CREATING A FRAMEWORK FOR SUPPLY CHAIN TRUST IN HARDWARE AND SOFTWARE

A Report of the Lawfare Institute's Trusted Hardware and Software Working Group

May 2022

TASK FORCE MEMBERS

Paul Rosenzweig, Chief Rapporteur
Red Branch Consulting PLLC

Justin Sherman, Assistant Rapporteur
Atlantic Council

Benjamin Wittes, Task Force Chairman
The Brookings Institution and *Lawfare*

Trey Herr, Task Force Member
Atlantic Council

David Hoffman, Task Force Member
Sanford School of Public Policy at Duke University

Herb Lin, Task Force Member
Stanford University

Bart Preneel
The University of Leuven (Belgium)

Samm Sacks, Task Force Member
Paul Tsai China Center at Yale Law School

Fred B. Schneider, Task Force Member
Cornell University

Daniel Weitzner, Task Force Member
MIT Internet Policy Research Initiative

TABLE OF CONTENTS

In a world of growing dependence on technology, consumers of information and communications technology (ICT) goods face increasingly important questions: How, and to what extent (if any), can they be confident that the systems on which they rely are worthy of trust? One need only think of the controversies surrounding hardware and software systems manufactured in China but used in Western commerce to understand the political and practical salience of the problem. To answer that question, the Lawfare Institute convened a working group of experts to articulate and justify a set of trustworthiness principles—concepts that, ex ante, would justify accepting a digital artifact as worthy of being trusted. Although we concluded that a dispositive assessment of trustworthiness would never be feasible, the report develops a comparative checklist of steps an organization can take that significant stakeholders might agree demonstrates its products to be trustworthy—what one might call a functional definition of trustworthiness. Even without the prospect of precisely assessable levels of trustworthiness, the report concludes that a framework for assessments can be made with a relatively high degree of confidence.

The value of a framework based on agreed-upon principles should be evident. Using these principles—as well as acceptable evidence—as a guideline, ICT manufacturers and users, including organizations and consumers, can analyze comparative risks and make reasoned risk-benefit and resource-allocation decisions. The framework identifies multiple principles of trustworthiness organized around four core values: maximize transparency, ensure accountability, allow for independence of evaluation and prefer provable analytic means of trust verification over axiomatic, unverifiable means of assessment.

INTRODUCTION[1]

In a world of growing dependence on technology, consumers of information and communications technology (ICT) goods face increasingly important questions: How, and to what extent (if any), can they be confident that the systems on which they rely are worthy of trust?

At the most basic level, when we determine the trustworthiness of a digital artifact (a component, a device, or a system), we are attempting to describe that artifact in terms of its security. We say that a

[1] In the fall of 2019, the Intel Corporation asked the Lawfare Institute to convene an expert panel to study the question of trustworthiness in hardware and software, to try to establish and describe the constituent elements of trustworthiness, and to identify a set of principles by which one might determine how trustworthy a product is. Intel provided funding for this working group, which was chaired by Benjamin Wittes. Two representatives of Intel—one of whom left the company during the working group's deliberations—were members of the group and involved in its proceedings. This report represents the work product and thinking of the working group as a whole and was not subject to editorial control or oversight by Intel or any other outside body or organization. The working group's rapporteur was Paul Rosenzweig, who chiefly drafted this document along with Justin Sherman.

digital artifact that *does what is expected of it and nothing more* is trustworthy.[2] The challenge is not so much in defining the desired end-state of trustworthiness but, rather, in defining how it is that one may demonstrate trustworthiness to a skeptical world. In modern systems, all artifacts are the sum of multiple parts—so the inquiry of trustworthiness is an inquiry into creation, into assembly out of diversely produced components, into distribution, and into use. In effect, it is to ask about the entire supply chain of ICT goods from conception to consumption.

Given this complexity, one may reasonably ask whether it is even possible to devise a means of validating whether a digital artifact is a trustworthy piece of hardware or software. Some might despair of the effort. But in doing so they would, necessarily, condemn a significant fraction of the world economy to perpetual indeterminacy about security. In our view, that prospect is too grim. So, we offer this report as an effort to begin conceptualizing a framework for trustworthiness. Our tentative answer to this problem resonates in three different dimensions.

First, the inquiry implicates questions of technical capacity and security: How are we to know that the manufacturers of a hardware or software system have designed and built it in a way that is trustworthy and, therefore, secure against deliberate internal or external attack? In other words, has the manufacturer performed competently in constructing the digital artifact? Has the enterprise performed transparently? And has it used only those suppliers that have, in turn, also performed competently and transparently?

Second, the inquiry implicates the question of organizational intent as reflected in a corporate governance structure: How are users to be assured that manufacturers have not constructed and marketed a system that affords either themselves or some other party privileged access and control, without the users' knowledge? In other words, does the manufacturer see a value to be gained for itself from the design, to the disadvantage of the consumer?

On yet a third axis of inquiry, the problem of trust can be seen as a question of policy and law: What protections exist against inappropriate intervention into the manufacture or operation of an ICT system? And, relatedly, what are the mechanisms for determining the appropriateness of such

[2] We note here, at the outset, that this security question is a subset of a broader reliability question, which asks whether the digital artifact is secure against error, mistake, or natural disruption. Those, too, might be reasons why a device does not do what it is supposed to do or does that which it is not supposed to do. But these questions are of a different character from the trustworthiness questions we address in this report, in which we focus on the security of an artifact and the willful attempt of actors to manipulate the artifact. Failures, natural disruption, and other accidents and errors are also part of trustworthiness, but they are not our concern here.

interventions to the extent they may exist? Are there flaws in the system that some third party, for whatever reason, can benefit from exploiting?

In suggesting this three-part structure for analysis, we recognize that a dispositive assessment of trustworthiness will *never* be feasible. Even so, a combined assessment of technology, corporate governance, and policy and law can provide insights into questions of trustworthiness. Even when decisively clear answers may not be possible in most cases, evaluating relative trustworthiness is possible.

It is also important to try to identify a framework for establishing trustworthiness, as the question lies beneath a great many ongoing controversies. As the global supply chain for ICT products expands, new producers enter the field, bringing novel and different risks to the security of the products they create. Everyone—from end users to major enterprises—must decide which products, and vendors, to trust.[3] Many make those decisions without a great deal of consideration, but the stakes in those decisions are often significant. And even those who choose to think about them carefully lack a system of thought for addressing the question. The ongoing discussion regarding the use of Chinese information technology products as components of Western systems is one example of a broader and deeper problem: How should we assess the degree of trustworthiness in ICT products?

The answer to that question can only be one of degree. It is the nature of ICT systems that risks of compromise can never be fully eliminated, but these risks can be mitigated to a degree that often depends strongly on the effort devoted to the task.[4] Sometimes, investments aimed at mitigating the risk of compromise may contribute more to the perception of trustworthiness than to increasing the relative trustworthiness of an artifact. Put another way, in some cases, risk assessment can often be a matter of perception rather than evidence. Yet perception is a poor basis for decision-making.

Trust also depends on context. What is a risk to one user could be an opportunity for another. A significant risk to one organization in a given setting can be considered negligible to a different organization operating in another environment. Product differentiation, market fragmentation, and

[3] This report focuses primarily on products (both hardware and software) and the vendors of those products. Though we have not addressed the issue directly, to a large degree much of what we offer here will be of utility in evaluating the trustworthiness of services and service vendors as well.

