences can affect parties directly, indirectly, or in combination. For example, strategy or operations changes may affect employees directly. Similarly, changes to products or support affect the customer directly. They may also affect customers indirectly through factors such as declining product quality or increasing prices. A supplier may be indirectly affected through customer feedback, changes in sales or profits, etc.

*Immediacy:* Effects can be obvious to all parties instantly, for example severe weather. Or they may be more difficult to recognize. Attacks, such as theft of intellectual property, may not be noticed—by the victim or its customers—until well after the event. A delay in responding to an event can cause harm until noticed, so delays can have negative consequences.

*Severity:* Trustworthiness failures can have consequences ranging from trivial to crippling and can affect parties and processes with differing severity depending upon context. Trustworthiness failures can lead to many unrelated effects on a wide range of participants. An undetected safety lapse, for example, can result in injured workers, product defects, production delays, declining financial performance and legal consequences. A less serious safety lapse can actually improve trustworthiness if caught and fixed before any serious accident or failure occurs.

These contextual factors can result in improvement or decline in overall trustworthiness.

Some additional points of trustworthiness in context are worth noting:

*Company differences:* Companies are different, and they use their own selection of standards and practices depending upon the goals of their business. Companies in the same industry may make different decisions about how to organize and conduct operations. Even when similar processes are used, the parameters of the process can differ. Suppliers that sell wood for fuel have different tree-selection standards from those who sell lumber, and different requirements for cutting and processing the wood, so risks and concerns differ as well.

*Local environment:* Local custom, culture, and law inevitably influence trustworthiness for suppliers and consumers by imposing different minimal requirements. One example is privacy requirements, where data usage in one country may be acceptable while illegal in another.

*Balancing aspects of trustworthiness:* Varying weights can be given to the five components of trustworthiness to achieve a balance appropriate to the business and context. For example, the design of production operations for mining may prioritize safety by including high levels of safety procedures in its process, while security may have lower priority. On the other hand, an IT-services firm may instead prioritize security and confidentiality over safety.

*Affected parties:* Any trustworthiness factor can affect multiple parties. For example, safety can affect production workers, consumers of the supplied product, or the general public.

Principle

**Principle 1: Trustworthiness characteristics must be considered holistically.**

Principle

**Principle 2: Understanding context is necessary for making Trustworthiness tradeoffs.**

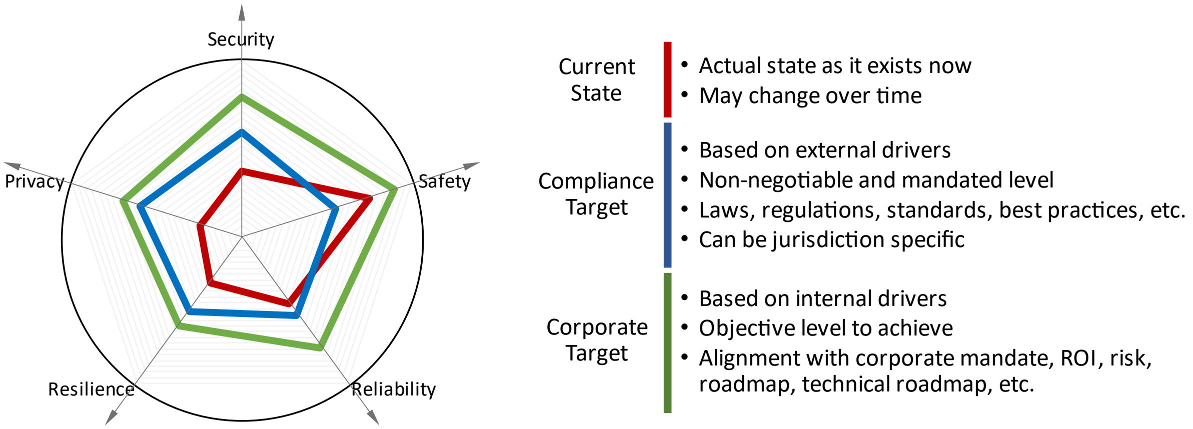
## 1.3 Trustworthiness Requirements

A trustworthy IoT system must meet compliance-mandated requirements for the trustworthiness characteristics throughout the lifecycle of the system. Defined by laws, regulations standards and industry-accepted best-practices that apply within the jurisdiction where the IoT system is deployed. Beyond these, some organizations may set higher trustworthiness targets, based on corporate vision, roadmap, market positioning and other objectives.

Privacy and safety examples of regulatory constraints include EU GDPR privacy law and OSHA workplace safety standards. Reliability and resilience are often driven less by laws than by competitive forces, though aerospace or healthcare, for example, are heavily regulated.

Organizations must assess and define mandatory requirements and subsequently implement systems and methods to track and measure the compliance with these requirements. These efforts must be also sustained throughout the lifecycle of the IoT systems.

The current state, compliance target and corporate target can be visualized as in Figure 1-3.



Current State
- Actual state as it exists now
- May change over time

Compliance Target
- Based on external drivers
- Non-negotiable and mandated level
- Laws, regulations, standards, best practices, etc.
- Can be jurisdiction specific

Corporate Target
- Based on internal drivers
- Objective level to achieve
- Alignment with corporate mandate, ROI, risk, roadmap, technical roadmap, etc.

Figure 1-3: Trustworthiness states (source: IIC)

### 1.3.1 Business Benefits of Trustworthiness

There are several major drivers for trustworthiness in an industrial context:[1]

*Create safe and secure systems:* Create systems that do not injure or cause deaths to people or harm the environment.

*Meet compliance* requirements of laws and regulations (industry-specific, governmental or regional). Compliance is necessary for product approval and has marketing value.

*Risk avoidance* and *mitigation*: Evaluating and addressing risk are helped by frameworks and guidelines, such as the IIC IoT Security Maturity Model,[2] which considers business, technology and operational concerns to achieve a good fit between business needs and security investment.

*Performance predictability* and *quality*: Operational efficiency and system-specific objectives drive this aspect. When collaborating with partners, this is subject to an agreement, such as data center service availability as contracted for in service-level agreements (SLAs).

Business drivers are related to trustworthiness characteristics. For example, regulations are direct drivers for safety, security and privacy. Risk analysis is also used for safety, security and privacy and to a lesser extent for reliability and resilience. Performance predictability and quality management are typically applied to reliability and resilience is considered as part of business continuity.

Sometimes the relationships can be more complex. Privacy regulations, such as GDPR, initially present compliance concerns but can also affect business-risk management because of the significant financial penalties and reputational damage for violation of regulations.

[1] see [IIC-MAT2019]
[2] see [IIC-SMPG2020]

Many trustworthiness requirements are initially driven by some concern within one category, yet indirectly motivate others. For example, if the reliability of a service is necessary for performance predictability, one way to support it is to enhance the security of the service to avoid delays or shutdowns caused by a denial of service (DOS) attack. Business performance and product quality directly drive reliability requirements and are affected by other characteristics such as security.

Failure to meet the compliance and trustworthiness requirements can lead to accidents, equipment damage, personal injury, data breaches, delays and operational interruptions and lack of compliance. These direct consequences may incur additional indirect costs such as revenue loss, fines, litigation costs, higher insurance costs, reputation damage and opportunity loss.

Greater levels of trustworthiness can generate business benefits in many ways, including reductions of costs related to failure, as well as other benefits to the business, such as:

- reduced rework, waste, and delays in supply chains,
- reduced levels of compensation payments to affected parties,
- reduction of fines to regulatory bodies for non-compliance with regulations,
- improved business performance due to stronger brand image, including increased sales, shareholder value and other key metrics,
- reduced insurance cost,
- lower legal costs and reserves,
- lower warranty cost and more favorable warranty terms and
- improved predictability of operational performance.

## 1.4 Consequences Of Inattention To Trustworthiness: An Automotive Example

Connected systems introduce risks to each other, so a single system cannot be effectively trusted until the other systems to which it is connected are also trusted. “Jeep Hack” provides a vivid example from the automotive industry of failures of trust in individual and aggregated systems.

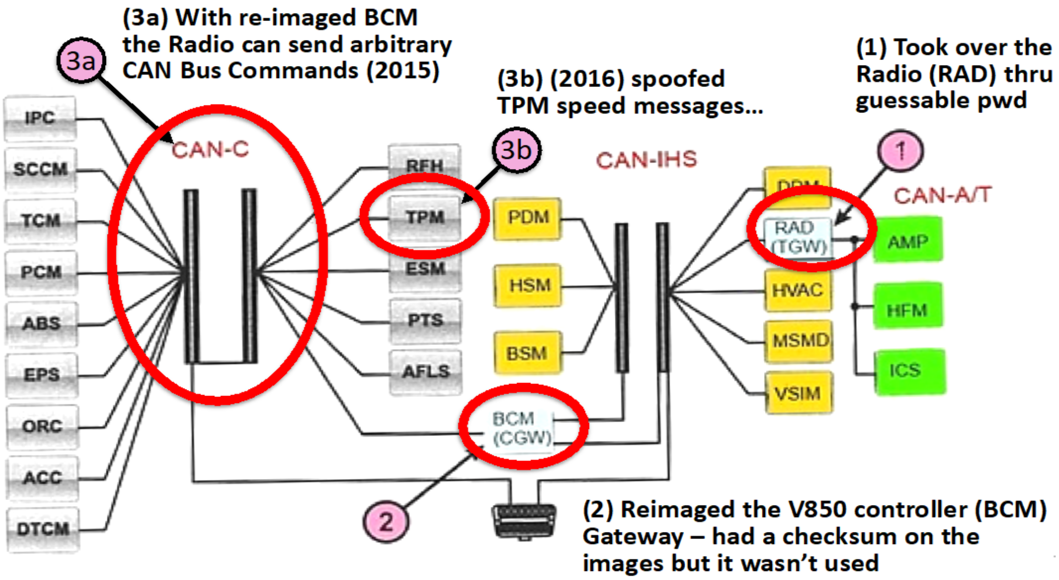

(3a) With re-imaged BCM the Radio can send arbitrary CAN Bus Commands (2015)
3a
(3b) (2016) spoofed TPM speed messages...
3b
(1) Took over the Radio (RAD) thru guessable pwd
1
IPC
SCCM
TCM
PCM
ABS
EPS
ORC
ACC
DTCM
CAN-C
RFH
TPM
ESM
PTS
AFLS
PDM
HSM
BSM
CAN-IHS
DDM
RAD (TGW)
HVAC
MSMD
VSIM
CAN-A/T
AMP
HFM
ICS
BCM (CGW)
2
(2) Reimaged the V850 controller (BCM) Gateway – had a checksum on the images but it wasn't used

Figure 1-4: Hacking a vehicle - Jeep CAN bus, annotated to show exploits

Cybersecurity researchers Charlie Miller and Chris Valasek in 2015[1] and 2016[2] created a series of technical security exploits that allowed near-total remote control of a consumer vehicle, a 2015 Jeep Cherokee. Their work demonstrates how alignment of assumptions about the operational context is necessary across the different components from the supply chain. The designers of the entertainment unit used different contexts from those engineering the CAN bus.

In cars like the Jeep, the radio/entertainment system head unit (labeled "RAD" in Figure 1-4) is an externally facing device that can receive commands through external interfaces such as USB and Bluetooth. Properly secured designs enforce strict separation between head-unit communications and systems related to life safety on the car. However, in 2015 these design requirements were not well understood. Miller and Valasek recognized this flaw.

They also found that the head-unit used an easily guessed password (Figure 1-4 ①). This is convenient for the dealer, service people or manufacturer, who might need those passwords to service the car. Miller and Valesek figured out how to apply a software update to the bus gateway through the head-unit. This bus gateway was supposed to arbitrate the connection between the CAN bus and the bus with the head-unit, but since the password was guessed this control was not effective. After the update, Miller and Valesek had access to all of the devices on the internal CAN bus including those that control the car.

Further, firmware updates to the gateway were applied through the head-unit. They required a signed checksum, but the feature was poorly implemented, and an illegitimate update was not stopped. The hackers' update was accepted by the gateway as a legitimate update without any

[1] see [Miller2015]

[2] see [Miller2016]

authentication (Figure 1-4). Once they had access to the CAN bus from the head-unit they could issue commands to others on the bus, including opening and closing windows, turning on blinkers and windshield wipers, changing speeds and turning the wheels.

The attacks in 2015 did not work at highway speed because they were based on the diagnostic system which did not allow changes to be made above 5 mph. The attackers learned that the tire pressure monitoring system was the source of the information about the speed of the vehicle so in 2016 the Jeep Hack evolved to spoof the tire-pressure monitor messages to tell the car that it was going slowly. This was possible because the protocol for the bus discarded duplicate messages. Once they knew how to get illegitimate message numbers onto the bus before the actual tire pressure monitoring systems messages through a spoof attack, they could go at highway speeds. The tire pressure management message was discarded as duplicative, so the car concluded it was going slowly when in fact it was not.

This example shows clearly how one untrustworthy system (the radio/entertainment head unit) in an otherwise trustworthy solution can render all connected systems untrustworthy. It is essential that all systems connected to other systems be trustworthy. Otherwise, a failure of trust in one can bring down the others and result in a failure across all the connected systems.

# 2 Business Foundations for Trustworthiness

## 2.1 Organizational Culture and Support for Trustworthiness

Digital trust is an extension of person-to-person trust. Trust is maintained over time, and there must be enough visibility and transparency to enable judgements that the system performs as designed throughout the lifecycle of the system.

A good reputation enhances trust in a system. The fact that system has historically performed as expected, and disruptions were minimal lends credence to trust in the system.

