Attributes and Dimensions of Trust in Secure Systems

Matthew Bradbury
School of Computing and Communications
Lancaster University
Lancaster, UK
m.s.bradbury@lancaster.ac.uk

Daniel Prince
School of Computing and Communications
Lancaster University
Lancaster, UK
d.prince@lancaster.ac.uk

Victoria Marcinkiewicz
IROHMS, School of Psychology
Cardiff University
Cardiff, UK
MarcinkiewiczV@cardiff.ac.uk

Tim Watson
The Alan Turing Institute
London, UK
tim.watson@turing.ac.uk

ABSTRACT

What is it to be trusted? This is an important question as trust is increasingly placed in a system and the degree to which a system is trusted is increasingly being assessed. However, there are issues with how related terms are used. Many definitions focus on one *attribute* of trust (typically behaviour) preventing that definition from being used for other attributes (e.g., identity). This is confused further by conflating what trustors measure about a trustee and what conclusions a trustor reaches about a trustee. Therefore, in this paper we present definitions of measures (trustiness and trustworthiness) and conclusions (trusted and trustworthy). These definitions are general and do not refer to a specific *attribute* allowing them to be used with arbitrary attributes which are being assessed (e.g., identity, behaviour, limitation, execution, correctness, data, environment). In addition, in order to demonstrate the complexities of describing if a trustee is designated as trusted or trustworthy, a set of *dimensions* are defined to describe attributes (time, scale, proactive/reactive, strength, scope, source). Finally, an example system is classified using these attributes and their dimensions in order to highlight the complexities of describing a system as holistically trusted or trustworthy.

CCS CONCEPTS

• **Computer systems organization → Sensor networks**; • **Security and privacy → Trust frameworks**; **Trusted computing**.

KEYWORDS

Trust, trustworthy, attributes, dimensions, limitations

ACM Reference Format:
Matthew Bradbury, Daniel Prince, Victoria Marcinkiewicz, and Tim Watson. 2022. Attributes and Dimensions of Trust in Secure Systems. In *Proceedings of 1st International Workshop on Socio-technical Cybersecurity and Resilience in the Internet of Things (STaR-IoT '22).* ACM, New York, NY, USA, 8 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Ensuring computer systems are trusted or trustworthy is important to many different entities [5], from the users of a system, to a system's owners and those involved with developing the system. In recent years we have seen many initiatives to provide trust such as with the development of trust and reputation systems in distributed systems [34], the provision of "Trusted Computing" which facilitates the restriction and attestation of software that can be run [22], and notions such as the "Trusted Computing Base" (TCB) which is "a small amount of software and hardware that security depends on" [24]. However, in attempting to provide elements of trust in computer systems, these initiatives have led to greater confusion and uncertainty due to different and conflicting ways in which entities in a system are said to be trusted or trustworthy. Additional confusion is added because many pieces of work redefine trust in order to suit their own requirements.

A further challenge is that many works, although not all, define trust in terms of behaviour (see Appendix A for a selection of definitions). This approach assesses if the *actions* taken by a system matches what was expected. Behavioural assessments typically exclude other attributes that may be useful to assess if something is trusted or trustworthy. An example is identity, which uses evidence of different types to demonstrate the identity of a system is as expected and if the identity evidence of a trustee can be verified.

In this paper we present a general definition of these concepts, with a clear difference between what it means to be trusted and what it means to be trustworthy. We define both trust and trustworthy as states (or beliefs) that a trustor holds about a trustee. Further we also define trustiness as a measure of the degree to which the trustor believes a trustee will meet its expectations and trustworthiness as a measure of the uncertainty a trustor has in the trustiness.

Using this general definition, trust and trustworthy are not specific to a single mode of assessment (behaviours, attribute, etc.). Therefore, researchers and developers are able to select multiple features of a system to be used to assess trustiness and trustworthiness to assign the respective states of trusted and trustworthy. We also highlight that because trusted and trustworthy states can be assessed using multiple features associated with different functionality or capability, it is undesirable to state a system is holistically trusted or trustworthy based on evidence for a limited number of features. Instead, trust and trustworthy state assignments should only apply to the set of functions or capabilities, under assessment using evidence sourced and analysed from features related to that function or capability.

These problems have been noted in other fields, such as with decentralised blockchain techniques [18], but there is need for a consistent and coherent framework to discuss entities being trusted and trustworthy in secure systems in general. This is especially important now due to legislative and regulation efforts in this area such as

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

STaR-IoT '22, November 07–10, 2022, Delft, Netherlands
© 2022 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-XXXX-X/18/06.
https://doi.org/XXXXXXX.XXXXXXX

UK efforts to standardise practice in securing IoT devices [9], building secure [10] and private [7] systems by design, and efforts in developing a decentralised self sovereign identity framework [30, 33]. Each of these initiatives require elements assessed to be trusted and trustworthy, but are unclear in their definitions of what this means. For example, the EU regulation on "electronic identification and trust services" [6] does not define what it means by *trust*. Future work in building trusted and trustworthy systems will be better served by a consistent approach to defining these concepts, being specific in terms of features used for assessment, and identifying techniques to assess trustiness and trustworthiness plus the interdependence of different evidence.

Therefore in this paper we set out our position on this topic to progress discussions on trust in secure systems. To do so we make the following contributions:

(1) we define the general concepts of trust, trustworthy, trustiness and trustworthiness,
(2) we explore a set of *attributes* which may be utilised for assessment of these concepts; identify the different ways these attributes can be evidenced, and their interdependencies,
(3) we provide *dimensions* to categorise each individual attribute, which can be presented in a *classification matrix*, and
(4) we identify limitations to applying trust in secure systems.