[4] The risk of compromise is not the only risk an end user may seek to mitigate. Expenditures may, for example, fail to mitigate actual risk of compromise but significantly mitigate other risks, such as the risk of liability or reputational risk. Our discussion of trustworthiness in this report is generally focused on the direct risks that arise from the use of digital products, not these related, but distinct, collateral risks.

context of deployment are intricately tied to these questions of trustworthiness. All of this means that there is no one answer to the question of whether and when trust is justifiable or appropriate.

Perhaps most importantly, we don't have an agreed-upon checklist of steps an organization can take that all significant stakeholders will agree demonstrates its products to be trustworthy—what one might call a functional definition of trustworthiness. Without such a functional definition, an evaluation of trustworthiness cannot be conducted, though the assumption is that trustworthiness is likely to be differentiated by technology, context of use, and user capabilities. Increasing trustworthiness can mean that increasing costs and commonly adopted solutions may add to the comfort of the system owners and customers, but they may not alter the actual security of the system in question. In addition to the need for objective measures that have not yet been developed, technology users and customers have differing perceptions of trade-off calculations based on their risk preferences and their business models.

That, in turn, leads to a fundamental problem: Today, we don't know how to appropriately assess and evaluate the trustworthiness of a digital system or component. We don't know how to construct a concrete description of acceptable features in systems behavior and agreed-upon measures of assurance, either in general or for any specific system. Though we likely have in mind a practical definition of trustworthiness (that a system does what it is supposed to do and does not do what it is not supposed to do), that definition is not concrete for a given system and we do not yet know how to embody a useful definition in policy or law. On a global scale, we have yet to reach consensus on how to establish trustworthiness in a single domain, much less a cross-domain consensus—not at the technical level, not at the level of corporate behavior and expectations, and not at the level of nation-state policy.

How, then, are ICT manufacturers to provide skeptical users and observers with assurance of the trustworthiness of their products (or of their own internal governance structures)? If a producer of ICT goods wishes to differentiate its goods by providing convincing assurances, how can it do so? Where can consumers or customers turn to if they wish to evaluate the trustworthiness of the items they are selecting, especially when products contain multiple elements originating from different manufacturers and developers? And what are the key characteristics of a framework that permits us to answer these questions?

To date, the technology, legal, and policy communities have each produced their own answers to the question of establishing trust, limited to their respective domains. But it should not be impossible to develop a broader set of principles, some of them based on empirically testable postulates, that can guide an assessment of trustworthiness. Even without the prospect of precisely assessable levels of trustworthiness, we believe that a framework for assessments can be made with a relatively high degree of confidence.

The value of a framework based on agreed-upon principles should be evident. Using these principles—as well as acceptable evidence—as a guideline, ICT manufacturers and users, including organizations and consumers, could analyze comparative risks and make reasoned risk-benefit and resource-allocation decisions.

To that end, the Lawfare Institute convened a working group to articulate and justify such a set of trustworthiness principles.[5] This report reflects our efforts. Our review proceeds in several parts:

- First, we outline the existing state-of-play in the current field of trustworthiness evaluation.
- Second, we identify a more formal set of definitions for what it means for an ICT product (whether hardware or software) to be trustworthy and, relatedly, what those definitions mean in terms of a trustworthy supply chain.
- Third, we attempt to apply those definitions along various axes of possible trust: trust in the technical components, trust in the manufacturer, and trust in the political and legal system within which a component is developed or deployed.
- Fourth, we integrate the disaggregated axes of trustworthiness into a more nuanced and holistic assessment of the concept.

In candor, our efforts with respect to this final effort were only partially successful. We believe that we have developed a more nuanced view of trustworthiness than others have previously articulated, but we recognize that a persistent and perhaps ineradicable challenge remains in applying our framework to specific instances: The various categories are stubbornly incommensurate, and consumers' risk preferences are heterogeneous in character. We believe, however, that, as described more fully below, useful cross-category comparisons can be made. To test our hypothesis, we include along the way a case study that, we hope, illuminates the utility of what we have attempted.

In the end, we have articulated a set of principles that, in our view, form a suitable structured framework for analysis to guide the assessment of ICT trustworthiness. Our framework distinguishes between trustworthiness criteria that are analytic in nature (that is, capable of definitive measurement) and those that are axiomatic (that is, procedures or rules that are indicative, but not dispositive, of trustworthiness). At the top level, we summarize this framework as follows:

- *Maximize transparency*—The single greatest engine of trust is transparency. Visibility into the production and the operation of any system (whether technical or not) increases the ability to test the system and verify its operation.

[5] The biographies of the working group members are included in Appendix A.

- *Ensure accountability*—Transparency can help users make better decisions, but from a trustworthiness standpoint, it must be paired with accountability. Those who put ICT artifacts into the stream of commerce must be accountable for their trustworthiness. Mechanisms for accountability may vary depending on the artifact, regulatory environment, and context of use, but their absence is a sign of systemic untrustworthiness.
- *Allow for independence of evaluation*—Often the transparency and accountability of a system are self-assessed by the manufacturer and/or designer of the digital artifact, both of whom have a conflict of interest in engaging in this analysis on their own. In general, independent (often outside) evaluation is superior.
- *Prefer provable analytic means of trust verification over axiomatic, nonverifiable means*—As a general matter, trustworthiness assessments that are analytic in nature are superior to those that are axiomatic. We recognize that analytic means of establishing trust are significantly more costly and difficult to implement, but where feasible, they are preferable.

SETTING A BASELINE: THE EXISTING STATE-OF-PLAY

We did not begin this effort writing on a blank slate. To paraphrase Isaac Newton, our work stands on the shoulders of others who have gone before us. Indeed, the seminal paper by Lee M. Molho, "Hardware Aspects of Secure Computing,"[6] which first looked at the hardware aspects of this problem, is now fifty years old, and the early information security work is likewise more than forty years old.[7] We therefore began our examination of the question by compiling an annotated bibliography that gives a baseline of existing works on the evaluation of trustworthiness.[8] We sought both to summarize the existing field and to characterize it, as a jumping-off point for our efforts.

We did not set out fully to define the field, nor did we try to plumb the depths of technical intricacy. The goal, rather, was to provide examples of important and relevant papers and reports, to assist readers in finding additional information on the subject.

---

[6] L. Molho, "Hardware Aspects of Secure Computing," in *International Workshop on Managing Requirements Knowledge*, Atlantic City, NJ, 1970, pp. 135, https://www.computer.org/csdl/pds/api/csdl/proceedings/download-article/12OmNweBUI1/pdf.

[7] See, for example, J. H. Saltzer and M. D. Schroeder, "The Protection of Information in Computer Systems," in *Proceedings of the IEEE*, 63, no. 9 (1975), 1278–1308, doi:10.1109/PROC.1975.9939. https://www.cs.virginia.edu/~evans/cs551/saltzer/.

[8] The full annotated bibliography is included in Appendix B.

The results revealed that the question of trustworthiness is stove-piped into subfields. We found ample work in subfields of technical trustworthiness, corporate compliance, and the legal and political domain. Much of the work in each of these subfields makes meaningful and substantial contributions to advancing our understanding of trustworthiness, but little that we identified crossed the subfield boundaries. It is our belief that a coordinated consideration of all of these subfields together will amount to more than a study of any of them individually can provide. Hence, one task of the working group was an effort to recharacterize the field in a fashion that allows for cross-connections between existing subfields. We began that effort by attempting to define the bases for trustworthiness in ICT products more rigorously.