Principle

**Principle 3: Organizational consistency over time enables reputation and trust.**

Principle

**Principle 4: Accountability is an essential underlying foundation of trustworthiness.**

Achieving trustworthiness may require that those ensuring trustworthiness must be independent from those they monitor, such as a development team, and free from undue influence due to budget or management, with the ability to stop production as appropriate. This is especially relevant to software trustworthiness and is similar to quality initiatives. See Boeing 737MAX MCAS design flaw[1] and other examples of related failures.

Principle

**Principle 5: A culture of trustworthiness is essential to achieving trustworthiness.**

## 2.2 Managing Trustworthiness as an Iterative Process

Concerns about establishing confidence that an industrial system meets the trustworthiness requirements must be addressed throughout the lifecycle of the system, an effort that must be powered by an established program for an extended period.

Figure 2-1 shows an example of such a trustworthiness lifecycle over time:

[1] see [FAA-737MAX]

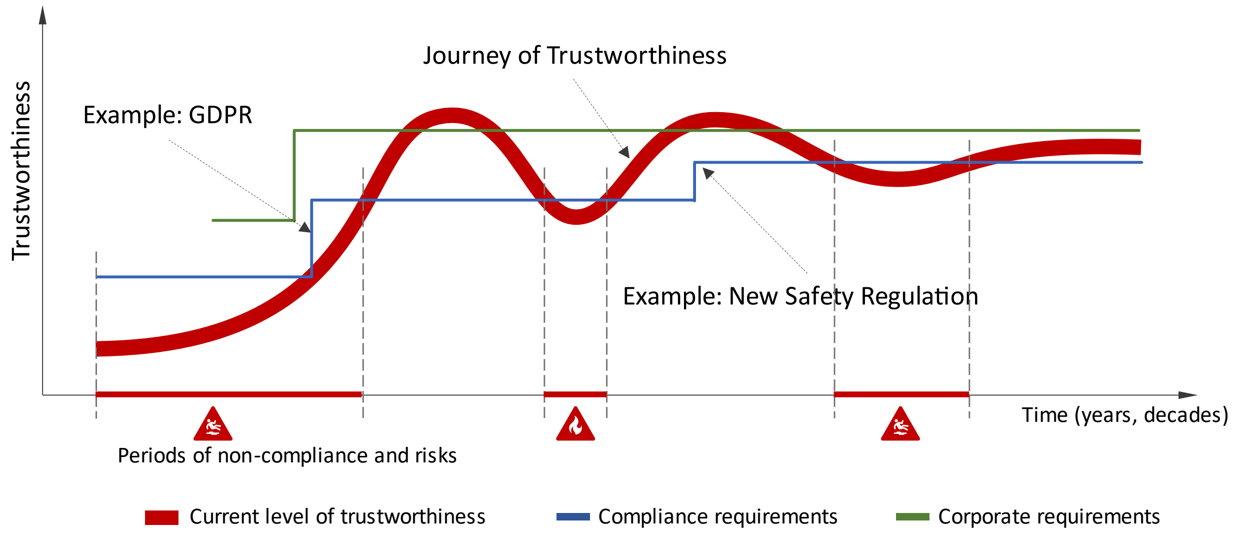


Figure 2-1: Trustworthiness journey over time

The blue line in this diagram represents legal and regulatory compliance requirements. These requirements must be transformed into technical and functional requirements that are weaved into the system design. The upward jumps in that blue line represent increases in requirements during the lifecycle that may result from new laws and regulations going into effect.

The green line represents the corporate requirements, based on internally defined, self-imposed drivers and objectives (business and technical). The upward jumps in the green line represent increases in corporate requirements for trustworthiness that match the upward jumps in the compliance-mandated requirements.

The red line shows the current state of trustworthiness of the system, as currently implemented, deployed and operated. A system is trustworthy if the red line is above the blue line.

### 2.2.1 Top-down versus Bottom-up versus Middle-out Approaches

Managing and controlling a trustworthy system may require a combination of top-down elements (as described above) and bottom-up elements, as Figure 2-2 shows:

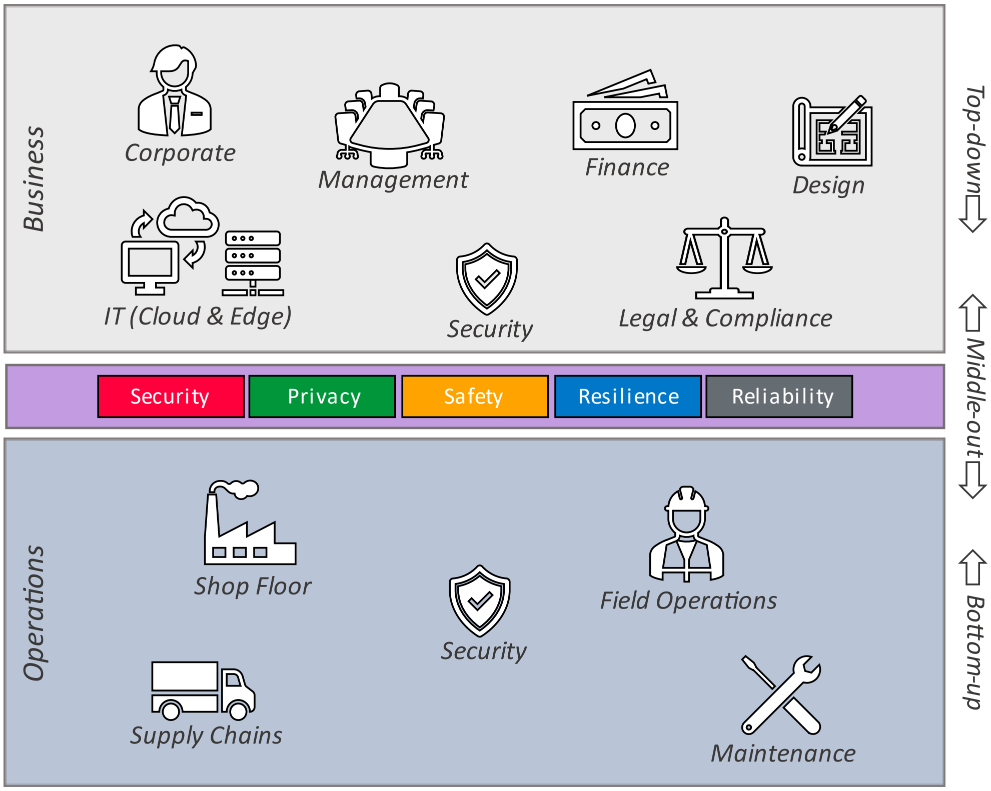


Figure 2-2: Middle-out management approach for trustworthiness

A cross-functional team should manage trustworthiness activities in the organization, due to the work's cross-functional nature and the need to coordinate top-down and bottom-up activities.

| Description | Bottom up | Top Down |
|---|---|---|
| Stakeholders | Operational, production and regional managers, with their partners and customers. | Corporate executives, business managers.<br>Trustworthiness should be assigned a corporate sponsor to mandate and track its realization. |
| Drivers | Safety and continuity of operations, risk mitigation with respect to production and operational objectives, reliability of equipment and services involved in operations, resilience of operational systems with respect to known risks. | Regulatory compliance<br>Global market requirements<br>Corporate-wide policies<br>Industry standards and practices. |
| Challenges | Identify objectives and governance across departments, business units.<br>Harmonize and integrate separate trustworthiness objectives and | Trustworthiness crosses group and departmental boundaries. |

| Description | Bottom up | Top Down |
|---|---|---|
| | measures to align them with and contribute to corporate and regulatory drivers.<br>Understand interdependencies of various trustworthiness objectives. | Organization-wide objectives must be translated in a consistent way across departments or units. |
| Implementation | Empower operational personnel and managers to establish trustworthiness objectives and metrics for local operations.<br>Establish a trustworthiness council across business units to address fragmentation and interdependency challenges.<br>Different stakeholders responsible for the various trustworthiness characteristics must define current states and identify compliance states of their respective domains.<br>They must also identify the requirements to move from the current to compliance states, including technical roadmaps, budget requirements and resource requirements. | Assign trustworthiness to a corporate sponsor who can mandate its realization.<br>Sponsor may mandate trustworthiness targets that can exceed compliance levels.<br>Steering committee comprising representatives from different groups and departments oversees the work on trustworthiness |

The system may be a system of systems deployed across organizational boundaries and jurisdictions. Ensuring the top-down-driven trustworthiness objectives are applied throughout the design, implementation, and operation of the system requires a middle-out approach that bridges the bottom-up and top-down approaches. The middle layer starts with the top-down requirements, then reconciles them against the specific complexity and specificity of each organization and jurisdiction, accounting for financial, compliance and operational risks.

## 2.3 MANAGE TRUSTWORTHINESS RISK–REMOVE RISK MANAGEMENT SILOS

The IIC defines[1] *risk* as the effect of uncertainty on objectives, often expressed in terms of a combination of the consequences of an event (including changes in circumstances) and the associated likelihood of occurrence.

[1] see [IIC-Voc2020]

This approach can be problematic when the likelihood of an occurrence is hard to estimate, such as with software, new activities (such as the first flight to the moon or use of a new technology), or for events that occur infrequently making for a lack of historical data.

Another approach is to start with some definitions, taken from the STP-Handbook[1] and Engineering a Safer World: Systems Thinking Applied to Safety[2]*, both* by Nancy G. Leveson et al:

Definition

**System**
A set of components that act together as a whole to achieve some common goal, objective, or end. A system may contain subsystems and may also be part of a larger system.

Definition

**Loss**
The negative consequence of an undesired or unplanned event.

Definition

**Accident**
An undesired or unplanned event that results in a loss, including loss of human life or human injury, property damage, environmental pollution, mission loss, etc.

Definition

**Hazard**
A system state or set of conditions that, together with a particular set of worst-case environmental conditions, will lead to a loss. A potential source of harm.

Note

Some of these terms may be used differently depending upon the discipline, but for this discussion we will use the definitions above as drawn from the field of safety.

A common mistake in defining hazards is to confuse hazards with causes of hazards: For example, “brake failure”, “brake failure not annunciated”, “operator is distracted”, “engine failure” and “hydraulic leak” are not system-level hazards but potential causes of hazards. To avoid this mistake, make sure the identified hazards do not refer to individual components of the system, like brakes, engines, hydraulic lines, etc. Instead, the hazards should refer to the overall system and system states.

To give an example, falling down a staircase can cause a loss, be it death, serious injury, or spilling one’s tea. The hazard is losing balance and falling on the stairs. Potential causes can include a rotted stair tread, slippery stairs from rain or a tea-

[1] see [Leveson2018]
[2] see [Leveson2012]

> spill, faulty shoes, inattention, or a combination of causes. Mitigations can include installing a handrail, inspecting stairs for quality, providing training to avoid distractions and to not carry too much on the stairs, a policy to disallow drinking tea while using stairs, installing a roof for rain, or installing plastic tread enhancements, to give a few examples.

When evaluating risk, the following questions are critical:

- What are the potential losses?
- Which hazards could lead to losses?
- What are the potential causes of those hazards?
- How might the hazards be addressed?
- Which decisions should be made to avoid, mitigate or manage the hazards?

When building a system, from conceptualization through design and implementation to operation and decommissioning, decisions are made regarding requirements and priorities of different choices, based on cost, goals, trustworthiness and others. Traceability is important so decisions are recorded and understood by all parties over time.

# 3 SYSTEM VIEW OF TRUSTWORTHINESS

## 3.1 TRUSTWORTHINESS APPLIED TO SYSTEMS AND SYSTEMS OF SYSTEMS

Stakeholders must define the system under consideration for which they want trustworthiness. Choosing that system relates to the goals and concerns of the stakeholders and the constraints associated with it. This system will interact with other systems, and their consequences should also be understood. Understanding trustworthiness requires a whole-system view.

The complexity of a system and its requirements is partly addressed by considering the concerns of different parties and looking at the system from different viewpoints, as discussed in the IIRA. There, a *system of systems* is defined as a collaborative system coordinating independent systems that, when combined, provide new capabilities and services, but maintain both managerial and operational independence from each other. This relationship may create new concerns about the trustworthiness of the system of systems in addition to the concerns of the independent systems.

Pre-planned system of systems (i.e., one designed from the start to be a system of systems), can consider trustworthiness aspects from inception. Ad hoc system of systems (i.e., systems not originally intended to work together) need to consider new trustworthiness effects that relate to the entire system and should not assume that independent analysis of the constituent systems will adequately address concerns of the system of systems.

Traceability should be treated as a key trustworthiness component in all system of systems architectures in order to be able to validate the understanding of the trustworthiness effects that relate to the entire system.

Every system can affect people, the environment, the business system and business operations as shown in Figure 3-1 with the *trustworthiness target model*.

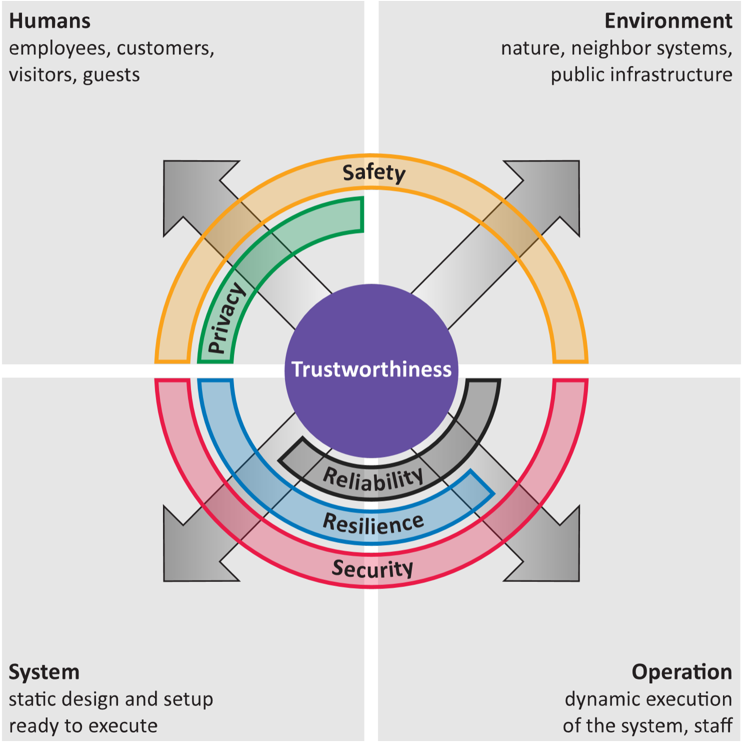

Humans
employees, customers, visitors, guests
Environment
nature, neighbor systems, public infrastructure
Safety
Privacy
Trustworthiness
Reliability
Resilience
Security
System
static design and setup ready to execute
Operation
dynamic execution of the system, staff

Figure 3-1: Trustworthiness target model

Trustworthiness is also affected by the environment or context of regulations, standards, engineering and best practices as shown in Figure 3-2 with *trustworthiness foundation model*. Note that sustainability and supply chain considerations can affect trustworthiness.