There has already been significant work on developing techniques to represent how much a system is trusted, what its reputation is, or how to represent and store evidence [34]. Therefore, this paper does not concern itself, with the representation of evidence. Instead the focus is to address higher-level concerns with secure systems being deemed to be trusted and trustworthy.

The remainder of this work is structured as follows, in Section 2 we define general concepts of trust, trustworthy, trustiness and trustworthiness; and also identify relevant attributes. To explore individual attributes in greater detail, we define dimensions for classifying attributes in Section 3 and present an example *classification matrix* in which we classify multiple attributes in an example system. Next, we discuss limitations to assessing trustiness and trustworthiness and claiming a system is trusted or trustworthy in Section 4 before presenting related work in this area in Section 5 and concluding in Section 6. In addition we provide a selection of definitions of trust from the literature in Appendix A.

2 WHAT DOES IT MEAN TO BE TRUSTED AND TRUSTWORTHY?

To begin, we will address the first problem which is the way in which trust and trustworthy have been used. There have been many approaches to defining trust (see Appendix A) and these definitions typically either reference the vulnerability an entity is willing to accept by trusting [27] or reference specifics of the problem that the work is addressing. For example, many of these definitions focus on a system performing specific actions within some constraints, i.e., that its behaviour matches expectations [1, 13, 27]. While these approaches work well for specific domains, these definitions are challenging to apply when evaluating if a system should be designated as trusted or trustworthy in a broad context.

Table 1: Example of Trust and Trustworthy being assigned to an entity in a system based on their *behaviour*.

Behaviour	Trusted	Distrusted
Trustworthy	The entity is believed to do as expected and will not deviate from that behaviour.	Do not believe that the entity will behave as expected, but expect them to reliably misbehave.
Untrustworthy	The entity will do as expected, but may deviate from expectations in how the action is performed.	The entity will not behave as expected and their misbehaviour is unpredictable and varied.

Table 2: Example of Trust and Trustworthy being assigned to an example system

	Trusted	Distrusted
Trustworthy	A bus will arrive on time at the correct stop and allow people on and off the bus.	Do not expect the bus to arrive on time, but do expect them to allow people on and off.
Untrustworthy	The bus will arrive on time, but may drive dangerously on the pavement.	The bus is not expected to arrive and has become a helicopter.

Therefore, we explicitly do not define trust in a similar manner to other works, as we intentionally do not specify *what* the assessor has belief in. Instead we rely on a general definition of trust being paired with an attribute, for which we present definitions in Section 2.1. This avoids a pitfall of works which define trust and trustworthy with an explicit focus on a subset of the attributes for which trustiness and trustworthiness could be assessed.

In this paper we use the same commonly used definitions that a trustor is an entity in a system who holds a belief about the trusted and trustworthy state of another entity in a system called the trustee [20, 34]. We now present these definitions. The first two — trustiness and trustworthiness — are what a trustor measures about a trustee. The second two — trusted and trustworthy — are the designations, or states, that a trustor holds for a trustee, based on the previous two measurements.

Definition 2.1 (Trustiness). A measurement of the attributes under consideration by the trustor to assess the ability of the trustee to meet the trustors trust expectations

Definition 2.2 (Trustworthiness). A measure of the uncertainty in the trustiness the trustor has in the trustee.

Definition 2.3 (Trusted). An entity in a system is deemed to be trusted when the trustiness is sufficiently high.

Definition 2.4 (Trustworthy). An entity in a system is deemed to be trustworthy when the trustworthiness is sufficiently high.

In summary, trustiness and trustworthiness are measurements assessed by the trustor on a trustee. Trust and trustworthy are states that a trustor holds about a trustee based on the trustiness and trustworthiness respectively. Different trustors will have different thresholds at which a trustee is considered trusted or trustworthy.

A key conceptualisation is that trust and trustworthy are not implicit properties of the trustee. They are value judgements made by the trustor. This means it is entirely possible that two trustor entities within a system can hold different states for trust and trustworthy of a third, trustee entity, despite presenting exactly the same evidence, for the same attributes being assessed. The concepts of trust and trustiness refer to the belief the trustor has in the trustee, and the concepts of trustworthy and trustworthiness refer to the level of uncertainty the trustor has in its assessment of trustiness. Trustiness and trustworthiness are what is measured when a trustor evaluates evidence that has been gathered about a trustee. An analogy with a statistical distribution is that trustiness relates to the mean and trustworthiness relates to the variance. As such this concept of uncertainty differs to “belief, desire, uncertainty” [19] where uncertainty forms part of what we have described as trustiness.

In Table 1 we have given examples of what these concepts could mean when evaluating if a system’s behaviour is trusted or trustworthy. An example of an actual system (a bus stopping at a bus stop) is given in Table 2.

2.1 Multi-attribute

These definitions have been defined in a general way because trustiness and trustworthiness need to be measured for specific attributes on a trustee. There is not one single attribute, but a set of multiple attributes that an trustor needs to measure on a trustee. In this section, we define seven attributes which will be important to assess on secure systems. However, there are likely to be other attributes for which trustiness and trustworthiness needs to be assessed that depends on the specific domain in which a system operates.

Identity The identity attribute is concerned with verifying *who* an entity is in the system is. This can be for a digital component of a system, a person, an organisation, or other types of entities with an identity. An trustee may have many identities.