DEFINING THE BASES FOR TRUST IN AN ICT DEVICE OR PRODUCT[9]

Those who would subvert the trustworthiness of a digital artifact have many plausible goals. They often involve a desire to intercept, degrade, disrupt, or destroy data for purposes that advance their own interests. They may also include significantly more aggressive efforts to modify the functionality of physical components in the world or even to destroy them.

To achieve these objectives, a malicious actor has any number of opportunities to intervene. As a Defense Science Board report put it, an attacker may seek to exploit vulnerabilities introduced at any point during the development, manufacture, or delivery of a component.[10] The digital systems on which individuals and nations increasingly depend are large and complex, comprising many devices and components; today, these systems are likely to be rife with vulnerabilities. Many of those vulnerabilities will be known, some unpatched, and others easily discovered by analysis. In short, digital systems are easy to compromise.

Whether users trust a digital system will be based on the beliefs they hold about that system's behaviors. But beliefs are not necessarily truths. Holding unsound or incomplete beliefs could lead to trust in a system whose behavior does not satisfy expectations. Consequently, it is crucial to understand the soundness and completeness of any beliefs that might be derived to justify trust in a digital system.

---

[9] This section of the report is derived from work originally published as Fred B. Schneider and Justin Sherman, "Bases for Trust in a Supply Chain," *Lawfare*, February 1, 2021, https://www.lawfareblog.com/bases-trust-supply-chain.

[10] Defense Science Board, *Resilient Military Systems and the Advanced Cyber Threat* (Washington, DC: U.S. Department of Defense, January 2013), https://dsb.cto.mil/reports/2010s/ResilientMilitarySystemsCyberThreat.pdf.

To use a digital device (or artifact) with justified confidence, individuals and nations must have some basis to trust that the device will do what is expected and that it will not do anything unexpected, despite attacks from adversaries in the environment in which that artifact is deployed. A digital artifact might range from a single electronic component to a networked information system; its environment could include humans (intended users and others with malevolent intent), utilities (such as electrical power and communications networks), computer hardware (from personal devices to desktop computers to cloud infrastructures), and software (operating systems, databases, and applications). Any and all of the various elements identified above making up the broader environment might be involved in an attack that both disrupts the operation of a device and erodes the trust one places in that device.[11]

We have identified three different broad categories of such trust: axiomatic trust, analytic trust, and synthesized trust.

Axiomatic Basis for Trust

In mathematics, an axiom is a statement that is accepted at face value. In this spirit, we define an *axiomatic basis* for trust in a digital system to be any rationale where the beliefs about a system's behavior are accepted without evidence derived from analysis of the system itself. Needless to say, by ignoring details about the system's implementation, an axiomatic basis for trust is necessarily a weaker source of assertions about a system's behaviors than one based on data from that system.

A common example of an axiomatic basis for trust presumes to predict attributes of a system's behaviors from attributes of that system's development: the country in which the company that created the system is located, for example, or the reputation of the company that sold the system, the ISO (International Organization for Standardization) certifications that company holds, or the certifications or degrees held by the people the company employs. The problem is that attributes of a system's developers are not sufficient to determinatively conclude anything about a system's behaviors. Rather, the attributes of the processes used to design and implement a system are used to support inferred conclusions about the system's behaviors. Consequently, trust that is based on attributes of development requires belief in the absence of vulnerabilities without evidence of that absence derived from an examination of the system itself.

Deterrence through accountability—whether that accountability is regulatory or reputational—is another example of an axiomatic basis for trust. Once it is established that developers who are being held accountable for their actions have built the system, users may choose to accept (again without

[11] Appendix C offers a more formal and technically in-depth definition of how to define a trusted component based on trust in the underlying supply chain of other components.

proof derived from the design or implementation of the system) that these accountabilities lead to a system that exhibits the right security behaviors. One need only think, for example, of the reputational harm experienced by SolarWinds or the regulatory response to the Colonial Pipeline ransomware attack to understand the nature of this sort of accountability. That said, some observers argue the opposite about these examples—that the reputational harm and regulatory responses were not sufficient to provide adequate security.

Many of the nontechnical measures being advocated to improve the trustworthiness of digital systems can be seen to be schemes that facilitate deterrence through accountability and, thus, are axiomatic bases for trust. An extensive list of such measures, for example, forms the body of a recent Center for Strategic & International Studies (CSIS) report concerned with trust in 5G telecommunications networks.[12] That CSIS list includes various ways to foster visibility into a company's operation, either to create accountability or to detect conditions under which incentives exist for introducing vulnerabilities into products. The CSIS list also identifies institutions or processes that can provide the incentives and disincentives needed for deterrence through accountability.

A final example of an axiomatic basis for trust is seen in defenses that are implemented through source selection during acquisition in a supply chain. For example, when it is too costly for an adversary to infiltrate and corrupt all the suppliers for a particular system, one possible defense against an attack is to prevent adversaries from learning who the supplier is for a specific system by making the purchase in secret or by making purchases from many suppliers and then randomly selecting only one supplier's systems for actual deployments. This is a rationale for trust that ignores how the system works, so, by definition, it is an axiomatic basis for trust. To be sure, special-purpose digital systems are unlikely to have many suppliers, though commodity software might. For this basis for trust to be effective, therefore, users must believe that purchases of the system can be made in secret or that a randomly chosen system is unlikely to have come from a supplier that the adversary has infiltrated.

Analytic Basis for Trust

In contrast to axiomatic bases for trust, trust might be acquired by studying a system's possible behaviors: (a) by running the system on selected sequences of inputs or (b) by deducing properties from the construction of the system itself. Testing is an example of the first approach. It is an

[12] CSIS Working Group on Trust and Security in 5G Networks, *Criteria for Security and Trust in Telecommunications Networks and Services* (Washington, DC: Center for Strategic & International Studies, May 2020), https://www.csis.org/analysis/criteria-security-and-trust-telecommunications-networks-and-services.

inductive form of analysis, wherein experiments inform beliefs about unobserved behaviors; inputs submitted and outputs they produce constitute evidence for broader beliefs about the system's performance. With the second approach, the analysis is deductive; a logical proof that relates an implementation to some formal specification constitutes evidence for beliefs about the system's behavior. Verification, model checking, and other formal methods are instances of this approach. Testing and formal methods exemplify an *analytic basis* for trust, which uses the system itself to derive beliefs about its possible and impossible behaviors.

The soundness and completeness of beliefs derived by using testing depends on what tests are performed. For nontrivial systems, it is infeasible to check all inputs, much less to check all the sequences of inputs necessary for a complete understanding of possible system behaviors. Even systems that include interfaces for direct access to the system's internal state (thus reducing the number of test cases that need to be observed) will typically require that a prohibitively large set of inputs be observed. With exhaustive testing likely to be infeasible, beliefs derived from testing could be inaccurate. Specific misbehaviors might be observed, but testing cannot be used to infer that other misbehaviors are *not* possible. In addition, whether vulnerabilities are revealed by testing will depend on what interfaces can be monitored when a test is run. For example, side channel attacks—data leaks involving unanticipated ways to monitor system activity—are unlikely to be detected by ordinary testing approaches.