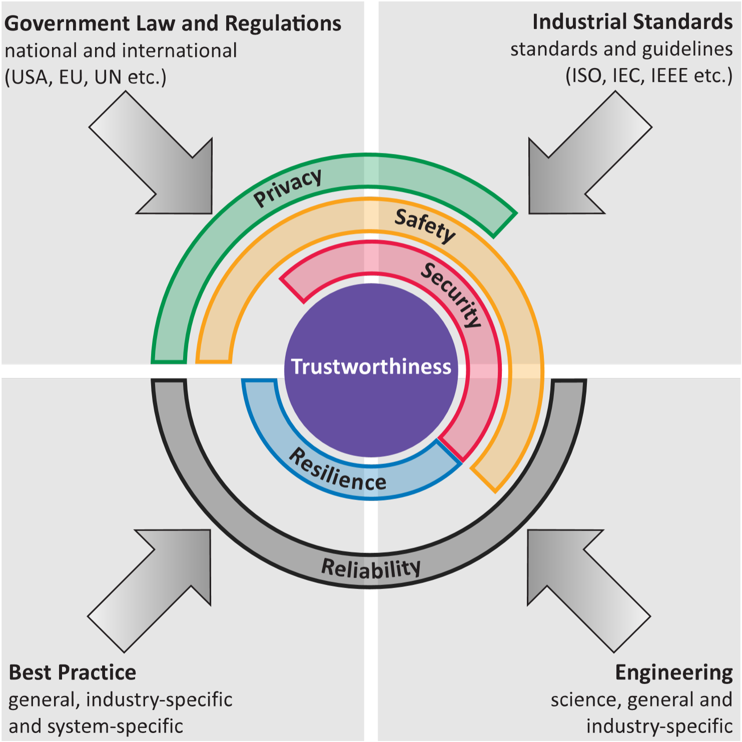


Figure 3-2: Trustworthiness foundation model

Figure 3-3 shows an example of how a change factor can have both positive and negative effects on system characteristics, including trustworthiness characteristics and operational characteristics. The changes and impacts need to correspond to strategic management decisions:

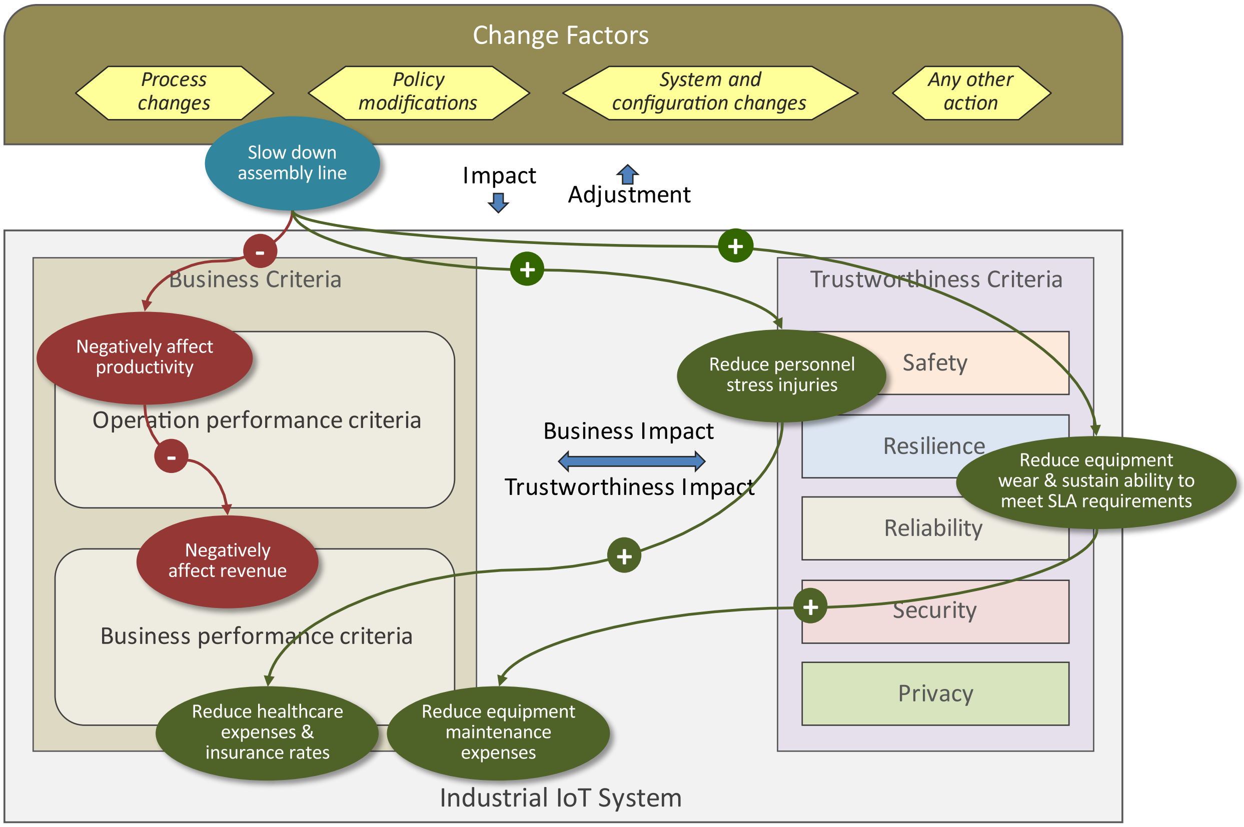

Change Factors
Process changes
Policy modifications
System and configuration changes
Any other action
Slow down assembly line
Impact
Adjustment
-
+
+
Business Criteria
Negatively affect productivity
Operation performance criteria
-
Negatively affect revenue
Business performance criteria
Reduce healthcare expenses & insurance rates
Reduce equipment maintenance expenses
Business Impact
Trustworthiness Impact
+
Trustworthiness Criteria
Reduce personnel stress injuries
Safety
Resilience
Reduce equipment wear & sustain ability to meet SLA requirements
Reliability
+
Security
Privacy
Industrial IoT System

Figure 3-3: Change factors system impact

Slowing down an assembly line in a factory to improve safety by giving workers more time for the tasks around the equipment may reduce the productivity of the line.

Time frames matter in making such tradeoffs. For example, if an accident stops a fast production line the overall productivity of the system could be lower than with a slower, safer line. This may not be apparent if accidents are infrequent, so different time frames must be considered.

As another example, an oil pipeline may detect leaks directly (e.g., human inspection or hydrocarbon sensing) or inferentially (also known as computational pipeline monitoring), using instruments to monitor internal pipeline parameters[1], such as pressure, flow, temperature. Detecting and closing leaks using these techniques increases safety for people and the environment. Resilience can be achieved by combining techniques, such as internal inferential methods and periodic visual inspection. Monitoring can detect a physical attack on the pipeline.

## 3.2 TRUSTWORTHINESS THROUGH THE SYSTEM LIFECYCLE

Security must be considered throughout the lifecycle, including concept, architecture, design and implementation. This is called “security by design”, and the concept extends to trustworthiness. *Trustworthiness by design*, considering the need for trustworthiness at the conceptualization and

[1] see [Argonne2007]

architectural design phase of a system, is a means to achieve assurance that the trustworthiness aspects have been addressed properly for the system. This requires trustworthiness of all aspects of the system and results in a flow of trust from system usage at its highest levels down to its smallest components: a *permeation of trust*—a term adopted from the IISF.[1]

Trustworthiness requires ongoing effort as systems and circumstances change. To maintain trust, assurance is needed that systems and components are trustworthy, that requirements and specifications are correct for the context of the system, and that the system is operating according to specifications even as the system changes.

Principle

**Principle 6: Assurance based on evidence is essential to establish trustworthiness.**

For example, a security attack on a TCP/IP stack implementation in a specific component may provide malicious intruders network access to other, more critical, components within the system. The system operates according to the specification but at same time intruders can collect critical information or prepare an attack to the physical control components.

Trustworthiness must be pervasive in industrial systems, which means there must be both trustworthiness by design and a means to achieve assurance that the trustworthiness aspects have been addressed properly for the system of concern.

### 3.2.1 Systems and Trustworthy Components

A complex industrial system is typically designed and built by people other than the operational user who controls the running system. The system builders select, install and connect components from different component providers, which may be again built out of components from other providers, defining a supply chain across the builders of the installed components. This chain may be complex, especially if such components are installed software modules or data-center-based services, and this chain does not end with the setup of the system: Many components need frequent maintenance or the assurance of quick and compatible replacement in the case of damage or loss. The supply chain of the components is an essential part of trustworthiness of the operated system.

Traditionally, supply chains have an important role in manufacturing processes for delivering parts that will be installed into a larger overall system being manufactured, for example seats for a car. Often the discussion of the trustworthiness of a supply chain for a system focuses on the things that make up the system, but it is also critical that the items used to construct, maintain, test, and run the system are considered to be components of that system and the trustworthiness of these components also be assessed and established. Thus, the components in industrial systems are not just the mechanical, electronic, and software enabled modules like pumps, IoT

[1] see [IIC-IISF2016]

sensors and data-center-services and consumables but also the aspects of these items that need frequent, planned replacements. The trustworthiness of the systems depends on all of them so the items in a system that are used (“consumed”) by the operated system must also determine to be trustworthy. To help address these often-overlooked aspects of making sure the running system is trustworthy we define two different types of components:

- *Installed components* that are rarely replaced except when they are at their defined end of life or have unexpectedly broken beyond repair.
- *Consumable components* that are replaced in a specific defined interval or just used until they are gone or no longer able to perform their intended function within the system.

Example

> Examples for consumable components are non-rechargeable batteries, brushes in electric motors, syringes in a hospital or gaskets between pipes in an oil refinery. An oil refinery needs crude oil—another delivered consumable component.
> Nearly all industrial systems use electric energy as consumable component. And again, trustworthiness characteristics can be easily assigned to this component:
>
> - Reliability: No delivery interruptions, within a specified voltage and frequency range.
> - Resilience: Short overloads should not result in a breakdown.
> - Safety: No high-power spikes that could create electric arcs, threatening people.

### 3.2.2 Roles in Trustworthiness

Figure 3-4 shows a simplified illustration of the roles of operational users, system builders and component builders. The operational user must define the trustworthiness requirements and tradeoffs, and must be able to verify, control and ensure that they are met in all possible states of the operated system. The operational user requires assurance of trustworthiness, demonstrated through *evidence*, so that there is confidence in the system and that it will adhere to the trustworthiness requirements based on the system-specific needs. System builders and component builders need to understand the requirements and provide assurance that they have been met through all stages of the supply chain.

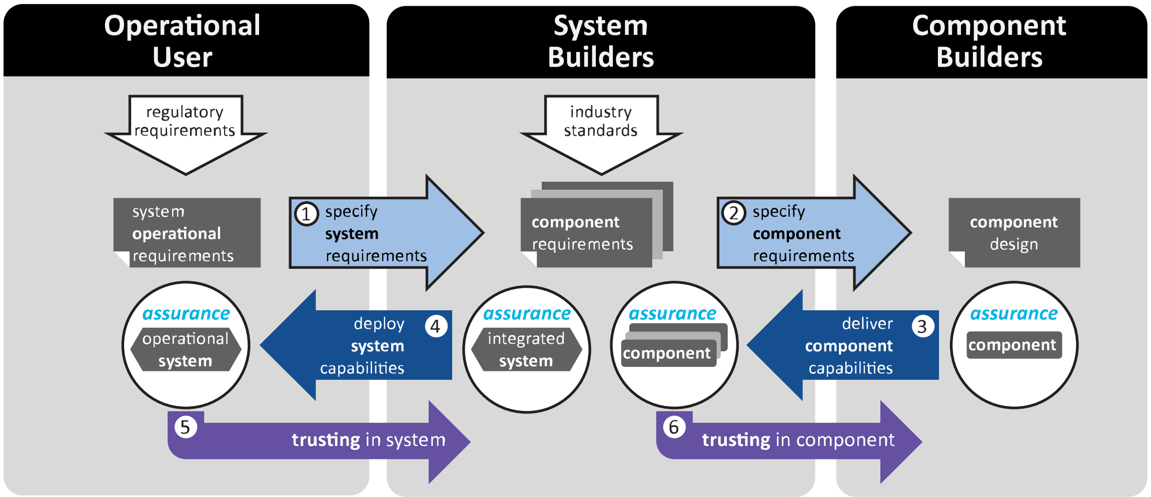


Figure 3-4: Integration and commonality of trust in the lifecycle of a complex system

The supply chain of components ends with the system builders: If a specific component is no longer in production and needs to be replaced with a similar component, the whole system architecture must be corrected and perhaps recertified, which is the duty of the system builders, not of the operational user.