Behaviour The behaviour attribute refers to whether the actions taken by the trustee in the system matches the expectations of the trustor, i.e., the trustee does what it is supposed to do.

Limitation The limitation attribute refers to whether the trustee does not exceed its capabilities and also has the ability to police and enforce those limits, i.e., the trustee does not do what it is not supposed to do.

Execution The execution attribute is where the software running on the trustee is the expected software.

Correctness The correctness attribute is where the trustee is implemented correctly and that the implementation is the *right* implementation for the purpose of the trustee.

Data The data attribute itself has sub-attributes. A trustor may need to assess the accuracy, integrity, provenance, and other sub-attributes of data provided by the trustee.

Environment The environment attribute is where the trustor expects certain conditions or state of the environment in which the trustee operates or with which it interacts.

Each of these attributes will have different degrees of importance to different systems which are deployed for different purposes. Some of these attributes are likely to be universally important, such as with identity and behaviour, as it is usually the case that an identity (whether real or a pseudonym is verifiable) is required for interactions and that a trustor will require that the behaviour of a trustee matches its expectations. From a security perspective, other attributes become critical. For example, in digital forensics it is important that a tool or system acts within specific limits. It is poor practice to trust that these limits will be followed and instead trustworthiness should be demonstrated by suitable evidence [28].

Of these attributes Behaviour, Limitation, Execution, and Correctness are the four most similar. We have divided trust into these categories because they respectively state: (1) does the system do what it is expected to do, (2) does the system not do what it is expected to not do, (3) does the system execute the expected software, and (4) is the system correctly implemented.

To highlight the differences, we now provide two examples. In the first example, a server with a memory leak is delivering content to clients, which crashes when the memory runs out. The server may initially be assessed to be behaviourally trusted before the out of memory error, but if the system was deemed to be trusted based on correctness, this would have been incorrect. In the second example, a system is behaving correctly with respect to the observations that a trustor is making. However, if the software is not as expected it may be performing other activities that were unexpected. Therefore, if the system was deemed to be trusted in terms of the software being executed this would have been incorrect.

That there are many attributes for which trustiness and trustworthiness can be assessed means there is complexity in determining if a system is trusted or trustworthy. However, when assessing just one of these attributes, there are a variety of considerations around the evidence that can be gathered. In the next section we present a collection of dimensions to describe a single attribute.

3 DIMENSIONS OF EVIDENCING AND MEASURING TRUSTINESS AND TRUSTWORTHINESS ATTRIBUTES

In order to describe these attributes, we will now define a six dimensions in which each attribute can be assessed. For each dimension, a hierarchy is presented of values that could be assigned.

3.1 Time at which Evidence is Gathered

The time over which evidence is gathered is important to consider, as the *freshness* of evidence will be important in a highly dynamic system. The hierarchy of this dimension is shown in Figure 1. Assumed means that the system is designed trusted and/or trustworthy at all times without evidence. Single means that trustiness and trustworthiness is derived from one-shot evidenced properties such as results from tests, formal proofs or other evidence that are gathered at a single point in time. Sampled is when the assessor periodically samples the subject or system to assess trustiness and trustworthiness. Continual is when the assessor continually evidences trustiness and trustworthiness.

Sampled and Continual evidence differ because when evidence is sampled it means that there are periods of time between the samples in which the trustee would otherwise lose trustiness or trustworthiness but not be detected, whereas this cannot be the case when continually evidenced [2].

Assumed ⟶ Single ⟶ Sampled ⟶ Continual

Figure 1: Hierarchy of evidence time

3.2 Scale

Trustiness and trustworthiness have been assessed on a variety of different scales [34]. However, these scales can be categorised into four levels of measurement shown in Figure 2. The nominal scale has variables without ordering, the ordinal scale has variables with ordering, the interval scale is the same as the ordinal scale but with fixed widths between variables, and the ratio scale is the same as interval scale but includes a notion of true zero. While the nominal scale has been included, we do not expect trustiness or trustworthiness to be measured using a scale that does not have an ordering of the variables which comprise the scale.

Trustiness and trustworthiness may be quantified using a probability measure (i.e., what is the probability that the behaviour will be as expected) which would fall under the interval scale. Alternatively, a number of finite variables with ordering such as "High", "Medium" and "Low" could be use, which would fall under the ordinal scale. Similarly a binary quantification of trustiness and trustworthiness would fall under the ordinal scale. Understanding which scale is appropriate, as if a composite metric is desired, it would be fundamentally invalid to arithmetically compose measurements from an ordinal and ratio scale for example. In this case suitable processes would need to be defined to evaluate whether thresholds of trust and trustworthy states had been reached.

Nominal ⟶ Ordinal ⟶ Interval ⟶ Ratio

Figure 2: Hierarchy of trustiness/trustworthiness scale

3.3 Proactive or Reactive

How trustiness and trustworthiness are assessed is important to understand what guarantees the evidence provides. Typically assessments are performed *reactively*, meaning that evidence is gathered after the trustor acts, and the *goodness* of the evidence is then evaluated. Alternatively, trustiness and trustworthiness can be assessed proactively, where the trustor takes actions before the trustee in order to obtain appropriate evidence. Neither proactive or reactive assessments (shown in Figure 3) are better than the other in general, although in different situations one may be preferred.