With formal methods, a property of the system that has been verified becomes a belief. Today's formal methods impose practical limitations on the size of systems that can be analyzed and the classes of properties that can be checked. So certain kinds of vulnerabilities—such as errors in the specifications for the system—will be missed when formal methods are being used to justify trust in a system. In addition, unanticipated vulnerabilities could be missed by failing to verify the right property. Moreover, use of a formal method is impossible without a description of the internals of the system to be analyzed. Such a description might not always be available if it is proprietary or if the system builder integrated existing subsystems according to specifications of its interfaces rather than descriptions of the system's internal operation.

Access to a system for testing and access to a system description are sometimes afforded to evaluators through special arrangements with vendors. Microsoft, for example, established its Government Security Program (GSP) in 2003 to provide national governments with controlled access to Windows source code and other technical information from within special facilities located throughout the world.[13] In late 2010, Huawei Technologies opened its Cyber Security Evaluation Centre in Banbury, England, to provide the U.K. government and U.K.

[13] Microsoft Corporation, "Program Overview," Government Security Program (GSP), https://docs.microsoft.com/en-us/security/gsp/programoverview.

telecommunications providers with an opportunity to analyze the security of Huawei products. This kind of privileged access to systems and information, however, leaves open the possibility that beliefs derived using it are not valid for all instances of the system. An inspected instance of a system might not have the same vulnerabilities as the system instances that are deployed in the field—for example, the code running in the field may not correspond to the code that was inspected. The accuracy of such an analysis is also limited by the capabilities of the available analysis methods, and evaluators may be left wondering whether they forgot to check something.

Synthesized Basis for Trust

For the sake of completeness, we identify one final basis for trust that depends on technical means but that is independent of testing or formal methods. A *synthesized basis* for trust enables trust to be justified in a whole device or system based on trust in its components and trust in whatever glue enables their interaction.

A synthesized basis for trust often will relocate what must be trusted to a smaller *trusted computing base*. For instance, users might seek to trust that multiple large and complex virtual machines together deliver some service. Given the complexity and size of this system, formal analysis is nearly impossible without unrealistic resource commitment.

But what if the user could establish trust in a small piece of the system that would prevent compromise of any one of the virtual machines from affecting the others? Establishing trust that a *hypervisor* (a potentially small piece of software) isolates the memory and other resources of each virtual machine makes it easier to establish trust in each individual virtual machine, since now no virtual machine can undertake actions that affect any other. The hypervisor can be seen as allowing a synthesized basis for trust by restricting possible behaviors of some target components.

A less technical example of a synthesized basis for trust is *split fabrication* for semiconductor chips, an approach proposed in research literature to mitigate provenance from potentially untrusted sources.[14] A semiconductor is a hugely complicated component, one that is foundational to the functionality and trustworthiness of all the components in which it is installed. Compromise it, and an adversary has potentially compromised all the downstream products of which it is a part. Here, however, different steps in producing a semiconductor chip can be performed by different manufacturers, limiting the potential impact of corruption of a single step in the fabrication process.

[14] Meenatchi Jagasivamani, Peter Gadfort, Michel Sika, Michael Bajura, and Michael Fritze, "Split-Fabrication Obfuscation: Metrics and Techniques," 2014 IEEE International Symposium on Hardware-Oriented Security and Trust (HOST), May 6–7, 2014, https://ieeexplore.ieee.org/abstract/document/6855560.

When the different manufacturers are independent, split fabrication increases the cost of attacks by requiring attacker involvement in many different organizations.

Replicated distributed systems—in which copies of the same data, and sometimes copies of the same software, are run simultaneously—implement a software analog to this kind of defense. The output from a replicated system is defined to be the output from a majority of its replicas. Provided that replicas are independent and, therefore, typically have different vulnerabilities, an attacker must do the work to compromise multiple replicas in order to corrupt a replicated system's output.[15] This approach has been used widely to enhance integrity in fault tolerance systems to control aircraft and trains, thereby ensuring continued system function even if some components fail.

The Challenge of Trust in an Integrated System

If it were possible to test that a component delivered the correct output for all possible inputs under all possible operating circumstances and with all possible histories, then it would not matter in the least how that component had been fabricated, who had fabricated it, or who had enjoyed access to it prior to the test. The component could be regarded as perfectly trustworthy.[16]

But it is not possible to do this—at least not in any practical environment. The reason, in essence, is that such a test requires using empirical evidence (the results of various tests) to prove a negative. Finding one improper output for a given input demonstrates a flaw. But as noted above, the number of possible inputs is so large in practice for most components of nontrivial systems that such testing is impossible. Indeed, even the *same* input might lead to a different output if, for example, component operation is sensitive to its environment—for example, operating under conditions of excessive heat or radiation, or operating with timing-dependent variation in the outputs it generates.

---

[15] Tom Roeder and Fred B. Schneider, "Proactive Obfuscation," *ACM Transactions on Computer Systems*, 28 (July 2010), 1–54, https://www.cs.cornell.edu/fbs/publications/ProactiveObfuscTOCS.pdf. Though this is generally true, independently written programs sometimes contain similar or identical flaws. See John Knight and Nancy Levenson, "An Experimental Evaluation of the Assumption of Independence in Multi-Version Programming," http://sunnyday.mit.edu/papers/nver-tse.pdf.

[16] The only other theoretical way in which one could have perfect trust in a component is if one were to have constructed the component from scratch—that is, from basic unaltered raw materials without reliance on the assistance or input of any third party—and from a design without any flaws. Needless to say, while conceptually feasible, the effort to construct modern-day hardware or software components by hand from a flawless design is, for all practical purposes, impossible. Anyone who has seen Thomas Thwaites's efforts to build a toaster from scratch will understand that contemporary reality precludes technological independence (see https://www.ted.com/talks/thomas_thwaites_how_i_built_a_toaster_from_scratch).

Consequently, no amount of testing suffices to fully establish trustworthiness in practical circumstances, and we must turn to other means to establish a reasonable degree of trustworthiness. That trustworthiness may be quite high, but it most assuredly will not be perfect. What counts as high enough is a matter of judgment or policy. It is not ultimately a technical question.

Another problem is that even perfect trustworthiness of components does not guarantee perfect trustworthiness of systems constructed from those components. Even if perfectly reliable components existed, these components could still be integrated in ways that result in systems that are not reliable. Consider a building constructed in Los Angeles using perfectly trustworthy building materials yet designed with a seismically stable region in mind; such a building might be made wholly of reliable components, but it would not be trustworthy in earthquake-prone Los Angeles. The problem of integrating perfectly trustworthy components into a trustworthy system is a simpler one than that of integrating not-so-trustworthy components into a trustworthy system, and by focusing on the former, a number of insights emerge more clearly.

Even once we establish an acceptable level of trustworthiness for individual components, ensuring adequate trustworthiness of a system is a matter of system integration—what we call the synthesized basis for trust. System integration involves how components are connected to and interact with each other.

Evaluating the trustworthiness of a system requires knowledge of how that system is intended to be used—its concept of operations. Embedded within a system's concept of operations are assumptions about the threats that the system will face—the threat model. All threat models capture certain threats and ignore or deemphasize other threats not relevant for the context. It thus stands to reason that interested parties cannot be trusted to perform system evaluation alone; such parties will either naturally ignore one or more threats or amplify the importance of other risks.