Example

Most software components and services have only a limited lifetime regarding updates, which are frequently necessary to fix security issues or functionality failures. If such a component is an operating system, its replacement can lead to major architectural changes.

In reality, the relation between operational user, system builders and component builders is much more complex:

- The operational user may outsource some or all of its operations to third-party subcontractors, for example an airline as owner of an airplane outsources the maintenance of an airplane. In the extreme, the system owner may outsource the whole operation. Either way, the operational user is responsible for ensuring the system delivers its business purpose while meeting operational requirements and maintaining stated levels of trustworthiness.
- System builders may be system integrators or solution providers, integrated components may be customized components, commercial components or interfaces to services in remote systems. For example, a document management system could be installed on-premises using Microsoft Office and SharePoint server or use Microsoft Team in the Azure Cloud. They have similar capabilities but the assurance regarding server availability or frequent software updates is fulfilled by different roles.

### 3.2.3 TRUST AT COMPONENT BUILDER ROLES

Manufacturers and vendors develop technical components to sell as standard. They can be adapted for specific usage, but this is the responsibility of the system builder. The deliverer of the component is responsible for delivering the capabilities that fulfill the anticipated and implicit requirements over the lifecycle of the component. The receiver of the component is responsible for assuring its trustworthiness at the next level of the trust hierarchy.

Trust must permeate down through all the components and their subcomponents, as shown in Figure 3-5. Component builders must ensure that trust requirements are applied to each of the subcomponents as they are integrated in their own components.

These components may be delivered as a service integrating and exposing both hardware and software components. The trust in service components is assured by the fulfillments of the requirements of the service-level agreements by these components and their subcomponents. For *infrastructure as a service* (IaaS) such subcomponents may include hardware and low-level software components such as firmware and hypervisors. *Platform as a service* (PaaS) usually includes operating systems, and system components such as databases and application frameworks. Finally, *software as a service* (SaaS) may have other software subcomponents running on a third-party platform. In all three of these service offerings, the main component builder is responsible for the permeation of trust through all the subcomponents of the service.

Vendors and manufacturers seek to implement incremental value-adds to products already in the market, and so maintain the return on investment on the research and development required to implement trust. However, if the manufacturer and vendor do not implement appropriate trust mechanisms, it is difficult for the system builders and equipment operational user to implement those mechanisms later on. The trust must be designed in from the beginning.

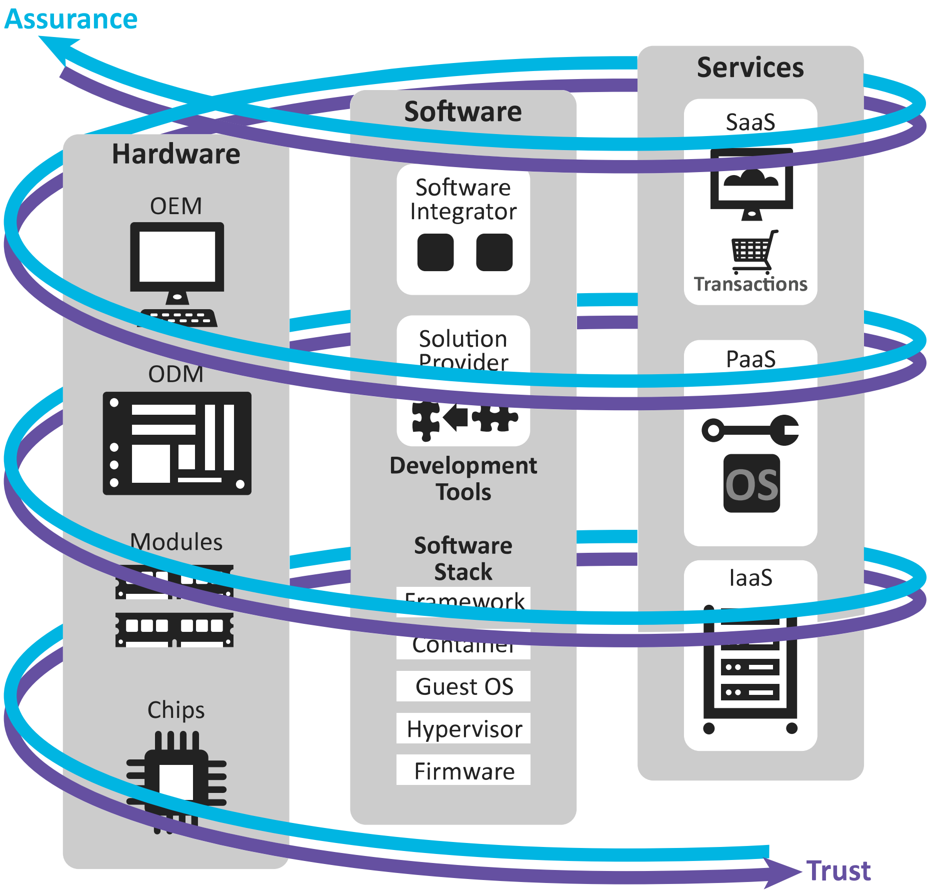


Figure 3-5: Permeation of trust between components

### 3.2.4 Trustworthiness Standards in Supply Chains

Supply chains are the basis for industry, services, and consumption in society. This is why trust and trustworthiness are the foundation of strategic national programs such as Society 5.0[1] in Japan and Industrie 4.0[2] in Germany.

As society and business increase their reliance on digital connectivity, supply chains are increasingly automated and digitized. Processes that were manual now require closely integrated systems connecting global partners digitally who may not be known to each other. Partners in a supply chain must therefore establish and maintain digital trust to have confidence in the trustworthiness of their offerings and operations.

Three key supply-chain business challenge elements are:

- difficulty of coordination and collaboration on data, applications, and system design of independent stakeholders across the global marketplace,

[1] see [Hitachi2017]
[2] see [Industrie-4.0]

- demands for reliable data, firmware and product integrity that can be trusted for the numerous transactions between the stakeholders and
- opacity of provenance due to the limited use of digital records, standardization and criteria for qualifying trustworthy suppliers.

A trustworthy system addresses these challenges at each stage in the supply chain.

Strategically injecting transparency and trustworthiness into the supply chain enables a network of independent stakeholders to trust a shared record of digital assets, transactions, software authenticity, and information. Without this transparency, enterprise supply chain participants have little or no evidence that their trust is justified.

At the enterprise level we find trust across a range of functions from the trustworthiness of data and analytics at the CXO level, to systems, to hardware connections and cyber-physical systems, to the need to connect enterprises and supply chain considerations (UST Global, Industrie 4.0 working group). Common among these views of trust are a set of trust values, including:

- Reliability
- Resilience
- Security
- Privacy
- Safety
- Quality
- Effectiveness
- Integrity
- Authenticity
- Transparency

Some of these are the trustworthiness characteristics defined in this framework, others are attributes that relate to them.

## 3.3 SOFTWARE TRUSTWORTHINESS

Software plays a critical role in both function and trustworthiness for most components. Software developers are highly trained within their specialized field, but they find themselves having to understand technical domains and disciplines that are different from their own. Software Trustworthiness Best Practices[1] contains actionable guidance to decision makers, managers and software practitioners who wish to improve the trustworthiness of their software product:

- Software design should describe the trustworthiness challenges.
- Various methods and techniques should be given to achieve software assurance.
- Software composition must be understood and managed, including the use of a software bill of materials (SBOM), a structured list of components included in the software, that can also convey origin, chain-of-custody, methods of construction, and ensure the integrity of that information.

[1] see [IIC-SWTW2020]

- Software protection can be used to enable trustworthy operation as software executes in untrustworthy environments. They can make software executables more resistant to modification as hackers attempt to alter software's expected functionality. Software protection hinders discovery of the intellectual property software contains.
- Attention to software throughout the lifecycle, including creation, operation, updating and decommissioning.

Principle

**Principle 7: Software trustworthiness must be managed throughout the entire software lifecycle.**

## 3.4 TRUSTWORTHINESS METHODS

The primary challenge of implementing trustworthiness is that none of the trustworthiness characteristics can be implemented separately and they cannot be simply combined: The characteristics may support or block each other; their combination results in new challenges.

The solution is instead of having the system design to be directly focused on the five trustworthiness characteristics to realize methods that are directly assigned to these characteristics. Such methods have been used for a long time but were not classified by trustworthiness characteristics. And this classification can be extended by other attributes.

Definition

**Trustworthiness Method**
A component, tool, technology, software application, operational procedure or management directive assigned to at least one trustworthiness characteristic.

Such methods are named as *trustworthiness safety method*, *trustworthiness resilience method* etc. If a method is assigned to several trustworthiness characteristics, the list of characteristics is separated with a slash, e. g. *trustworthiness security/privacy method*. A trustworthiness method implements features that support one or more trustworthiness characteristics.

Examples of trustworthiness methods are:

- *Fire extinguisher*: a tool and a trustworthiness safety method.
- CO2 fire suppression system*:*[1] a tool and a trustworthiness resilience method (the main purpose is to protect the system not the environment or humans; carbon dioxide[2] is indeed dangerous for humans).
- *Network firewall*: a tool and a trustworthiness security method.
- *Melt-resistant steel*: technology and a trustworthiness resilience method.

[1] see [Wiki-GaFiSup]

[2] see [Wiki-CarDio]

- *Windmill restart:* operational procedure for airplanes during an engine flameout[1] and a trustworthiness resilience method.
- *Electric motor brush replacement*: operational procedure and a trustworthiness reliability method.
- *Brushless motor*: technology and a trustworthiness reliability method.
- *Encryption of all social security numbers on servers*: management directive and a trustworthiness privacy method.

Examples of trustworthy methods assigned to several trustworthiness characteristics are:

- *Fire-resistant plastic*: technology and a trustworthiness safety/resilience method: it prevents a fire from spreading and endangering humans (safety) but also damages the system (resilience).
- *Using encrypted hard disks*: management directive and a trustworthiness security/privacy method

Most of these trustworthiness methods have existed for many years in industrial systems. The novelty is assignment to one or more of the trustworthiness characteristics and the name and such these methods can be used to implement trustworthiness in a system practically.

Principle

**Principle 8: Implementing trustworthiness means implementing trustworthiness methods**

## 3.5 TRUSTWORTHINESS SYSTEM STATUS

The *trustworthy system status* defines the health of an existing system from *normal* to *ruined* as the result of specific levels of loss of functionality. Only in the *normal* status does the system work as specified. In the next sections we delve into this status definition, ending with a universal *trustworthy system status model* (TSSM).

### 3.5.1 IDEALISTIC VIEW: A SYSTEM WITH NO THREATS

The *normal* status meets everyone's expectations on how the system should work and everyone has full trust in this system.

Even without threats, trustworthy methods are necessary. For example, most systems need maintenance, and many systems have to fulfill privacy requirements. The *normal* status may be *challenged* by operations of the system itself as shown in Figure 3-6, in terms of needing maintenance and support for example. The specific trustworthiness reliability and privacy methods ideally reject every of these challenges (shown with the *succeed* arrow) and the *normal* system status is established again. Examples for such trustworthiness methods are:

[1] see [Wiki-FlaOut]

- A combustion engine, challenged by frequent oil consumption, needs the implementation of an oil change trustworthiness reliability method.
- Standard software products, challenged by coding and design flaws, need the implementation of a software update trustworthiness reliability method.
- The privacy regulations of the system may be challenged by an operational staff error which is blocked by a trustworthiness privacy method.

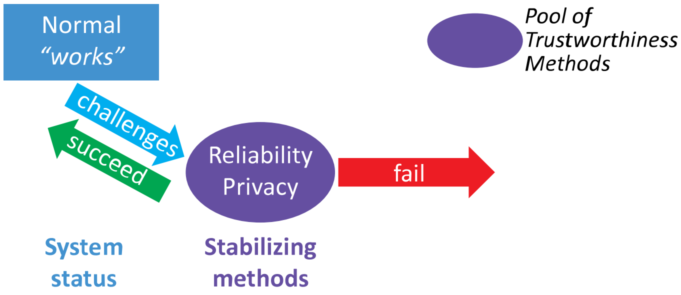


Figure 3-6: Trustworthiness in a system with no Incidents

In Figure 3-6, the purple "Reliability/Privacy" circle contains all types of trustworthy methods that are necessary to *stabilize* the trustworthy system status *normal* as long as possible.

If a challenge cannot be rejected by trustworthy methods (not working as expected or not provided), then the challenge will impact the system and the system status *fails to address the challenge* as shown with the red arrow in Figure 3-6. In this case the system leaves the *normal* status, entering the *disrupted* status.

### 3.5.2 Defending the System Against Incidents

After this idealistic core system design is completed, all potential threats must be addressed. Such threats can come from outside e.g. a hurricane, loss of power or a hacker attack, or from inside e.g. an overheated motor or a design flaw that results in erroneous behavior of the system.

A threat in general is not a problem in and of itself. For example, every electric motor has the threat of overheating and every internet access the threat of a hacker attack. Only threats actually reaching the system are relevant and need to be addressed: Such threats create incidents at the system, as shown in Figure 3-7. By implementing trustworthiness, all incidents should be rejected by trustworthiness security or safety methods. For example, a trustworthiness safety method reducing the speed of the overheated motor so it can cool down or the firewall in the router blocking the hacker attack as a trustworthiness security method. If protection is successful, the system status returns to *normal*. If the threat cannot be rejected–either because the trustworthiness methods are not working as expected or an oversight by design failure–the system status switches from *normal* to *disrupted*.