Proactive ↔ Reactive

Figure 3: The ways in which trustworthiness is assessed

An example of the difference between these two is now given for an example scenario of two approaches to assess behavioural trust of Edge nodes who are executing tasks that resource-constrained IoT devices have offloaded to them. The first is the reactive assessment [3] where evidence is stored in a trust and reputation system after observations have been made An alternate approach is to instead proactively send challenges from IoT devices to Edge nodes [2], however, this assumes that responding to the proactive challenge accurately reflects the trustiness of the behaviour that needs to be assessed. Hence, there is an *assumed trusted* correlation between the a correct response to the challenge and a correct response to a task. If this assumption does not hold, then the approach to assess trustiness does not work.

3.4 Evidence Strength and Scope

To determine the trustiness or trustworthiness of a specific attribute evidence needs to be provided. Different attributes will typically require different evidence. Not all evidence will be of the same quality or provide the same guarantees. For example, both hashed message authentication codes (HMACs) and digital signatures can be used to demonstrate authenticity of data, however, HMACs cannot be used to evidence non-repudiation whereas digital signatures can. Secondly, not all evidence will be sourced from the same breadth of sources. Thirdly, evidence may not be directly gathered and may be provided by another party as *reputation.*

We do not specify a hierarchy for the strength of evidence. The strength of evidence could be similar to scale on which trustiness and trustworthiness are measured. In reality strength is likely to be complicated to measure and will require statistical techniques such as with Bayesian approaches and hypothesis testing [14].

None ⟶ Local ⟶ Distributed ⟶ Global

Figure 4: Hierarchy of scope of evidence

For the evidence scope in Figure 4, None is where evidence comes from no sources, Local is where evidence comes from the single trustor assessing trust, Distributed is where evidence comes from multiple entities in a system, and Global is where evidence comes from all entities in a system. Typically, the greater the scope of evidence the stronger the evidence will be. However, depending on the type of evidence, sufficient proof may be provided with a smaller scope from which evidence is presented.

Indirect ⟶ Direct

Figure 5: Hierarchy of who has gathered evidence

Finally, some evidence will be gathered directly by the trustor. Alternatively, another entity could have gathered their own evidence and shared it with the trustor. This is reputation information which has been indirectly gathered. A challenge with reputation is that the trustiness and trustworthiness of the entity providing accurate and reliable information needs to be assessed.

3.5 Evidencing Attributes

For each of these attributes there exist specific techniques to evidence trustiness and trustworthiness which typically cannot be used to assess another attribute. This section provides a summary of the difference techniques which can be used to evidence the attributes presented in Section 2.1.

Identity can be assessed in an IoT system via the use of centralised approaches via certificate authorities which issue digital certificates and allow the verification of the authenticity of digital signatures. Alternate techniques include HMACs or signatures produced by physically uncloneable functions (PUFs). Decentralised approaches include the web of trust [37] where the trustiness of an identity is derived from a chain of peers each of whom trust the next peer in the chain. Higher-level approaches include digital identity systems (such as the EU eIDAS regulation [6]).

Behaviour can be assessed by IoT systems making observations on the actions an entity in a system takes and evaluating how well those observations match the expected observations of good behaviour. Various ways to record this information have been developed [34], typically by maintaining extensive logs.

Limitation is very similar to behaviour, as it can also be assessed by making observations on a system and evaluating if no observed action exceeds what is expected from the system.

Execution is typically assessed by checks on the code a system is executing. For example, where only digitally signed code is allowed to be executed in a privileged context. Alternately, when evidence needs to be provided to a remote entity, remote attestation can be used to evidence that specific software is running on an entity.

Correctness can be assessed by Verification and Validation (V&V) techniques. There are a wide variety of V&V different techniques from verifying properties on an abstract model of the system [21], to runtime monitors [12].

Data trustiness/trustworthiness is highly complex and there are many sub-attributes that need to be assessed, these include integrity, authenticity, availability guarantees, management of the data, the collection procedure and many others. Integrity can be protected via a variety of techniques including cyclic redundancy checks (CRCs) and also techniques used to provide identity trust (such as HMACs and digital signatures). These identity trust techniques also allow the authenticity of the data to be verified.

Environment can be assessed by directly monitoring the environment. The assessors will need to sense the environment or used evidence provided by other entities sensing the environment to assess if the conditions match expectations.

A trustor may have multiple attributes to evaluate. For example, a trustee is unlikely to have a single behaviour but multiple of them. Evaluating trustiness and trustworthiness in a single behaviour may not demonstrate trustworthiness in other behaviours, so these need to be considered separately as the evidence does not apply to multiple attributes. In addition, there are inter-dependencies among the evidence for different attributes where evidence for one attribute depends on another attribute having been evidenced. For example, assessing behaviour will rely on an assessment of identity as without knowing *who* is being assessed no behavioural evidence can be associated with the appropriate trustee.

3.6 Classification Matrix

In summary, we have proposed that for each attribute that trustiness and trustworthiness are being assessed for, there are multiple dimensions along which how trustiness and trustworthiness are being assessed can be described.

In order to demonstrate the complexity, we now present a categorisation of an example IoT system in which resource-constrained devices offload tasks to resource-rich edge devices. In this system [2], trustiness and trustworthiness is assessed by observing and recording evidence about how an edge device responds to a task. The edge device is considered to have acted badly when a task is (i) ignored, (ii) late, or (iii) has an incorrect response. Trustiness is calculated using the expected value of the Beta Reputation System, which is a continuous value between 0 and 1. An edge is considered trusted when the trustiness level is greater than or equal to the highest trustiness level of any edge minus 0.25.