Another source of untrustworthiness is the result of human behavior: how system users can behave in ways that violate the concept of operations for the system. Because such violations are by definition outside the concept of operations, system integration cannot account for such behavior. But considering the system as one element of a larger socio-technical entity can and should account for unexpected user behavior. Socio-technical integration considers how people interact with the system in the context of the organization in which the system is embedded and is the foundation for the appropriate organizational constructs that influence user behavior. Such constructs may include incentives and penalties for doing the right and wrong things or record-keeping to facilitate accountability and deterrence. Here, too, it stands to reason that an adversary should not have an opportunity to touch the socio-technical integration of a system that must be trustworthy.

System integration for security requires that human operators find it easy to do the right thing. But difficulties often arise when the "right" and "wrong" things refer to actions that enhance and detract from security. Butler Lampson, a well-known computer software architect, often quotes General Benjamin W. Chidlaw, a former NORAD commander, who said, "If you want security, you must be prepared for inconvenience."[17] Considering that "easy" and "difficult" map quite well onto "convenient" and "inconvenient," and "right" and "wrong" map onto "secure" and "insecure," the dilemma becomes clearer: System integration for security is faced with an internally contradictory set of requirements, as a system cannot be maximally secure and maximally convenient at the same time.

As an example, system integrators generally consult with users to better understand the users' operational needs. This relationship is intended to be one in which the system integrator acts as a facilitator for users' needs to be met, that is, the system integrator helps users to articulate their needs so that the system can be integrated in a way that is aligned with those needs. But if the system integrator, unbeknownst to the users, is working for his or her own ends rather than trying to help the users, he or she may subtly guide the users to make decisions regarding system integration that do not serve the users' security interests. For example, the system integrator may go along with a user request for a functionality that the integrator knows will lead to poor security, whereas a trustworthy system integrator would point out the security problem inherent in the request.

Lastly, one can sometimes go outside the system itself to improve trustworthiness. It is often possible to create an external monitor that does nothing but check system outputs for specific catastrophic outputs—outputs that, were they to occur, would be regarded as a catastrophic failure of system security. Should the monitor detect such an output, the monitor intervenes and suppresses the output, perhaps with a note to users announcing the action.

Best practice would call for an external monitor to be developed and operated independently of the entity being monitored. Any external monitor should be small and simple (and therefore more likely to be trustworthy). But for this approach to make sense, the number of catastrophic outputs to be checked must remain small—a fact that demands that customers make difficult judgments about what the catastrophic outputs would be and how they would be recognized. External monitors cannot transform insecure systems into secure ones. But they can monitor for and mitigate certain kinds of security problems and transform certain types of misbehavior into benign behavior.

---

[17] Butler Lampson, "Perspectives on Security," Microsoft Research, Symposium on Operating Systems Principles, October 4, 2015, https://www.sigops.org/s/conferences/sosp/2015/history/02-lampson-slides.pdf.

ANALYTIC TRUST: TRUST IN TECHNICAL PERFORMANCE CRITERIA

Each of the various bases for trust in an ICT component can be used as a lens through which to evaluate the strengths and limitations of various trust proposals. In this section, we examine the technical lens and outline some of these technical considerations and how they play into a trustworthiness evaluation.

Technical Context

First, consider the context in which a technology is deployed, which is to say the factors that are related to the technology itself but are, in some ways, exogenous to the technology.

Technological sophistication: The harder it is to understand a technology's external functionality (what it advertises it can do) and its internal functionality (what goes on inside the component), the more important it becomes to take into account its provenance. Conversely, the easier it is to understand a technology's external and internal functionality, the more there is a basis for a trust judgment independent of the technology's origin and manufacture.

Application: The more sensitive a technology's uses, including the other systems with which it is connected, the greater the potential harm from a compromise of that technology.

Vendor contact after deployment: The more a vendor can access software or hardware fabrication in the development process, and the more software or hardware updates and patching take place after the technology is deployed, the more important it becomes to factor in the attributes and circumstances of the vendor and the vendor's relationship with the country in which it produces the component. Likewise, the degree to which a vendor is (or is not) transparent about its activities may affect an assessment of its products. This consideration must include both legal and illegal means that the vendor can employ.

Technical Evaluation

Given these considerations, several measures can be deployed to technically evaluate and increase trust in systems.

Testing: Testing a system's inputs and outputs is a way to better understand its operations in making evaluations of trust. But testing should also evaluate a system's internal processes and constituent components—looking within the black box, so to speak—at requisite levels. The depth at which such an inquiry can successfully occur remains unclear. For example, the Huawei Cyber Security Evaluation Centre (HCSEC) in the United Kingdom was designed to look *within* Huawei

components. But the HCSEC's practical success has yet to be fully demonstrated. An ideal testing system is one that enables noninvasive testing of all units, rather than testing of designs, and allows testing of instances of systems in a reliable and standard way. In other words, it allows the tester to determine that the component is working as expected rather than needing to know all the details of its operation.

One caveat related to this measure is what we call the Volkswagen scenario: a corporation with knowledge of the testing system that was able to "game" the results precisely because of the reliability and standardization of the testing regime.[18] To some degree, this scenario is possible in any standardized testing model, and it may be viewed as a necessary "cost" of testing. The cost may be ameliorated by some randomization and regular auditing. In the end, however, although the Volkswagen experience is a challenge to the concept of testing structures, it can also be viewed as a partial validation of them. The testing structures played a role in exposing the scope and scale of the improper behavior.

Isolation: Within a single system, isolating certain subcomponents, such as with hardware security modules in processors, can enable the user to establish a smaller trust zone to simplify the trust calculus for the broader system.

Secure attestation: Secure attestation is a mechanism for software or hardware to verify its identity through cryptographic signing. This mechanism should be paired with isolation measures, because isolation itself involves a supply chain problem—for example, who is trusted to build the subcomponent for the smaller trusted computing base—that can be addressed by the manufacturer implementing secure attestation mechanisms for system subcomponents.

Distribution: Across an entire digital supply chain, distributing trust across multiple entities is a way to simplify the trust decision for a single system. If more than one entity, or even all of them, would have to be compromised for the entire end system to be fully compromised, that design approach raises the costs of compromising the system and, therefore, increases our reasons to trust it. To use an inexact analogy, assume that an enterprise has a choice between building three separate bridges across a river at cost or making the same investment to construct one super-bridge across the same river. If we can make assumptions about the independence of the failure rates of the three bridges, we don't need mathematical proofs or anything else to know that the failure of one system will not result in failure in the others.

Innovative approaches to patching: Patching is often thought of as necessarily all-or-nothing. But a more sophisticated technical and human process for certifying patches, deciding what kinds of

[18] For a short summary, see "Volkswagen Says 800,000 Cars May Have False CO2 Levels," BBC, November 4, 2015, https://www.bbc.com/news/business-34712435.

patches are permitted on a system, and handling patches in a more innovative fashion can increase trust in the user's ability to check the impact of a vendor's contact with the system after deployment. For instance, a more innovative approach to patching might consider what part of the system the patches will impact and what rules they will change. To be sure, there is some complexity to this approach. The current model, for example, for hardware patching is indirect. We use multiparty coordinated vulnerability disclosure to prevent malicious actors from leveraging an unpatched vulnerability. But a more innovative approach is, in practical terms, achievable.

One further note on this point: Patching policy is not exclusively technical. For example, a country that is no longer allied with another country whose firm produced a given component may find itself at greater risk of targeted compromise of a particular system. Likewise, there have been examples of patches that render some useful functionality or critical application inoperative. These counterexamples suggest that a "patch-always" policy has some edge-case limits.