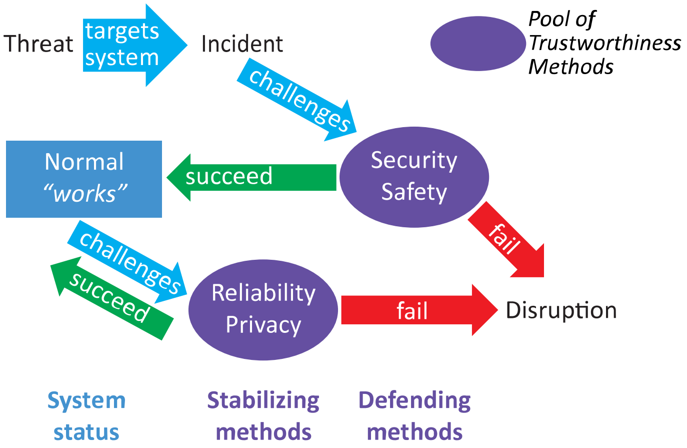


Figure 3-7: Trustworthiness in normal system status receiving an incident

### 3.5.3 Disrupted Systems

A *disrupted* system is not necessarily a serious problem. The trustworthy system status just defines this as a condition that the system is outside the *normal* status and needs some handling to be brought back to *normal*.

Example

An airplane engine flame-out situation would cause the captain to react by bringing the airplane to a lower altitude so he can try a windmill restart. After that maneuver, the pilot needs to check the entire system to find out why the engine flamed out, bring the airplane back to the original altitude and declare the problem as *solved*, and thus change the status back to *normal*.
For example, Figure 3-8 demonstrates this case: The pilot's action to bring the airplane to a lower altitude is a trustworthiness safety method, reaching the safe status of *disrupted*. The windmill restart is a trustworthiness resilience method. If one of these methods fails, the *disrupted* status cannot be continued, and the system status moves to *damaged* (because now one of the engines cannot be started again—an issue that needs deeper analysis and probably repair after a safe emergency landing).

The trustworthiness status model has symmetry, shown in Figure 3-8: *Defending methods*, assigned to security and safety, try to protect the current system status from incidents to avoid failures e.g., from *normal* to *disrupted* or from *disrupted* to *damaged. Stabilizing methods* try to defend challenges that are coming from the current status. Moreover, trustworthiness reliability or privacy methods are replaced by resilience methods as soon as the trustworthy system drops out of the *normal* status. This replacement lies in the original definitions of reliability and resilience: All reliability methods target well-known issues inside the *normal* operation of the system. As soon as the *normal* status moves to the *disrupted* stage or below, the system reaches a status of exception which needs attention to prevent further escalation: Trustworthiness resilience methods are taking over to stabilize the current status.

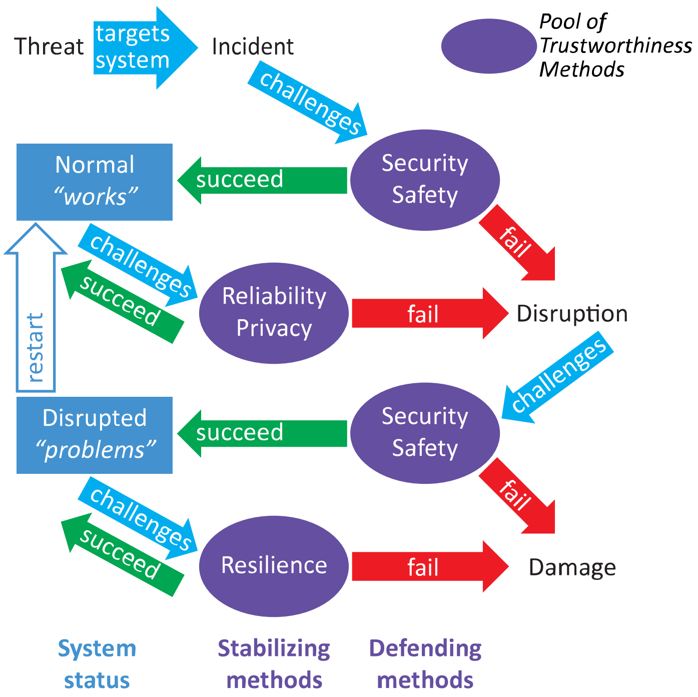


Figure 3-8: Trustworthiness in normal and disrupted system status

### 3.5.4 The Trustworthy System Status Model (TSSM)

The change from *normal* to *damaged* can be extended to further states that bring the system more and more into a fatal situation. This extension results in the *trustworthy system status model* (TSSM), shown graphically in Figure 3-9: The relationship between stabilizing and defending methods, presented by resilience resp. security and safety is extended by two more levels: *damaged* and *disastrous.* If the latter status fails, the system is permanently *ruined*.

Traditional alert colors are used to demonstrate status: *green* for *normal*, *yellow* for *disrupted*, *orange* for *damaged*, *red* for *disastrous* and *magenta* for *ruined*. The graphic also shows the required effort to move from a lower system status to higher one. A status change can also make jumps—for example from *damaged* to *normal*; to keep this graphic simple these were omitted.

Example

An example is when one airplane engine flames-out. If the trustworthiness method of bringing the airplane to a lower altitude to execute the windmill restart fails, the status would move to *damaged*. If the other engine flames out too, perhaps because the airplane ran out of fuel, the status degrades to *disastrous*. If the pilot is able to make an emergency landing (another trustworthiness safety method), the status will stay as *disastrous,* but the airplane could fly again after repair. Otherwise, the plane will crash and end as *ruined* making it clear that there is no way back to *normal*.

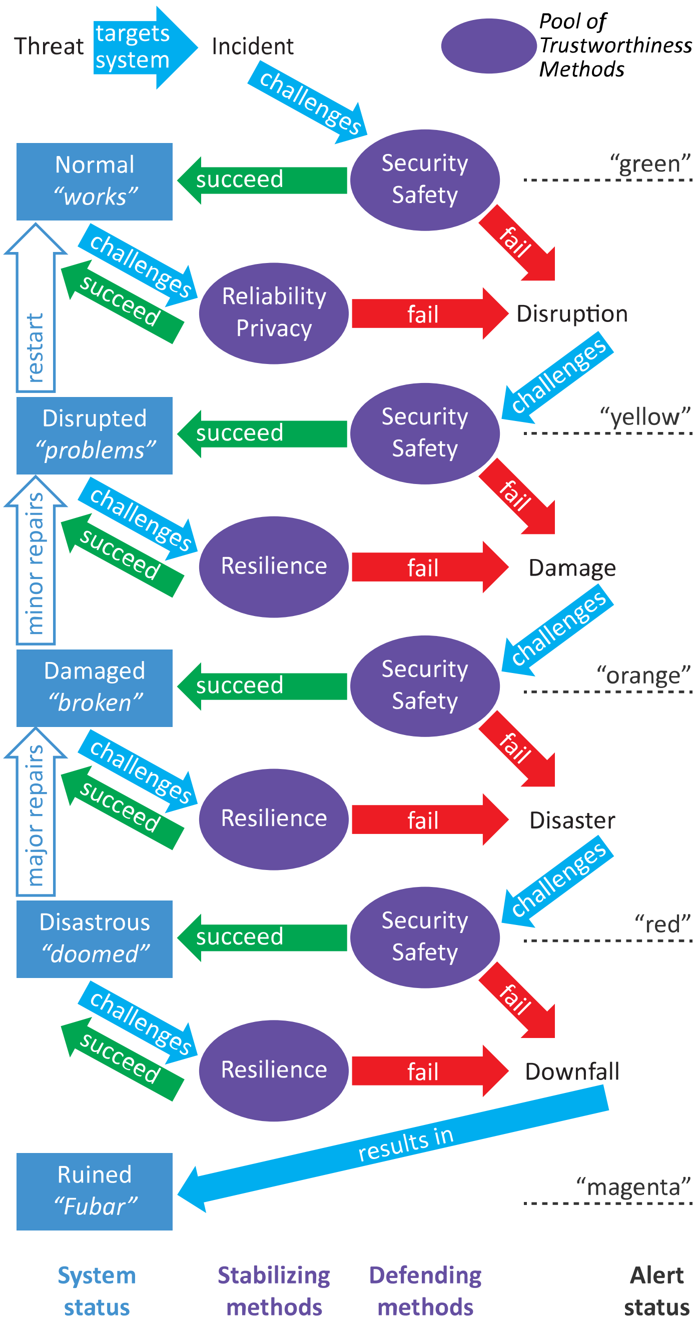


Figure 3-9: Trustworthy system status model (TSSM)

## 3.6 ASSURANCE AND EVIDENCE

A user will be able to maintain or enhance their trust in a system, when they have confidence in at least the following:

- design requirements and specifications fully and correctly address trustworthiness of the system for the system's context,

- design requirements are free of incomplete, contradictory, or untestable requirements that could cause issues with the trustworthiness during normal performance as designed,
- the system performs as designed, and continues to do so throughout the lifecycle of the system, which can be verified,
- the system has historically performed as expected, and disruptions were minimal (through historic evidence and documentation),
- the data generated by the system is authentic, timely, and has not been tampered with (through continuous trust certification) and
- the system and vendor have an ongoing good reputation based on evidence.

These items serve as a blueprint for building a trustworthy system.

Evidence provides assurance about the claims of the system's trustworthiness. An effective method to gather that evidence is *assurance case* that combines evidence from various sources, activities, subsystems and usually from across organizational boundaries. There are two main prerequisites in developing assurance cases: explicit statement(s) of the assumptions and prerequisites needed for the claims about the system's trustworthiness and second the ability to iteratively decompose the claims about the system's trustworthiness into sub claims that can, at the lowest level, be matched with evidence about the sub-claim.

Safety and trustworthiness are similar disciplines, and assurance cases can be thought of as generalizations of safety cases. Safety cases have a long history of use in such critical areas flight systems, medical devices and the rail industry. Structured assurance cases are assurance cases made in a standardized and exchangeable manner following well-known standards, like OMG Structured Assurance Case Metamodel (SACM)[1] and SCSC Goal Structuring Notation Community Standard[2] This provides a comprehensive method for addressing the safety, reliability, security, and functional requirements of systems and provides assurance that these requirements are being met.

Another important example of evidence of trustworthiness is the gathering, storage, and analysis of records from the processes that affect trustworthiness. Much of that evidence is in the form of log records or monitoring data from the components and subsystems that make up a system. These logs and data are called *digital evidence* and are useful for alerting operators to systems and components that have gone out of spec and seeing and anticipating changes over time. Such evidence can be used in a variety of ways. *Predictive maintenance* uses digital evidence to optimize maintenance based on trend analysis. *Trust ratings*, discussed in the following pages, are an emerging concept for using digital evidence to quantify and convey the trustworthiness of the rated item at a particular moment in time as a single number.

[1] see [OMG-SACM]

[2] see [SCSC-GSNCS]

# 4 Trustworthiness Conceptual Framework

## 4.1 Conceptual Framework Objectives

A *conceptual framework* for trustworthiness highlights important connections between system context, key stakeholder concerns, information sources and the process, data, and techniques to achieve evidence of trustworthiness to establish trust.

There is no one way to represent trustworthiness, and a flexible scheme is required to adapt to the relevant context. It is also important to avoid an overly complex and noisy model that obscures the key trust factors that are most important to the stakeholders and administrators of the system.

An adequate trustworthiness framework should:

- be flexible enough to include or exclude metrics as appropriate to the target context and audience,
- collect evidence to account for both the historical and the current state of the system's trustworthiness,
- allow algorithmic evaluation of trustworthiness evidence,
- support decomposition to evaluate the effects of trustworthiness in different layers of the system and
- facilitate a simple visualization with relevant information.

Such a trustworthiness framework will support basic operations such as *fusion* (positive trust factors combine to increase trustworthiness*)* and *discounting* (negative trust factors that are less important to trustworthiness). It will also allow evidence to be included, excluded and prioritized during algorithmic computation of trust ratings (section 3.6). Collection, aggregation, normalizing and adjustments to account for aging and wear are just a few of many possible methods for algorithmic weighting.

The overall approach is illustrated in Figure 4-1:

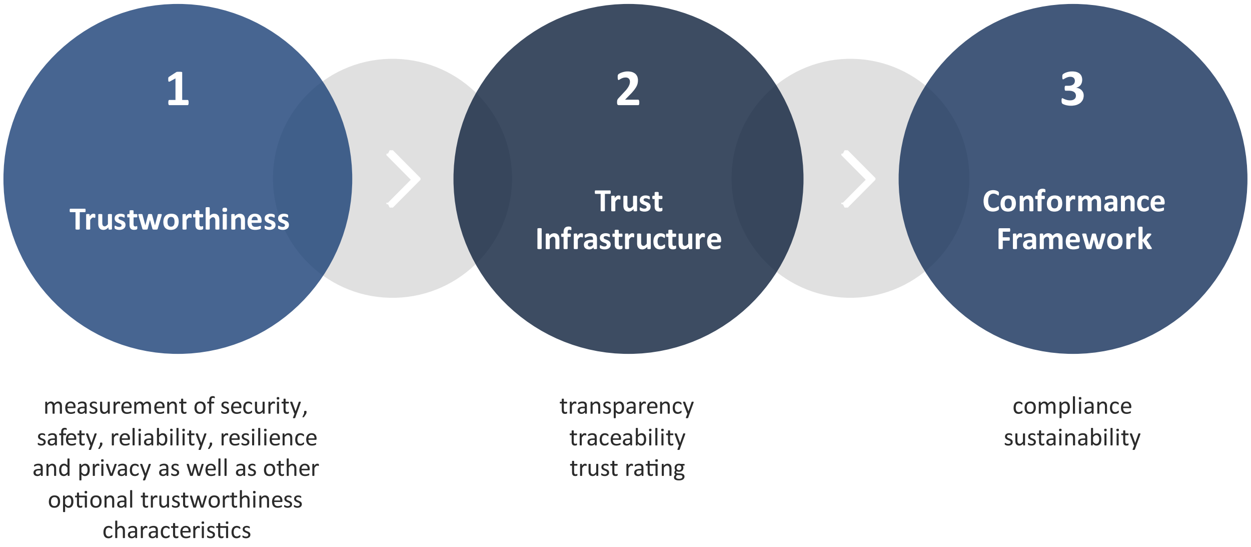

1
Trustworthiness
2
Trust Infrastructure
3
Conformance Framework
measurement of security, safety, reliability, resilience and privacy as well as other optional trustworthiness characteristics
transparency
traceability
trust rating
compliance
sustainability

Figure 4-1: Trustworthiness conceptual framework approach

Any model that meets the above requirements will need to quantify performance of a system in many operational areas, some of which may be unique to the system under consideration. A rating for each of these areas can be defined and used to track performance. These ratings can be aggregated and used as the basis for a single numerical value of trust: the *trust rating.*

When a trusted system is adopted, an initial trust rating computation establishes a baseline. During operations and through the trustworthiness journey, as the trust rating adjusts, the operator must either restore the trust rating to its original value or accept the newly generated trust rating by accepting the conditions that resulted in the new computation. In the latter case, the newly accepted level of trust represented by the changed trust rating becomes the new baseline level of trust.