The classification of this system is shown in a *Classification matrix* in Table 3. This matrix is complex with different attributes having different classifications for the different dimensions. This complexity shows that in this simple system, it is not possible to summarise it with statements such as “the system is trusted” or “the system has been demonstrated to be trustworthy” because many attributes differ in their classification and some attributes have not had trustiness evidenced. However, one approach may be to specifically define the set of attributes that are needed for a system to be considered trusted or trustworthy. In this case, it is important that the limitations are clearly identified and that such a claim does not mean that a system is holistically trusted or trustworthy.

This matrix does not include other attributes such as (i) the design of the system or hardware being correct, (ii) lack of tampering during manufacture or delivery, or if being re-used that (iii) previous owners correctly used the devices and did not damage or alter them. These are only a selection of possible additional attributes and there are likely to be many others worth considering depending on the domain in which trustiness and trustworthiness is considered.

4 LIMITATIONS TO TRUST

The goal of a system should be to have a minimal set of entities within it that need to be implicitly trusted. Instead the system should aim to demonstrate its trustiness and trustworthiness via appropriate evidence. However, due to IoT systems unique characteristics (such as limited resources) it means that challenges can often be faced by an IoT system when evidencing trustiness and trustworthiness. This may necessitate a balance between what should be assumed to be trusted/trustworthy versus demonstrated as trusted/trustworthy based on the costs to evidence an attribute and the potential impact an assumption may have.

4.1 Cost to Evidence Trustiness and Trustworthiness

Evidencing trustiness and trustworthiness can be expensive. Different techniques to provide evidence will incur costs such as: additional computation, memory usage, time delays, energy, communication, and others. Blockchain and its different approaches to distributed consensus have different costs depending on the approach used, for example, proof-of-work has very high computation costs but others such as proof-of-stake eliminate the need to solve the same problem, so have lower computation costs.

For resource-rich systems, these costs are often appropriate in order to obtain the required evidence demonstrating trustiness and trustworthiness. However, systems such as in the Internet of Things

Table 3: Example classification matrix of a proactive task offloading technique [2]

Attribute	Scale	Activity	Scope	Strength	Source	Time of Evidence
Identity[1]	Ordinal	Reactive	Distributed	High	Direct	Sampled
Behaviour[2]	Ratio	Proactive	Local	Medium	Direct	Sampled
Limitation[3]	—	—	None	—	—	Assumed
Execution[4]	—	—	None	—	—	Assumed
Correctness[5]	Varies	Proactive	Global	Low	Indirect	Single
Data Accuracy	—	—	None	—	—	Assumed
Data Integrity[6]	Ordinal	Reactive	Local	High	Direct	Sampled
Data Provenance[7]	Ordinal	Reactive	Local	High/Medium	Direct	Sampled
Environment[8]	Ratio	Reactive/Proactive	Distributed	Varies	Direct	Sampled/Continual

[1] Verify identity via digital signature and certificates, a certificate authority is needed who is trusted.
[2] Assumes behaviour to mean correctly responding to a challenge will mean the edge will correctly respond to task.
[3] While behaviour has been assessed, no attempt has been made to check that the system does not do actions it should not do. Behavioural trustworthiness assessment is limited to checking if an Edge responds correctly to a challenge.
[4] No evidence is provided about software running, nor is signed firmware required.
[5] Correctness was evidenced via experimentation, there exists the potential for incorrect operation in other scenarios not tested.
[6] Data integrity is guaranteed by protections provided by CoAP/OSCORE (HMAC) or Group OSCORE (digital signatures).
[7] Data provenance is guaranteed by protections provided by CoAP/OSCORE (HMAC) or Group OSCORE (digital signatures). HMAC has weaker guarantees than digital signatures.
[8] Aspects of the environment are checked (radio links, edge availability) via different mechanisms. However, many aspects are not checked.

are typically resource-constrained. This means that there will be a trade-off between the capability to gather evidence, the capability to check the evidence, and the capability to provide the evidence to another party within the limited resource of the device or system.

4.2 Changing over Time

Trustiness and trustworthiness are not static, they will change over time as the environment, system, beliefs of the trustor, and other considerations change. This means any designation of a trustee being trusted or trustworthy needs to be associated with (at minimum) a time at which this becomes true, but ideally will have a maximum duration for which the belief is held. An example of systems in which this is the case are X.509 digital certificates which bound the time at which the certificate is valid.

Another aspect to trustiness and trustworthiness changing over time is that if a trustor deems a trustee to be trusted or trustworthy then this may take a significant amount of time. However, a loss of trustiness or trustworthiness can occur over a much shorter period of time and then be harder to restore once diminished or lost. This means that any trustee in which trustiness and trustworthiness is assessed needs to consider the impact of choices they make on how trusted or trustworthy they will be considered in the future. However, these properties are not guaranteed and specific implementations [31] may be required to have these properties.

4.3 Bootstrapping Trust

A root of trust in a computer system is a fundamental component which needs to be trusted. A trusted platform module (TPM) is often used as a root of trust for cryptographic operations and random number generation and is commonly used to verify the integrity and authenticity of firmware used to boot an operating system [17]. Bootstrapping trust for a TPM [29] and evidencing that a TPM is trusted and trustworthy is a challenging problem as reaching a state of trust in a root of trust comes with disadvantages [25]. This is concerning as a loss of trust in this component would mean the invalidation of the evidence it has produced and previously trusted components would no longer be considered as trusted.

This issue is not limited to roots of trust in hardware, this issue occurs in other domains too. For example, certificate authorities (CA) who issue digital certificates need to provide a root certificate that is trusted. The CA needs to be trusted along multiple dimensions, including that they only issue certificates to valid entities, that they manage their private keys correctly, that correct certificate revocation lists are updated and published, plus many other attributes. The root certificate is needed to validate the authenticity of certificates which claim to be issued by that CA, hence trust of identity is built upon a root of trust whose trustiness cannot be assessed using the same techniques as other certificates [4].