Costs

We acknowledge that costs sometimes accompany efforts to provide assurance of technical trustworthiness. These cost considerations are relevant to any potential technical measures for assuring trustworthiness.

Technical costs: Secure multiparty computation, for example, would impose high computation and communications costs on both vendors and users if widely deployed on consumer systems. Understanding the technical costs of a particular solution, and whether those costs are worth it, is essential.

Competitiveness: The impact of government regulations on industry competitiveness is not the only competitiveness consideration in supply chain security policy. The introduction of security controls by a particular vendor could increase the vendor's control of a market chokepoint and thus be a mechanism to undermine fair competition. In the alternative, it can serve as a market differentiator, for good or ill, in competition. Microsoft's years-ago plan for a trusted computing base is an example; the company's proposed security controls got external pushback because Microsoft could also theoretically use them to enforce all kinds of other policies, including copyright rules.[19]

Transparency: Ensuring trust in technology is about not just transparency, but transparency to *whom*. Transparency is often highly beneficial. However, there are cases where transparency to particular vendors or within a particular subset of the industry bolsters security in ways that would be undermined by sharing that information publicly (and thus, for example, with potential

[19] See, for example, Bruce Schneier, "Palladium and the TCPA," August 15, 2002, https://www.schneier.com/crypto-gram/archives/2002/0815.html.

adversaries). We would not want to be seen as endorsing a "security by obscurity" paradigm, but obscurity may be beneficial in some contexts. International standards for vulnerability disclosure already provide some of the necessary transparency mechanisms to the necessary parties and may be sufficient.

AXIOMATIC TRUST: TRUST IN CORPORATE GOVERNANCE

It would be nice to hope and expect that some forms of analytic trust and technical controls could be used to automatically, reliably, and completely verify the trustworthiness of a digital component or system, but that is not likely to be possible in most cases. This form of technical control is essentially "turtles all the way down." In other words, even if we can have confidence in a tamper-proof technical control and in a technical means of verifying that the technical control has been implemented properly, we then need a means of verifying that our verification methodology is technically appropriate and that it, in turn, has been implemented correctly. This requires yet another technical iteration, it seems. At some point, the chain must end.

Put simply, even the best engineered process in the world cannot inspire trust if we do not in some way trust the producers of the system itself. Thus, an analytic basis for trust cannot exist in a vacuum. It must, at some point, connect to more axiomatic aspects of trust.

Indeed, at some point, establishing a system's trustworthiness will rely on the confidence we have in how technical controls have been implemented by humans. Thus, the question will ultimately be how confident we are in the governance of the enterprise that is implementing the controls. If we trust the enterprise, we are more apt to trust the hardware or software it produces and the analytic trust it builds. Conversely, the more we mistrust the enterprise, the more we should mistrust its products, even when there is an apparently strong basis for analytic trust.

This type of trust is necessary but not sufficient to have a trustworthy system. A system may be produced by a very reliable organization with fine people, but it still may have fatal security flaws (as the 2021 Microsoft Exchange Server vulnerability suggests).[20] These problems do not arise because of malicious action by the organization, but they can be just as harmful (if not more so given the blow to trust that such failures engender).

In the context of a corporate enterprise, we are familiar with the idea of establishing process controls for the implementation of technical activities; auditing those processes; assessing compliance; doing quality assurance; and, in the end, enforcing penalties of one form or another for

[20] For a description of the Hafnium attack, see "Hafnium Targeting Exchange Servers With 0-Day Exploits," Microsoft Security, March 2, 2021, https://www.microsoft.com/security/blog/2021/03/02/hafnium-targeting-exchange-servers/.

noncompliance. This is the paradigm of the administrative state and also of many self-governing private certification authorities. It has been used for the past hundred years for everything ranging from environmental mandates to health care privacy requirements to voluntary accident reporting systems.

There is reason to think that this structure could be helpful for the implementation of trustworthy information technology systems (though we note a few concerns below), and there is every reason to expect that a properly constructed corporate governance and compliance system for trusted hardware and software will mirror, to a large degree, existing structures.

Perhaps that conclusion is unexceptional and obvious. But the implications of the observation suggest that a number of significant, yet-to-be-answered issues remain regarding how corporate governance for trusted hardware and software would function.

To begin with—and to state the most obvious point—no comprehensive structure of this sort currently exists. Although we do have experience with any number of corporate-level governance structures that address cyber issues (such as the National Institute of Standards and Technology Cybersecurity Framework, or any number of ISO standards), most observers would agree that they do not form a consistent framework.

This is unsurprising. Regulatory and voluntary compliance structures are commonplace in most national and international systems, but most structures with which we are familiar have grown up organically and as the product of years of development, refinement, and revision. To cite just one simple example, the architecture of domestic American environmental regulation has grown from nonexistent to a mature structure, but that development has taken more than fifty years. And the structure of international environmental governance is still relatively incomplete in many regards, despite sustained attention from the world community over an extended period of time. In short, creating a governance structure for a complex and heterogeneous field from whole cloth is no easy task and will require significant effort.

Today, there is no generally agreed-upon, transparent, auditable, and scalable structure for assurance of hardware and software implementation. There are a number of nascent standards, created by various standard-setting organizations, but there is no single framework—akin to, say, generally accepted accounting principles (GAAP)—that would allow one to define the requirements of acceptable organizational governance and then audit for compliance with those requirements.

That does not mean that creating such a set of requirements is impossible. On the contrary, it appears to be achievable. But preliminary questions need to be answered first. For example, should the standard be voluntary, or the product of governmental rule-making? Then the standards themselves need to be articulated.

As we've already indicated, the question of corporate accountability and governance is not new. Any number of frameworks exist, some quite general in nature and others more specific to the operation of ICTs of the sort we are considering.[21] As a general matter, these programs are designed around three types of questions about a corporate effort to demonstrate trustworthiness:[22]

- Is the governance well designed?
- Is it being applied in good faith?
- Are the results satisfactory?

Each of these questions requires the articulation of principles that can guide an evaluation of a corporate trustworthiness or compliance program. Here are a few of the considerations that might go into such an assessment:

Design: Obviously, one size does not fit all corporations. But many aspects of a well-designed corporate compliance program to provide assurance of trustworthiness (as with most other compliance programs) apply across enterprises. For example:

- Is the program grounded in a solid understanding of the business model of the enterprise, and is it grounded in a realistic risk assessment of enterprise vulnerability?
- Is it grounded in existing industry practice and consistent with domestic law?
- Is it comprehensive?
- Is it based on international standards?
- Is the program new, or is it a long-standing one with settled practices and procedures?
- Does it provide adequately for communications and training?
- Does it cover the entire enterprise, and (consistent with our view of the supply chain) does it impact relevant third-party vendors?

Operations: The most well-designed system in the world is of little value if it is not implemented in good faith and with an appropriate level of effort. Here we might ask questions of the following nature:

- Are the compliance efforts fully integrated into operational activity, or do they have the structure of an add-on component?

[21] The Information Accountability Foundation website, https://informationaccountability.org/.
[22] "Evaluation of Corporate Compliance Programs," U.S. Department of Justice, Criminal Division, June 2020, https://www.justice.gov/criminal-fraud/page/file/937501/download.