Principle

**Principle 9: Maintaining change and audit records is necessary for trustworthiness.**

## 4.2 What Makes a Good Conceptual Framework?

By providing clarity on the tasks required to adjust misalignments, the conceptual framework is actionable. As trust is subjective, the conceptual framework has to incorporate the user perspective so that the modeled system mirrors the expected behavior. The flexibility can fine-tune stakeholder adjustments to reflect the desired state of the environment worthy of trust.

The goal is to establish confidence that the system performs as expected. The conceptual framework establishes and reflects the foundational trustworthiness characteristics that contribute to the system's ability to realize transparency and traceability to the key stakeholders of the business process. Conformance of the business process and structure is then attainable.

A means for measuring trustworthiness must be rigorous, pragmatic, and able to address the specific operating concerns of each organization. For example, a systemic measure of trustworthiness may rely more heavily on security attributes in a large, distributed industrial environment such as energy production than in an access-controlled production facility such as food manufacturing, where (product) safety may be more important.

A conceptual framework for trustworthiness is useful only if it can help make decisions in a timely manner, so it must be practical, quantifiable and lead to useful business results. And it must be intuitive and actionable so that organizations will come to rely on its usefulness, rather than seeing it as an interesting measure with limited value to daily operations.

Principle

**Principle 10: A trustworthiness framework must enable timely business decisions.**

Note that in a conceptual framework we do not "implement safety" (for example) but rather implement methods to achieve safety.

## 4.3 Trustworthiness Framework Components

When incomplete system requirements and design prevent the definition of a complete trustworthiness framework upfront, one can still work through some of the fundamental elements to start the process. By organizing the trust system into inputs, outputs, and systems, one can create a general understanding of components as the framework grows.

We follow a process with five essential stages associated as shown in Figure 4-2.

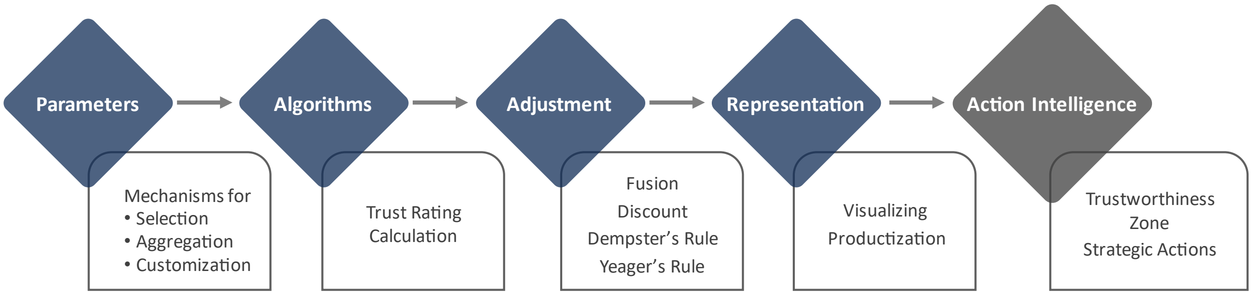


Figure 4-2: Trustworthiness dynamic flow

As shown in the figure above, the flow comprises:

*Parameters* inherent to the system, combined with mechanisms for:

- selection: A method for selecting the relevant items of trustworthiness elements that impact the overall trust of the system, i.e.
  - system levels: The compilation of component level, device level, enterprise level (including communications), supply chain network, and stakeholders and
  - device trust attributes.

- aggregation: A technique to represent the relationships between all these elements.
- customization: A way to customize the parameters to reflect the "customer's" point of view, for example through weighting.

*Algorithms*: Functions to calculate the trust rating at both micro and macro level.

*Adjustment*: Mechanisms to reflect inherent "subjective" properties of trust:

- use of fusion and discounting to allow adjustments and learning,
- whether it meets the trustworthiness zone specifications and
- traceability and transparency drive evidence and assures that the trust system has been behaving as expected and required.

*Representation*: A means to show the relative score against expected values. A method to represent and visualize the final trust scoring.

*Action intelligence*: What data is necessary to take the appropriate action to ensure trustworthiness? What actions should be taken when data values are out of specifications?

## 4.4 PARAMETERS

The trustworthiness of a system can be evaluated by decomposing it into smaller components and then evaluating the individual trustworthiness of each component as it contributes to the overall system. The framework provides the flexibility to select attributes based upon the context in which system is operating in and the aspects important to that specific use case.

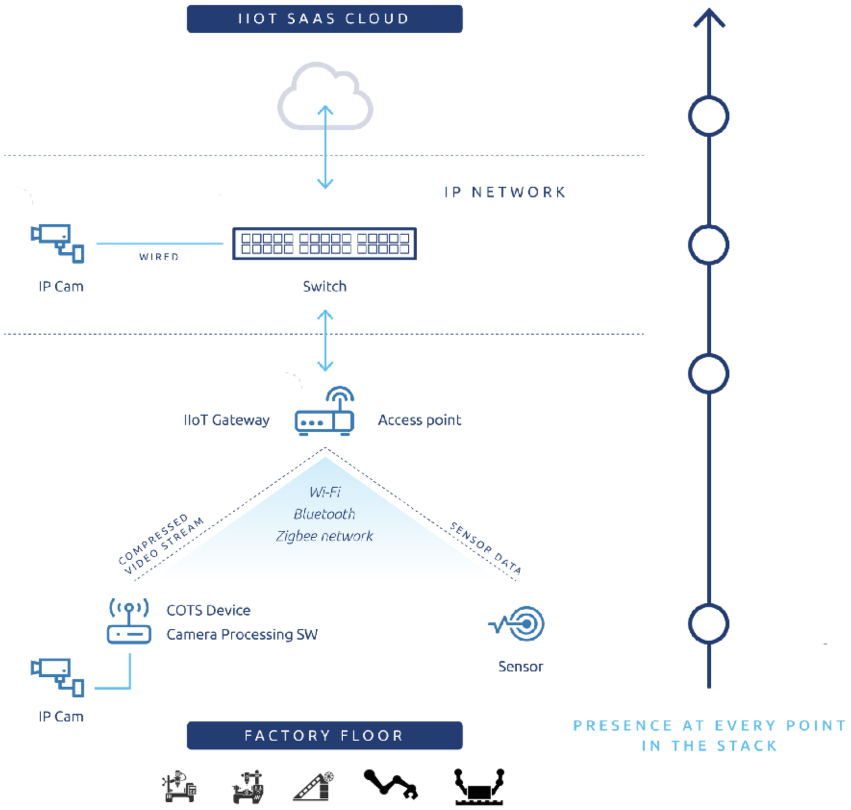


Figure 4-3: Trustworthiness framework for the collection of attributes

A system design needs to reflect the contribution of the different trustworthiness parameters from all components that make up a system correctly, such as

- devices and associated components,
- communications network devices and connectivity,
- system of systems and
- enterprise systems (management servers, software, etc.)

The following is an example of building a conceptual framework. The collection of attributes in Figure 4-3 can apply to multiple inputs towards having confidence and trust in the trustworthiness framework of the system.

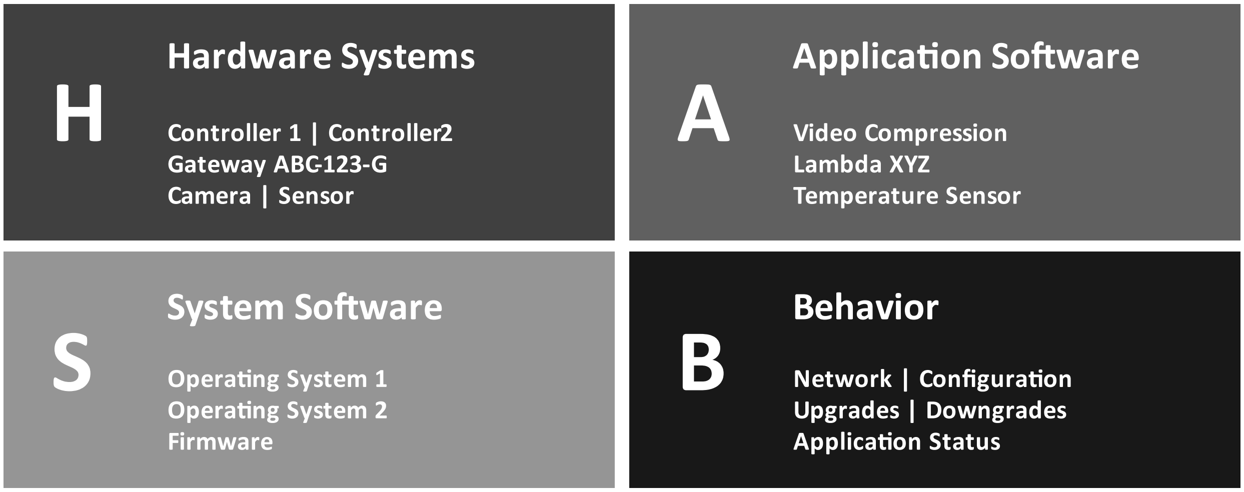
In this example, the attributes are collected from the different components including:

- cameras, sensors, end devices, edge compute devices,
- gateways and
- communication devices, e.g., WIFI AP, WAN switches and routers.

All of these components have their own trustworthiness characteristics that contribute to the total system trustworthiness by a bottom-up aggregation process.

### 4.4.1 EXAMPLE: BUILDING DEVICE UNDERSTANDING

To build device understanding the device attributes are grouped into logical sets (Figure 4-4).



H
**Hardware Systems**
Controller 1 | Controller2
Gateway ABC-123-G
Camera | Sensor

A
**Application Software**
Video Compression
Lambda XYZ
Temperature Sensor

S
**System Software**
Operating System 1
Operating System 2
Firmware

B
**Behavior**
Network | Configuration
Upgrades | Downgrades
Application Status

Figure 4-4: Grouping of device trust attributes

In the manufacturing solution in Figure 4-3, the attributes of a component within a manufacturing solution can be grouped into the following logical sets:

| | | | |
|---|---|---|---|
| System hardware attributes | System software attributes | Application software attributes | System access attributes |
| System temperature attributes | System vibration attributes | System configuration attributes | Observed pattern attributes |

Each attribute can be assigned a different weight based on how it affects the overall calculated trust rating of the component and can then be aggregated to calculate the system trust rating. Further, an application can assign weights to logical sets of attributes.

Approaches for the other system components proceed in a similar way.

## 4.5 ALGORITHMS

The framework flow involves a set of input values to compute a single trust rating value. First, compute a trust rating for each component and then the overall trust rating is a combination of the trust inputs of each component.

Trust ratings combine individual scores of all observed attributes. These attribute value inputs can be devices, group of devices, a manufacturing line, a factory, a warehouse, etc. that represent the foundational trustworthiness characteristics of what is being measured.

The approach supports fusion functionality (consolidating sources) and discounting for adjusting information based on confidence in the source. The output comprises the component trust ratings and the combined overall trust rating. This provides a measure that is straightforward to interpret and enables the user to address the root cause of issues that diminish trust.

The key benefit of this approach is that it addresses the goals of a variety of solutions flexibly.

## 4.6 Adjustment for Subjective Point of View

We must be able to customize, adjust and correlate the inputs to indicate what is important to the specific business process. The supply chain network, Figure 4-5, is a representative business process. Each business process in the supply chain has unique considerations of contextual factors and attributes that drive the analysis and management of the trustworthiness flow.

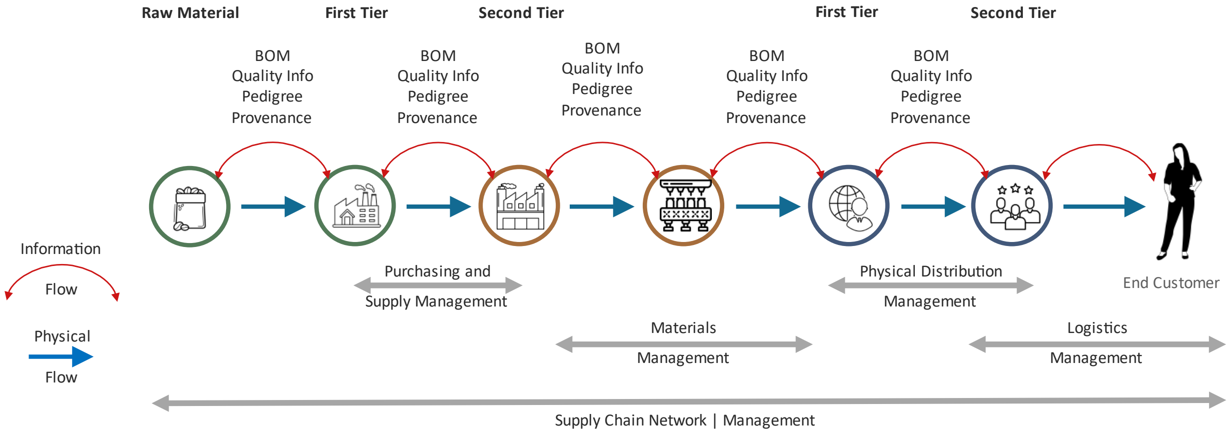


Figure 4-5: Supply chain network stakeholders

For example, a factory can designate that security, reliability, and privacy are important, while the raw material vendor is concerned with safety and reliability. We must be able to input, adjust and correlate appropriately at each stage. The conceptual framework establishes, measures, and adheres to pre-defined targets of desirable trustworthiness for different stages and states of the system.