A variety of other system components encounter the same issue. In 1984 Thompson wrote "You can't trust code that you did not write yourself" [35] which followed from his work where compilers can intentionally miscompile code to introduce vulnerabilities or behaviour different to the behaviour expected. The problem is that compilers are typically trusted without users checking the compiler's output to assess its trustiness and trustworthiness in generating compiled code. This problem is partially mitigated via Diverse Double-Compiling [36]. In this technique a diverse set of compilers is used to compile code. However there still exist challenges when compilers do not output reproducible builds (either due to non-deterministic behaviour or the output is based on varying inputs such as time) and the potentially limited number of compilers for less common languages[1].

[1] https://www.awelm.com/posts/evil-compiler/

The problem with bootstrapping trust is that assessing trustiness and trustworthiness of a trustee in which trust is being bootstrapped means there is a circular dependency, i.e., the trustee needed to be trusted in order for its trustiness and trustworthiness to be assessed. This means that bootstrapping trust often needs to evaluate trustiness and trustworthiness outside of the established trust framework. For example, by tracking if CAs correctly issue certificates [23], which is outside of the scope of verifying the authenticity of a digital certificate using a CA's root certificate.

4.4 Unknown Unknowns

It is the case of evaluating trustiness and trustworthiness (as it is with security) that unknown issues that we do not know we don't know will not be able to be addressed. For security, this means that there are vulnerabilities (or classes of vulnerabilities) we have not yet discovered and do not know to mitigate. This impacts assessment of trustiness and trustworthiness, as there is evidence that should be gathered to demonstrate the level of trustiness and trustworthiness in specific attributes, but it is not known that this should be done. This means any assessment of a system being trusted or trustworthy is performed based on incomplete knowledge.

4.5 Systems of Different Criticalities

Different domains will have different thresholds to which trustiness and trustworthiness must be demonstrated. A smart bulb, autonomous car, and nuclear power plant each has different levels of trustiness and trustworthiness that need to be demonstrated in order for a system to be deemed trusted or trustworthy. In more critical domains, more information will typically be provided to demonstrate trustiness and trustworthiness to stakeholders. Therefore, a designation of trust or trustworthy in one system is not necessarily comparable to another due to the different requirements for a system to be designated as trusted or trustworthy.

4.6 Freedom

Another side to assessing trustiness and trustworthiness is that in some scenarios it is not a benefit to all parties. For example, an organisation may want to place restrictions on the software that can be run on certain hardware either to prevent the misuse of the hardware or to mitigate vulnerabilities (e.g., rootkits) that the users may face. This could be achieved via a number of mechanisms such as requiring signed firmware or remote attestation. However, this means that users may be limited in how they can use their device, potentially preventing them from making legitimate choices in terms of the software that they runs [32]. This also has a sustainability impact as there is the potential for the reuse of devices to be prevented as maintenance may not be able to be turned over to open source communities.

5 RELATED WORK

There have been several other works that have proposed similar frameworks for discussing and reasoning about trust. many of these related works focus on the techniques to assess trustiness or trustworthiness [34]. However, other works have tried to classify the way in which trust can be assessed.

A set of trust classes was proposed by Jøsang et al. [20] (provision, access, delegation, identity, context, purpose trust) based on, but different to, the classes proposed by Grandison and Sloman [15] (resources, provision of service, certification, delegation, and infrastructure). These trust classes are similar to the attributes proposed in this work and overlap in some cases (such as identity, and context mapping to our environment attribute). Many of the classes proposed fall under our behaviour attribute. For example, "provision" relates to a service or resource provided, "access" is the ability to access resources, "delegation" is concerned with tasks delegated to another agent, these can all be attributed to the behaviour of the trustee. The execution, limitation, and correctness attributes we have proposed encapsulate different elements of trust compared to the proposed classes. However, as indicated by this previous work and our work, these lists are not exhaustive and each framework can make use of additional attributes as needed.

Daubert et al. [8] proposed a different set of trust attributes of: device, processing, connection, system, and multidimensional trust. Each of these dimensions have their own definition as to what trust in that context means. However, some of these definitions do relate to trust, for example, device trust "refers to the need to interact with reliable devices such as sensors and actuators" [8]. A "need" does not describe what needs to be assessed as trusted or trustworthy in the devices being interacted with. In addition, this paper circularly defines trust as "a measurement for the need of trust".

6 CONCLUSIONS

We have presented an alternate view for describing trust. Instead of using previous definitions, we have defined trustiness and trustworthiness as measures of belief and uncertainty made by a trustor about a trustee respectively, and trusted and trustworthy as states placed by the trustor on the trustee based on these measures. This alternate definition is intentionally general, such that it can be paired with an appropriate attribute. We have defined 7 attributes we believe to be important to secure systems: identity, behaviour, limitation, execution, correctness, data, and environment. Next we defined six dimensions to evidencing and measuring the trustiness and trustworthiness of a specific attribute: time, scale, proactive/reactive, strength, scope, and source. The complexity in deciding if a system is trusted or not was highlighted using an example system of task offloading based on an assessment of trustiness.