- Are the compliance efforts auditable?
- Do they incorporate well-known best practices for compliance programs, including
  - Autonomy of the investigation?
  - Confidentiality for the reporting and investigation of problematic events or policies?
- Does the program have adequate resources?
- Is there evidence of C-suite commitment to the process?
- Is the operational function subject to the oversight and audit of the board of directors or an equivalent outside review function?
- Do adverse consequences flow from error? At a corporate level, does the enterprise respond to and mitigate reported risks? At an individual level, are there appropriate positive and negative incentives for behavior?
- Can the process work on a time and cost scale that does not render its output commercially meaningless?

Effectiveness: It is worth asking whether or not the structures are in place to assess whether or not an assurance program actually works:

- Does the enterprise engage in continuous improvement in response to review?
- Is the operation of the assurance program subject to both internal and external audit?
- What is the enterprise's historical record of practice? Has it previously produced untrustworthy products? For what reasons, and to what degree?

Besides identifying the standards to be adopted, any corporate governance system will need to address at least two significant subsidiary process questions: defining a compliance mechanism and incentivizing compliance with its requirements.

The first of these may be especially problematic in the hardware- and software-assurance space. Any governance system will require a defined compliance mechanism. Compliance mechanisms operate on a spectrum from self-regulatory systems to government mandates with significant compliance verification oversight and obligations. In the middle of these two extremes are many options for the use of standards and verification mechanisms that mix governmental, private-sector, and nongovernmental organization roles. Where a particular system falls on the spectrum should be designed to maximize flexibility, minimize transaction costs, allow for necessary business confidentiality, and demand the appropriate level of accountability.

The distributed, technically diverse, and dynamic nature of the information technology domain creates challenges for mandatory and harmonized governance structures, as they may be slow to adapt to changed technology or business needs. Governance systems must be defined to be sufficiently flexible so as not to chill innovation and to adapt to the evolving technology environment. An example of a system that followed this goal was the voluntary approach of the National Institute of Standards and Technology Cybersecurity Framework (CSF). The CSF adopts a risk management vocabulary designed to be flexible to different industry sectors and changes in technology. Although the CSF is voluntary, there are opportunities for organizations to impose adoption of the CSF on other parties, particularly through procurement and vendor contracts. Similar governance models could be pursued for trustworthiness that preserve flexibility while creating a framework that can demand accountability.

There are other government mechanisms to require accountability of corporate governance. One example of such oversight would be U.S. securities law obligations for disclosure of financial reports with certification by a corporate executive. Another would be the ability for the U.S. Federal Trade Commission (FTC) to bring unfair or deceptive trade practices actions under Section 5 of the FTC Act to determine whether a company is acting in concert with representations it has made about its corporate governance or the trustworthiness of its products and services. A third example in U.S. law would be the use of the criminal prohibition against false statements to hold corporate executives accountable for representations they make to the government in official filings.

Beyond the need for the development of a recommended governance structure, any corporate governance structure will need to incentivize compliance with its requirements. We can imagine that such incentives should, properly, include both positive inducements and negative consequences.

For positive inducements, we believe that, in principle, any system should include a reward for successful implementation. Some of that reward will likely be reputational (with associated economic benefit). In legal systems where it is appropriate, that reward may also include some form of safe harbor against liability for enterprises that are adjudged to be substantially compliant with whatever governance standards are developed.

Conversely, we expect that negative incentives may include significant reputational and economic penalties, ranging from public identification, to exclusion from certain markets, and even civil fines. In principle, the proper balance between positive and negative incentives cannot be determined *ex ante*. Their development and deployment will, necessarily, be based on experience and reason. Along with practical positive and negative incentives, governments must have the political will and the supervision, monitoring, and enforcement resources to effectively hold large technology

companies accountable. Laws on the books generate sufficient trust only if they translate to action in practice.

AXIOMATIC TRUST: TRUST AS A FUNCTION OF NATION-STATE POLICY AND LAW

Even companies seeking earnestly to produce trustworthy products may be constrained in doing so by the laws and regulatory environments in which they operate. Except in rare cases of technical constraints that cannot be avoided, a state's policies or laws may mandate actions that overtly undermine a producer's efforts to provide a system that we would deem untrustworthy (as with recent Chinese regulations that require companies to disclose vulnerabilities to the Chinese government and keep data in China, and some American encryption proposals that would require law enforcement access to stored data and data in transit), or it may discourage corporate governance of the type that would foster trustworthiness (the opposite effect of the phenomenon just described, where government mechanisms encourage more trust in corporate governance). In other words, state law and policy may make systems vulnerable and untrustworthy either by allowing the state (overtly or covertly) to subvert the trustworthiness of the system or by disincentivizing the creation of trustworthy systems and thereby rendering them vulnerable to outside attack.

This form of axiomatic trust—in which the consumer bases a decision on whether to trust a product, in part, on the legal and policy context of its design, creation, and distribution—is, in many ways, the most difficult to define and the most contextual in nature. It depends not only on the legal and policy environment in which a good is created but also on exogenous political risk preferences that are incapable of resolution. In Europe, many observers see the American legal system as insufficiently protective against governmental intrusion, as evidenced by the repeated court battles over cross-border data transfers involving the United States. In the United States, China's legal architecture is thought inadequate. Even as the European Union chides America for its surveillance practices, some Americans think that European government surveillance authorities are opaque and overbroad. And although few might, say, place trust in the restraint of the North Korean government, there may well be situations in which legal and policy constraints are almost irrelevant to considerations of trustworthiness. Consider the question, for example, of whether you would trust a pencil made in North Korea for routine purposes.

Informal Influence

A central concern of supply chain integrity is the extent to which governments may influence the behavior or actions of companies in a key product's supply chain. One form of influence, of course,

is a legal mandate for the company to take action to avoid a penalty, possibly a draconian one such as the CEO being put in jail (we address that issue below). But the concept of trustworthiness in this context is broader than formal law. In our view, influence is best conceptualized along a spectrum of coercion. The following list is ordered by degree in descending order of coerciveness:

- An enterprise is compelled to construct a digital artifact in a manner prescribed by government mandate or directive, creating a product that the consumer would deem to be untrustworthy.
- An enterprise is compelled to act in accordance with government directive without an independent appeals process (for example, to an independent judiciary) that might reverse the government directive. Government can impose any penalty to compel compliance.
- The enterprise is compelled to act in accordance with government directive, subject to or pending an independent appeals process. That is, the enterprise is compelled to act according to government diktat until an independent appeals process reverses the government directive. Power is with government because the government can slow-walk the appeals process.
- The enterprise is compelled to act in accordance with government directive, but only after an independent appeals process upholds the government directive. It has freedom of action until the appeals process is complete.
- An enterprise is persuaded to act without government directive but in accordance with an expressed government preference. In this scenario, the government explicitly makes its preferences known but does not demand compliance. It offers or hints at other possible inducements in a variety of informal and off-the-record venues, such as over friendly cups of tea or at cocktail parties.
    - Positive inducements include various financial incentives, such as promises of government contracts, favorable tax treatment, better access to domestic markets, advocacy of the company's interests in international trade, favorable enforcement of regulations that might target the enterprise, promulgation of favorable regulations, and support for standards desired by the enterprise.
    - Negative inducements include threats of negative regulation, investigation, civil litigation of the sort that is just beginning with data breach torts, harassment, and stringent "by the book" enforcement regulations.
- The enterprise operates without regard for any expressed government preference but takes into account implied, implicit, or unstated government preferences ("it's just understood what to do").