## 4.7 Representing Trust Numerically

The trust rating reflects the interrelationships between the various activities of the supply chain for a system. A good visualization enables the business organization and its stakeholders to evaluate their business processes continuously. Trustworthiness helps the supply chain network strive and adhere to the common goal of a trusted supply chain—a business worthy of trust.

A business and its stakeholders must understand the importance of trust and the variety of factors that can impact that trust of the system or its supply chain. Transparency, conformance, and value creation of its business processes are the key fundamental elements to ensure that operations and its value chain (raw materials, manufacturing, logistics, service, and retail) are strategically aligned towards a demonstrable and measurable common goal of trustworthiness.

### 4.7.1 TRUST SEAL/TRUST CERTIFICATION

Innovation and technology provide the platform to drive guidelines to emphasize sustainability, traceability, transparency and ultimately compliance across the supply chain network. Supply chain stakeholders must have absolute confidence in its network, the dissemination of information, and the ability to onboard a new member. Moreover, the customer requires confidence in its products and suppliers so that with a product defect there is a transparent path to ensure feedback, discovery and compliance.

There is a requirement for industry certification that ensures trustworthiness of the product and its suppliers throughout the supply chain network. A company, a product, a stakeholder that is trustworthiness certified, receives a stamp of trust excellence. A trustworthiness certification is the pinnacle of trustworthiness to consumers and supply chain stakeholders.

Certifying the authenticity of components and sub-systems, both hardware and software, is one way of providing evidence of trustworthiness. Verification and certification that the system and the data it produces are authentic and has not been tampered with is possible through continuous trust certification. A systems viewpoint is necessary when creating evidence supporting trustworthiness. Certification of identity, data, processes, and systems may all be part of the evidence needed.

Principle

**Principle 11: Assurance requires a systems viewpoint with evidence of multiple factors.**

Stakeholders, employees, and consumers will look to brands for trust, assurance, and the confidence that the products are worthy of trust. The trustworthiness seal provides that confidence—the foundation of bringing trustworthiness to life.

### 4.7.2 ACTION INTELLIGENCE

Action intelligence is a measure of how much control an organization has over an activity. The trust-rating approach continually provides intelligence to help make better decisions and take actions in specific contexts that will provide valuable insights on the trustworthiness of a system.

Trustworthiness fluctuates over time, and it interacts with multivariate elements such as conformance, sustainability, integrity, quality, and assurance to name a few. Trust ratings provide data that stakeholders can adjust in the right place, at the right time and in the right way to reach the desired trustworthiness state. The real measure of trusted intelligence is the ability to act.

### 4.7.3 TRUST SYSTEM

A trustworthiness system should be able to support both the exploitation and exploration of the known and undiscovered factors and attributes of the trustworthiness flow and its business process shown in Figure 4-6. The ability to compare the desired value-creation model to the realized value-creation state is the fundamental foundation of the trust score approach. This moves beyond incremental improvement of processes to a complete trusted business model.

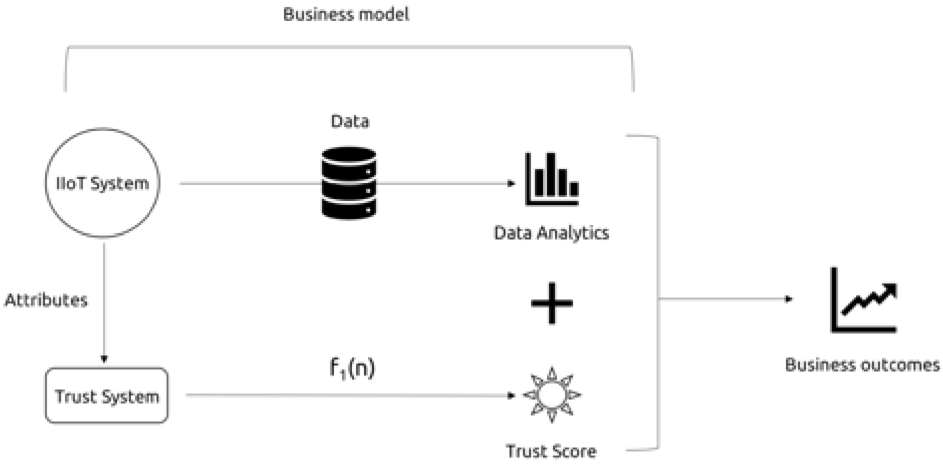


Figure 4-6: Quantifying the trustworthiness system

As shown in the figure above, the trust system is operating in parallel to the IIoT system, enabling it to collect trust data, and provide the trust context associated with the produced IoT data, and realizing a system that is worthy of trust.

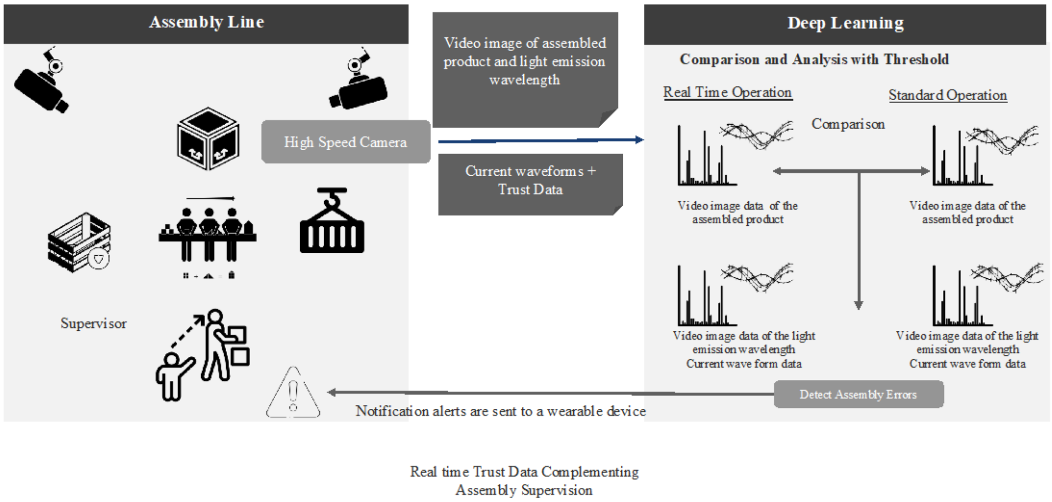

Assembly Line
Video image of assembled product and light emission wavelength
High Speed Camera
Current waveforms + Trust Data
Supervisor
Notification alerts are sent to a wearable device
Deep Learning
Comparison and Analysis with Threshold
Real Time Operation
Standard Operation
Comparison
Video image data of the assembled product
Video image data of the assembled product
Video image data of the light emission wavelength Current wave form data
Video image data of the light emission wavelength Current wave form data
Detect Assembly Errors
Real time Trust Data Complementing Assembly Supervision

Figure 4-7: Real time trust data complementing manufacturing supervision

Figure 4-7 illustrates how trust data complements manufacturing supervision in real time. In this scenario, cameras monitor the different manufacturing lines, and collected data is analyzed continuously for any anomalies. When anomalies are detected, the supervisor is notified and can intervene appropriately.

The trust system on the manufacturing site, in turn, observes the trustworthiness of the monitoring system for anomalies. When trust is determined to be low, corrective actions can be taken, either manually or automatically.

### 4.7.4 TRUST STORE AND TRUST CHAIN

Finally, to operate a supply chain with ongoing trust certification and verification, participants need an automated trust system for trust operations. Figure 4-8 illustrates how the trust system can be implemented for a supply chain scenario.

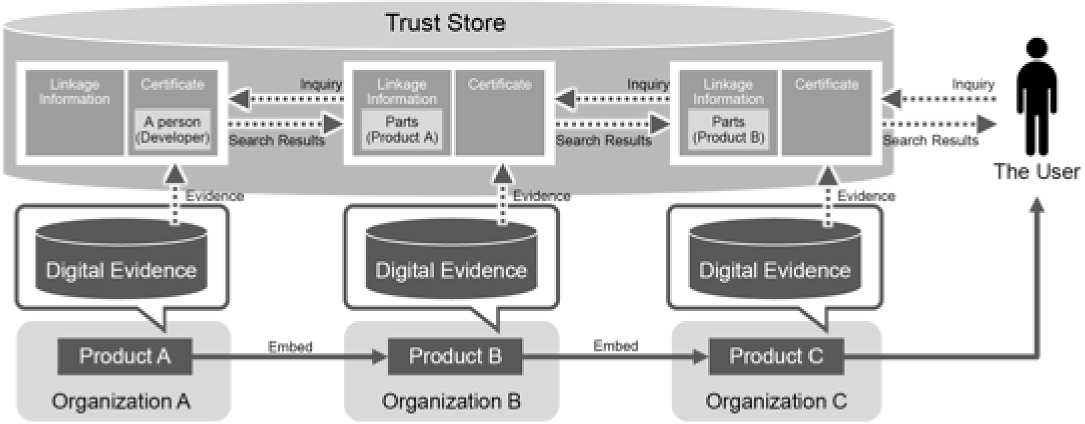


Figure 4-8: Trust system implementation for a supply chain scenario

For each organization that is part of a trustworthy supply chain, digital evidence (i.e., collected contextual trust information during production) is published by their respective trust systems to a shared information system known as a *trust store*. A trust store stores trustworthiness data and makes it available to participants.

The three main models for trust store development, operation, and governance are:

- a trusted third party, similar to an internet certificate authority or industry consortium,
- a participant in the supply chain, such as an automotive OEM, primarily for the benefit of its suppliers and partners and
- a public system that provides immutable secure record exchange between untrusted parties such as blockchain Open Transactions[1] or IOTA.[2]

A properly implemented trust store allows any authorized supplier to register, assert its trustworthiness using a model agreed by all participants and maintain its trustworthiness data over time. It allows participants to verify the trustworthiness of their suppliers based on agreed criteria, to view the history of the suppliers' trustworthiness, and be alerted to any change to the trustworthiness of a supplier.

[1] see [BCL-OpTr]

[2] see [IOTA-RoIn]

# 5 CONCLUSION AND NEXT STEPS

Trust and trustworthiness are crucial for confidence in the proper operation of nearly all systems and are especially important for industrial systems that can have negative safety and environmental impacts on people and society. The risks and negative consequences increase with increasingly networked and connected industries. In many modern industrial systems, multiple parties and systems interact automatically with minimal knowledge of each other and system functions may be hidden from users. Suppliers and service providers each have their own standards. Globally connected systems are subject to local rules and politics. Many systems have the potential for dangerous, expensive failure. For these reasons industrial internet systems require trustworthiness at every phase of design, operations, and management.

*Trustworthiness Framework* Foundations presents the basic elements of trust and trustworthiness, and the important factors and considerations needed to achieve them:

- It defines the terms trust and trustworthiness, describes essential requirements, and emphasizes the importance of context, in which critical factors of trustworthiness differ depending upon a particular system.
- It shows the interconnection between organizations and trustworthy operations, including the dependence of a system on the trustworthiness of organizations responsible for it, and approaches for organizations to increase trustworthiness over time.
- It notes that the interactions between systems can affect trustworthiness.
- It defines trust as flowing from a consumer to a supplier with the corresponding assurance and evidence of that trust as flowing from supplier to consumer.
- It illustrates how trustworthiness status can change under both correct and incorrect operation of a system.
- It emphasizes the importance of software to trustworthiness in modern systems.
- It provides a sample framework for active management of trustworthiness between the different actors (for example executive and operations or operational user and component builders). This illustrates possible methods of evaluating and rating the trustworthiness of parties, accumulating and using digital evidence, and creating a management system for enabling interaction.
- It provides examples of trustworthiness from the automotive industry, supply chain, software, battery manufacturers and other examples.
- It defines a set of principles for trustworthiness in theory and practice.

While *Trustworthiness Framework Foundations* is an overview of the IIC view of trustworthiness, it does not cover the design, construction, and operation of trustworthy systems in detail. Future papers, including the planned *Trustworthiness Practitioners Guide*, will expand upon the foundations presented here.

The IIC Trustworthiness Task Group will continue to explore the principles and practices of trustworthiness as applied to industrial systems. Participation of all IIC members and feedback and comment from all readers is welcome and encouraged.

## A. Glossary

The IIC Vocabulary[1] provides terminology and definitions for this document and other documents.

This document defines following terms at the specified pages:

**Trustworthiness ........................................ 5**

The degree of confidence one has that the system performs as expected. Characteristics include safety, security, privacy, reliability and resilience in the face of environmental disturbances, human errors, system faults and attacks. (IIC Vocabulary  )

**Safety ........................................ 8**

The condition of the system operating without causing unacceptable risk of physical injury or damage to the health of people, either directly, or indirectly as a result of damage to property or to the environment.

(ISO/IEC Guide 51:2014   as referenced in the IIC Vocabulary).

**Security ........................................ 9**

The property of being protected from unintended or unauthorized access, change or destruction ensuring availability, integrity and confidentiality.

**Reliability ........................................ 9**

Ability of a system or component to perform its required functions under stated conditions for a specified period of time.