Our conclusions are that systems should avoid being described as holistically trusted or trustworthy, but trusted or trustworthy for specific attributes and that an assessment of trustiness and trustworthiness in a specific attribute is complex due to the multiple dimensions involved with assessing them. For future work, we intend to explore the implications of defining trustiness and trustworthiness as analogous to the mean and variance of a probability distribution. We believe that this opens up scope to perform further rigorous analysis and quantification.

A TRUST DEFINITIONS

To give an understanding of the diversity of trust definitions that have been used in the literature, a selection are included here.

- "I trust you because your interests encapsulate mine" Hardin [16]

- "willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party" Mayer et al. [27]
- "Trust is the expectation of an entity with respect to certain properties or actions of another entity under a specified context and time, considering the risks, incentives, and historical information." Lee [26]
- "Risk, or meaningful personal investment, is a prerequisite of trust. The need for trust only arises in risky situations, and the trustor must be cognizant of the risks involved" Deutsch [11]
- "trust (or, symmetrically, distrust) is a particular level of the subjective probability with which an agent assesses that another agent or group of agents will perform a particular action, both before he can monitor such action (or independently of his capacity ever to be able to monitor it) and in a context in which it affects his own action" Gambetta [13]
- digital trust as "a trust based either on past experience or evidence that an entity has behaved and/or will behave in accordance with the self-stated behaviour." Akram and Ko [1]

ACKNOWLEDGMENTS

This work was supported in part by Security Lancaster and the PETRAS National Centre of Excellence for IoT Systems Cybersecurity [EPSRC Grant EP/S035362/1].

REFERENCES

[1] Raja Naeem Akram and Ryan K. L. Ko. 2014. Digital Trust - Trusted Computing and Beyond: A Position Paper. In *IEEE 13th International Conference on Trust, Security and Privacy in Computing and Communications*. 884–892. https://doi.org/10.1109/TrustCom.2014.116

[2] Matthew Bradbury, Arshad Jhumka, and Tim Watson. 2021. Trust Trackers for Computation Offloading in Edge-Based IoT Networks. In *IEEE INFOCOM*. Vancouver, Canada, 1–10. https://doi.org/10.1109/INFOCOM42981.2021.9488844

[3] Matthew Bradbury, Arshad Jhumka, Tim Watson, Denys Flores, Jonathan Burton, and Matthew Butler. 2022. Threat-Modeling-Guided Trust-Based Task Offloading for Resource-Constrained Internet of Things. *ACM Transactions on Sensor Networks* 18, 2, Article 29 (2022), 41 pages. https://doi.org/10.1145/3510424

[4] Jeremy Clark and Paul C. van Oorschot. 2013. SoK: SSL and HTTPS: Revisiting Past Challenges and Evaluating Certificate Trust Model Enhancements. In *IEEE Symposium on Security and Privacy*. 511–525. https://doi.org/10.1109/SP.2013.41

[5] Stacey Conchie. 2022. Trust In Security Contexts. *CREST Security Review* Summer (2022), 4–5. https://crestresearch.ac.uk/comment/trust-in-security-contexts/

[6] Council of European Union. 2014. Regulation (EU) No 910/2014 of the European Parliament and of the Council. https://eur-lex.europa.eu/eli/reg/2014/910/oj

[7] Andrew Crabtree, Hamed Haddadi, and Richard Mortier (Eds.). 2021. *Privacy by Design for the Internet of Things: Building accountability and security*. Institution of Engineering and Technology. https://doi.org/10.1049/PBSE014E

[8] Jörg Daubert, Alexander Wiesmaier, and Panayotis Kikiras. 2015. A view on privacy & trust in IoT. In *IEEE International Conference on Communication Workshop (ICCW)*. London, UK, 2665–2670. https://doi.org/10.1109/ICCW.2015.7247581

[9] Department for Digital, Culture, Media and Sport. 2018. *Code of Practice for Consumer IoT Security*. Guidance. HM Government. https://www.gov.uk/government/publications/code-of-practice-for-consumer-iot-security

[10] Department for Digital, Culture, Media and Sport. 2018. *Secure by Design: Improving the cyber security of consumer Internet of Things Report*. Report. HM Government. https://www.gov.uk/government/publications/secure-by-design-report

[11] Morton Deutsch. 1977. *The Resolution of Conflict: Constructive and Destructive Processes*. Yale University Press. 448 pages.

[12] Angelo Ferrando, Rafael C. Cardoso, Michael Fisher, Davide Ancona, Luca Franceschini, and Viviana Mascardi. 2020. ROSMonitoring: A Runtime Verification Framework for ROS. In *Towards Autonomous Robotic Systems*. Springer, Cham, 387–399. https://doi.org/10.1007/978-3-030-63486-5_40

[13] Diego Gambetta. 1988. *Trust: Making and Breaking Cooperative Relations*. Basil Blackwell Ltd, Oxford, UK, Chapter Can We Trust Trust?, 213–237.

[14] Steven N. Goodman. 2005. Introduction to Bayesian methods I: measuring the strength of evidence. *Clinical Trials* 2, 4 (2005), 282–290. https://doi.org/10.1191/1740774505cn098oa PMID: 16281426.

[15] T. Grandison and M. Sloman. 2000. A survey of trust in internet applications. *IEEE Communications Surveys Tutorials* 3, 4 (April 2000), 2–16. https://doi.org/10.1109/COMST.2000.5340804

[16] Russell Hardin. 2002. *Trust and Trustworthiness*. Russell Sage Foundation, New York, USA.