**Resilience ........................................ 9**

Ability of a system or component to maintain an acceptable level of service in the face of disruption.

**Privacy........................................ 9**

The right of individuals to control or influence what information related to them may be collected and stored and by whom and to whom that information may be disclosed.

**Accountability ........................................ 11**

Obligation of an individual or organization to account for its activities, for completion of a deliverable or task, accept responsibility for those activities, deliverables or tasks, and to disclose the results in a transparent manner.

[1] see [IIC-Voc2020]

**Assurance** ........................................................................................................ **11**

Grounds for justified confidence that a claim has been or will be achieved.

**System**.............................................................................................................. **22**

A set of components that act together as a whole to achieve some common goal, objective, or end. A system may contain subsystems and may also be part of a larger system.

**Loss** ................................................................................................................. **22**

The negative consequence of an undesired or unplanned event.

**Accident** .......................................................................................................... **22**

An undesired or unplanned event that results in a loss, including loss of human life or human injury, property damage, environmental pollution, mission loss, etc.

**Hazard**............................................................................................................. **22**

A system state or set of conditions that, together with a particular set of worst-case environmental conditions, will lead to a loss. A potential source of harm.

**Trustworthiness Method** .................................................................................. **34**

A component, tool, technology, software application, operational procedure or management directive assigned to at least one trustworthiness characteristic.

## B. References

[Argonne2007] Argonne National Laboratory: Overview of the Design, Construction, and Operation of Interstate Liquid Petroleum Pipelines, Argonne National Laboratory, 2007-November, retrieved 2021-04-29 *https://corridoreis.anl.gov/documents/docs/technical/apt_60928_evs_tm_08_1.pdf*

[BCL-OpTr] Block Chain Lab, Chris Odom: Open-Transactions: Secure Contracts between Untrusted Parties, retrieved 2021-04-29 *https://blockchainlab.com/pdf/open-transactions.pdf*

[CD-TR] Cambridge Dictionary, Cambridge University Press, retrieved 2021-04-28 *https://dictionary.cambridge.org/us/dictionary/english/trust*

[CD-TW] Cambridge Dictionary, Cambridge University Press, retrieved 2021-04-28 *https://dictionary.cambridge.org/us/dictionary/english/trustworthy*

[FAA-737MAX] Boeing 737 MAX Flight Control System Observations, Findings, and Recommendations Submitted to the Associate Administrator for Aviation Safety, U.S. Federal Aviation Administration, 2019-10-11, Recommendation R6, retrieved 2021-04-29 *https://www.faa.gov/news/media/attachments/Final_JATR_Submittal_to_FAA_Oct_2019.pdf*

[Hardin2002] Hardin, Russell: Trust and Trustworthiness, 2002, 256 pages, retrieved 2021-04-29 *https://www.jstor.org/stable/10.7758/9781610442718*
PDF copy can be requested at *https://www.researchgate.net/publication/-227390885_Trust_and_Trustworthiness_by_RUSSELL_HARDIN_Russell_Sage_Foundation_2002_xxi_234_pages*

[Hitachi2017] Hitachi Review: Society 5.0: Aiming for a New Human-centered Society, Japan's Science and Technology Policies for Addressing Global Social Challenges, 2017-06, retrieved 2021-04-29 *https://www.hitachi.com/rev/archive/2017/r2017_06/trends/index.html*

[IETF-RFC2119] IETF: Bradner, S.: Key words for use in RFCs to Indicate Requirement Levels, 1997. Accessed 2016-06-29. *http://ietf.org/rfc/rfc2119.txt*

[IIC-IIRA2019] Industrial Internet Consortium: Industrial Internet Reference Architecture Technical Report, v 1.9, 2019, retrieved 2021-04-28 *https://www.iiconsortium.org/IIRA.htm*

[IIC-IISF2016] Industrial Internet Consortium: Industrial Internet of Things Security Framework, version 1.0, 2016-09-26, retrieved 2021-04-28
*https://www.iiconsortium.org/IISF.htm*

[IIC-JOI2018] Industrial Internet Consortium, Journal of Innovation, September 2018: articles directly relevant to Trustworthiness Framework, retrieved 2021-04-28
*https://www.iiconsortium.org/news/journal-of-innovation-2018-sept.htm*

[IIC-KSC2017] Industrial Internet Consortium: Key Safety Challenges for the IIoT white paper
*https://www.iiconsortium.org/pdf/Key_Safety_Challenges_for_the_IIoT.pdf*

[IIC-MAT2019] Industrial Internet Consortium: The Industrial Internet of Things: Managing Assessing Trustworthiness IIoT in Practice, version 1.0, 2019-07-29
*https://www.iiconsortium.org/pdf/Managing_and_Assessing_Trustworthiness_for_IIoT_in_Practice_Whitepaper_2019_07_29.pdf*

[IIC-MILS2021] Industrial Internet Consortium: MILS Architectural Approach Supporting Trustworthiness of the IIoT Solutions, 2021-04, version 1.0, retrieved 20214-29
*https://www.iiconsortium.org/pdf/MILS-Architectural-Approach-Supporting-Trustworthiness-of-IIoT-Solutions-Whitepaper.pdf*

[IIC-SMMD2020] Industrial Internet Consortium: IoT Security Maturity Model - Description and Intended Use white paper, version 1.2, 2020-05-05, retrieved 20214-29
*https://www.iiconsortium.org/pdf/SMM_Description_and_Intended_Use_V1.2.pdf*

[IIC-SMPG2020] Industrial Internet Consortium: IoT Security Maturity Model (SMM): Practitioner’s Guide, version 1.2, 2020-05-05, retrieved 2021-04-28
*https://www.iiconsortium.org/smm.htm*

[IIC-SMMPG2020] Industrial Internet Consortium: IoT Security Maturity Model Practitioner's Guide, version 1.2, 2020-05-05, retrieved 2021-04-29
*https://www.iiconsortium.org/pdf/IoT_SMM_Practitioner_Guide_2020-05-05.pdf*

[IIC-SWTW2020] Industrial Internet Consortium: Software Trustworthiness Best Practices, Version 1.0, 2020-03-23, retrieved 2021-04-29
*https://www.iiconsortium.org/pdf/Software_Trustworthiness_Best_Practices_Whitepaper_2020_03_23.pdf*

[IIC-Voc2020] Industrial Internet Consortium: The Industrial Internet of Things Vocabulary, version 2.3, 2020-10-05, retrieved 2021-04-28
*https://www.iiconsortium.org/vocab/index.htm*

[Industrie-4.0] Federal Republic of Germany: Federal Ministry for Economic Affair and Energy, Federal Ministry of Education and Research: What is Industrie 4.0?, retrieved 2021-04-29
*https://www.plattform-i40.de/PI40/Navigation/EN/Industrie40/WhatIsIndustrie40/what-is-industrie40.html*

[IOTA-RoIn] IOTA Foundation, IOTA Foundation Block: Getting Started, the Road to Integration, Part 1, 2020-01, retrieved 2021-04-29
*https://blog.iota.org/getting-started-the-road-to-integration-part-1-95cd17f92563/*

[ISO-Guide-51] International Organization for Standardization: ISO/IEC Guide 51:2014: Safety aspects—Guidelines for their inclusion in standards, 2014-04, retrieved 2017-05-29
*https://www.iso.org/standard/53940.html*

[ISO-15026-1] International Organization for Standardization: ISO/IEC 15026-1:2013: Systems and software engineering—Systems and software assurance—Part 1: Concepts and vocabulary, 2013-11, retrieved 2021-04-29
*https://www.iso.org/standard/62526.html*

[ISO-17574] International Organization for Standardization: ISO/TS 17574:2009: Electronic fee collection—Guidelines for security protection profiles, 2009-September, retrieved 2017-05-29
*https://www.iso.org/standard/52387.html*

[ISO-21089] International Organization for Standardization: ISO/TS 21089:2018: Health informatics—Trusted end-to-end information flows, 2018-04, retrieved 2021-04-19
*https://www.iso.org/standard/66936.html*

[ISO-27040] International Organization for Standardization: ISO/IEC 27040:2015: Information technology—Security technique—Storage security, 2015-01, retrieved 2017-05-29
*https://www.iso.org/standard/44404.html*

[Leveson2012] Nancy G. Leveson, John P. Thomas: Engineering a Safer World: Systems Thinking Applied to Safety, Engineering Systems, MIT-Press, ISBN 9780262016629, 2012-01, retrieved 2021-04-29
*https://mitpress.mit.edu/books/engineering-safer-world*

[Leveson2018] Nancy G. Leveson, John P. Thomas: STPA Handbook, 2018-03, retrieved 2021-04-29
*http://psas.scripts.mit.edu/home/get_file.php?name=STPA_handbook.pdf*

[Miller2015] Dr. Charlie Miller and Chris Valasek: Remote Exploitation of an Unaltered Passenger Vehicle, 2015-08, retrieved 2021-04-29
*http://illmatics.com/Remote%20Car%20Hacking.pdf*

[Miller2016] Dr. Charlie Miller and Chris Valasek: Advanced CAN Injection Techniques for Vehicle Networks, presentation at Black Hat 2016, 2016-11, retrieved 2021-04-29
*https://www.youtube.com/watch?v=4wgEmNlu20c*

[NIST-Gl-TW] National Institute of Standards and Technology (NIST), Information Technology Laboratory, Glossary: Trustworthiness, retrieved 2021-04-28
*https://csrc.nist.gov/glossary/term/trustworthiness*

[OMG-SACM] Object Management Group: Structured Assurance Case Metamodel (SACM) Version 2.1, 2020-04, retrieved 2021-04-29
*https://www.omg.org/spec/SACM/2.1/PDF*

[Schneider1999] Schneider, Fred B. (Editor), Committee on Information Systems: Trustworthiness: Trust in Cyberspace, National Academic Press, Washington D.C., 1999, retrieved 2021-04-28
*http://www.nap.edu/catalog/6161/trust-in-cyberspace*

[SCSC-GSNCS] Safety-Critical Systems Club: Goal Structuring Notation Community Standard, version 2, 2018-01, retrieved 2021-04-29
*https://scsc.uk/scsc-141B*

[Wiki-CarDio] Wikipedia: Carbon dioxide, retrieved 2021-04-29
*https://en.wikipedia.org/wiki/Carbon_dioxide*

[Wiki-FlaOut] Wikipedia: Flameout, retrieved 2021-04-29
*https://en.wikipedia.org/wiki/Flameout*

[Wiki-GaFiSup] Wikipedia: Gaseous fire suppression, retrieved 2021-04-29
*https://en.wikipedia.org/wiki/Gaseous_fire_suppression*

## C. Authors and Legal Notice

This document is a work product of the Industrial Internet Consortium Trustworthiness Task Group, co-chaired by Marcellus Buchheit (WIBU-Systems), Frederick Hirsch (Upham Security) and Robert A. Martin (MITRE).

*Editors:* Marcellus Buchheit (WIBU-Systems), Frederick Hirsch (Upham Security), Robert A. Martin (MITRE).

*Authors:* The following persons contributed substantial written content to this document:

Marcellus Buchheit (WIBU-Systems), Frederick Hirsch (Upham Security), Robert A. Martin (MITRE), Dr. Vincent Bemmel (Corlina), Antonio J Espinosa (Corlina), Bassam Zarkout (IGnPower), Charles F. Hart (Hitachi), Mitch Tseng (Tseng InfoServ).

*Contributors:* The following persons contributed valuable ideas and feedback that significantly improved the content and quality of this document: Ekaterina Rudina (Kaspersky), Daniel Young (Toshiba), Simon Rix (Irdeto), Michael Pfeifer (TÜV SÜD).

*Technical Editors:* Stephen Mellor (IIC staff) and Michael Linehan (IIC staff) oversaw the process of organizing the contributions of the Authors and Contributors into an integrated document.

## In Memoriam: Antonio J Espinosa

Our good friend Tony Espinosa died the morning of April 8, 2021, a Thursday, at a hospital in eastern Washington state at the age of 60. His loving family was at his side when he died.

Tony was a high achiever and a spirited, joyous participant in everything he pursued. He was a varsity athlete, holder of two master's degrees, and a serial entrepreneur who built and sold three companies and was working on the fourth. He traveled to all corners of the country and the world, often with his eager family in tow, and he was a lover of technology, business, the outdoors, and photography.

But while his achievements were impressive by any standard, the real Tony was far beyond the reach of any resumé. It is on the greater qualities of humanity that Tony spent most of his time, and that is how we will fondly remember him. He simply got a bang out of life. He spread elation and goodwill wherever he went, and his meetings were never tense, because he was too gracious and diplomatic to allow it. Tony was quick to laugh, instantly and deeply sympathetic, loyal, helpful, and personable to a fault. His love for his family was legendary. He stayed in close touch with his college friends and saw them often, and for newer friends like me, he was quick to trust and fully worthy of trust in return.

In fact, there was nothing more natural than Tony's interest in trustworthiness. We in the IIC knew Tony as the prolific contributor to "Trustworthiness Framework Foundations" who met every deadline for his many pages and edits. We also knew him as the founder of Corlina, Inc. and as a true supply chain innovator and seasoned practitioner. His experience ensures our work will itself be trusted, and that it will go beyond theory to steer the real-life activities of industry.

Tony and I had close to a hundred meetings over the course of our short time working together. In this time of rancor and fear, I looked forward even to his texts and voicemails, which always began with an effusive "Mon ami!" My faith in human nature was affirmed knowing a friendship can bloom and great work can be done even in the face of hardship and bad times. But we never met in person, and it is among my greatest disappointments.

Tony will live on in his lasting contributions, including in this paper. The Trustworthiness Task Group dedicates "Trustworthiness Framework Foundations," to the memory of our friend and colleague Antonio Espinosa. Thank you for everything, mon ami!

*Charles F. Hart, Hitachi, 2021-05-13*