[17] International Organization for Standardization 2015. *Information technology — TrustedPlatform Module Library*. Standard ISO/IEC 11889:2015. International Organization for Standardization, Geneva, CH. https://www.iso.org/standard/66510.html

[18] Mattis Jacobs. 2021. How Implicit Assumptions on the Nature of Trust Shape the Understanding of the Blockchain Technology. *Philosophy & Technology* 34, 3 (2021), 573–587. https://doi.org/10.1007/s13347-020-00410-x

[19] Audun Jøsang and Roslan Ismail. 2002. The Beta Reputation System. In *15th Bled Electronic Commerce Conference*. University of Maribor Press, Bled, Slovenia, 14 pages.

[20] Audun Jøsang, Roslan Ismail, and Colin Boyd. 2007. A survey of trust and reputation systems for online service provision. *Decision Support Systems* 43, 2 (2007), 618–644. https://doi.org/10.1016/j.dss.2005.05.019

[21] Jack P.C. Kleijnen. 1995. Verification and validation of simulation models. *European Journal of Operational Research* 82, 1 (1995), 145–162. https://doi.org/10.1016/0377-2217(94)00016-6

[22] Boyu Kuang, Anmin Fu, Willy Susilo, Shui Yu, and Yansong Gao. 2022. A survey of remote attestation in Internet of Things: Attacks, countermeasures, and prospects. *Computers & Security* 112 (2022), 102498. https://doi.org/10.1016/j.cose.2021.102498

[23] Deepak Kumar, Zhengping Wang, Matthew Hyder, Joseph Dickinson, Gabrielle Beck, David Adrian, Joshua Mason, Zakir Durumeric, J. Alex Halderman, and Michael Bailey. 2018. Tracking Certificate Misissuance in the Wild. In *IEEE Symposium on Security and Privacy*. 785–798. https://doi.org/10.1109/SP.2018.00015

[24] Butler Lampson, Martín Abadi, Michael Burrows, and Edward Wobber. 1991. Authentication in Distributed Systems: Theory and Practice. In *Proceedings of the Thirteenth ACM Symposium on Operating Systems Principles* (Pacific Grove, California, USA) *(SOSP '91)*. Association for Computing Machinery, New York, NY, USA, 165–182. https://doi.org/10.1145/121132.121160

[25] Hagen Lauer, Amin Sakzad, Carsten Rudolph, and Surya Nepal. 2019. Bootstrapping Trust in a "Trusted" Virtualized Platform. In *Proceedings of the 1st ACM Workshop on Workshop on Cyber-Security Arms Race* (London, United Kingdom) *(CYSARM'19)*. Association for Computing Machinery, New York, NY, USA, 11–22. https://doi.org/10.1145/3338511.3357347

[26] Insup Lee. 2016. Trust Management for Cyber-Physical Systems. In *Robotics: Science and Systems 2016 Workshop — Social Trust in Autonomous Robots*. http://qav.comlab.ox.ac.uk/trust_in_autonomy/img/LeeTrustWorkshop16.pdf

[27] Roger C. Mayer, James H. Davis, and F. David Schoorman. 1995. An Integrative Model of Organizational Trust. *The Academy of Management Review* 20, 3 (1995), 709–734. https://doi.org/10.2307/258792

[28] Christopher Neale, Ian Kennedy, Blaine Price, Yijun Yu, and Bashar Nuseibeh. 2022. The case for Zero Trust Digital Forensics. *Forensic Science International: Digital Investigation* 40 (2022), 301352. https://doi.org/10.1016/j.fsidi.2022.301352

[29] Bryan Parno. 2008. Bootstrapping Trust in a "Trusted" Platform. In *Proceedings of the 3rd Conference on Hot Topics in Security* (San Jose, CA) *(HOTSEC'08)*. USENIX Association, USA, Article 9, 6 pages.

[30] Nick Pope, Michał Tabor, Iñigo Barreira, Nicholas Dunham, Franziska Granc, Christoph Thiel, and Arno Fiedler. 2022. *Digital Identity: Leveraging the SSI Concept to Build Trust*. European Union Agency for Cybersecurity (ENISA). https://doi.org/10.2824/8646 TP-09-22-024-EN-N.

[31] Guntur Dharma Putra, Volkan Dedeoglu, Salil S. Kanhere, and Raja Jurdak. 2020. Trust Management in Decentralized IoT Access Control System. In *2020 IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*. 1–9. https://doi.org/10.1109/ICBC48266.2020.9169481

[32] Seth Schoen. 2003. *Trusted Computing: Promise and Risk*. Technical Report. Electronic Frontier Foundation. https://www.eff.org/files/20031001_tc.pdf

[33] Manu Sporny, Dave Longley, Markus Sabadello, Drummond Reed, Orie Steele, and Christopher Allen. 2022. *Decentralized Identifiers (DIDs) v1.0*. W3C Recommendation. W3C. https://www.w3.org/TR/2022/REC-did-core-20220719/

[34] Phillip Taylor, Lina Barakat, Simon Miles, and Nathan Griffiths. 2018. Reputation assessment: a review and unifying abstraction. *The Knowledge Engineering Review* 33 (2018), e6. https://doi.org/10.1017/S0269888918000097

[35] Ken Thompson. 1984. Reflections on Trusting Trust. *Commun. ACM* 27, 8 (aug 1984), 761–763. https://doi.org/10.1145/358198.358210

[36] David A. Wheeler. 2009. *Fully Countering Trusting Trust through Diverse Double-Compiling*. Ph. D. Dissertation. George Mason University. https://mars.gmu.edu/handle/1920/5667

[37] Philip R. Zimmermann. 1995. *The Official PGP User's Guide*. MIT Press, Cambridge, MA, USA.