- The enterprise does as it pleases in accordance with its own technical or business judgments.
- The enterprise does things that might please the government to curry favor.

Along this spectrum, almost all of the soft-power factors are inherently unmeasurable. Some observers, for example, in the United States have a view of the problem that emphasizes the Chinese government's influence on major Chinese suppliers, but some Europeans look at the United States and see structurally equivalent soft-power influence wielded by the American government.

Formal Law

Even the harshest European critics of the United States would acknowledge some sense in which the considerations of trustworthiness differ from those they apply to, say, Chinese or Russian products. Our hypothesis, borne out by experience and anecdote but unprovable empirically, is that the distinction arises from differences in the nature of the formal legal structures at issue.

Put another way, though one may have soft-power concerns about a government's degree of influence over a technology vendor, those concerns can sometimes be mitigated by hard-power, formal considerations regarding the law and policy adopted by a nation-state. Those formal structures are more readily assessed and characterized and provide some assistance in assessing the trustworthiness of hardware and software systems.

Recognizing that the effort is, as we say, constrained by unmeasurable informal structures, we nonetheless believe that the following framework for assessing the national laws and policy governing systems of an originating state are relevant to, and valuable in, the effort to assess trustworthiness.

We start by recognizing that, at a minimum, any widely agreed-upon governance structure will come with a national security carve-out of some sort. No government, not even those of the Western democracies, would agree to wholly immunize hardware and software providers from the reach of national security law. Properly drawn, such a national security provision would be sensible and welcome; if improperly drawn or administered, it will be viewed as a loophole subject to abuse by another party through which a proverbial tank may drive. The degree to which there can be transparency to independent overseers of the government agencies that act under these national security legal authorities will be a significant indicator of whether the national security provision has been drafted too broadly. We offer two high-level principles that should guide the consideration of any such authorities and an assessment of whether or not they negatively impact trustworthiness:

- First, the law should leave as little as possible to the government's discretion and judgment. To that end, laws should be narrowly drawn to focus exclusively on national security needs

and to define those needs in a way that excludes economic nationalism or routine law enforcement.

- Second, the scope and application of the provision should be adjudicated by an independent neutral arbiter in the event of dispute between the parties involved.

Those principles can be described more particularly as follows:

Transparency: An effective rule-of-law structure is one that all actors can understand and refer to publicly. It is also one where the government's use of the legal system is apparent. The idea, naturally, is that transparency facilitates accountability by enabling others to assess how law is implemented in a particular country. Consequently, transparency can be defined by a government's public openness about the existence of a legal authority (Are there secret laws?) and the use of that legal authority to exert influence over a hardware or software system (Are there secret legal proceedings?).

To be sure, secrecy is sometimes an essential component of national security, and sometimes transparency is nothing more than voyeurism. But at a comparative level, we can judge degrees of transparency. A government that is publicly open about the existence of legal authorities to influence a technology vendor, but not about when those powers are used, has one degree of transparency—as distinct from and lesser than a government that is publicly open about the legal authority as well as its use. This transparency better enables assessment of the government's likelihood of exerting influence over a hardware or software system.

Existence of a legal authority—Transparency about the existence of a legal authority increases transparency into a government's capabilities to influence a technological system. The existence and disclosure of legal authorities may also provide insight into the government's intent vis-à-vis use of those authorities.

Use of that legal authority—Transparency about the use of a legal authority increases transparency into a government's likelihood of using its authorities to influence a system. It can also better enable assessment of the law's implementation (discussed next) and how its use impacts trust.

These two requirements can work in concert or independently. A legal authority that is rarely used by a government but whose existence and possible use greatly degrades trust in a system has different trust implications than a frequently used legal authority that degrades trust less severely.

Implementation of the law—When a government resorts to legal means to exert influence over the design, implementation, or use of a hardware or software system, the result of that effort matters both in how the result is achieved and in what results occur.

It is possible that through implementation we may increase trust in a hardware or software system, decrease trust, leave our trust in the system unchanged, or sometimes even make it more difficult to establish trust. Factors that might relate to the implementation of the law would include:

> *Source of law.* As an initial matter, the source of the law in question is important. Can the legal authority or policy in question fairly be said to have been the product of a public, open, and inclusive process? Can the laws readily be changed, or are they inviolate? Such processes (not all of which necessarily meet the definition of democracy) are inherently more trustworthy than sources of law that are arbitrary and invariant.
>
> *Scope of law.* Next, one may ask what the scope or domain of the legal system is. Who does it cover, and what does it oblige them to do? When we say that a law "governs" the conduct of a company, a more nuanced assessment can allow us to discern what that means in practice. For example, does the government have regulatory jurisdiction over an affected party directly, or does the regulatory nexus influence only the supply chain used by the party?

Right to contest—Government intervention in a hardware or software system can be either unilateral or subject to dispute by affected parties. A characteristic of a more trustworthy structure is that it is one where the nation-state's actions may be challenged by others.

Contest timing—Similarly, it is relatively more trustworthy for a legal system to permit that challenge to occur before the government acts than for it to occur after the fact. This factor is, of course, of less importance in times of imminent or emergent threat.

Selectivity of enforcement—Consistent with our views on the use of soft-power influence, one critical factor to consider in the implementation of law is the use or abuse of enforcement discretion. It is a truism that not all laws can be enforced in all cases. How a government selects when to enforce its laws (whether, for example, it rewards allies and condemns critics) is a factor in assessing the neutrality of the law's operation.

Independence of the decision-maker—Both the formal and the practical independence of the ultimate decision-maker are notable factors. Systems that foster the independence of adjudicators of a dispute enhance trustworthiness in the results of that dispute resolution mechanism.

Acceptance—The last consideration is whether or not the government and the affected parties accept the dispute resolution as definitive. In untrustworthy systems of

government, adverse decisions that go against government interests are sometimes ignored, or they are undermined, or they are never permitted to occur in the first place. In trustworthy systems, they are accepted and implemented.

Substance of the law: Finally, the exact content of a nation's laws matters a great deal. The authorities specified in a law for the government to exert influence over a hardware or software system allow very different things in different countries. In the end, it is necessary to assess the specific powers granted to specific government entities that allow them to exert influence over a technology. Even in a country without an independent judiciary and in which a government's de facto powers regularly exceed those on paper, understanding what authorities are specified in a law is an important consideration in evaluating trust.

Of equal importance to understanding the law is the need to understand the ambiguities in the law. No legal system is perfect. Much as we need to assess the authorities a legal system gives the government, we also need to understand the gray areas of the law and how ambiguities may affect the enforcement of law by the government in question. Almost by definition, this question is especially difficult to answer, but no assessment of trustworthiness would be complete without at least noting the difficulty.

A CASE STUDY IN TECHNICAL, CORPORATE, AND STATE POLICY LEVELS OF TRUSTWORTHINESS

Much of what we have described so far may seem overly theoretical. To make more vivid our points about trustworthiness as a function of state policy and corporate governance, we sought to apply our thinking to a recent case with sufficient information available publicly to make the analysis possible.[23]

In 2019, the Committee on Foreign Investment in the United States (CFIUS), which reviews foreign investments in sensitive U.S. companies for national security risks, initiated a review of Chinese company Beijing Kunlun Tech's 2016 and 2018 investments in the dating app Grindr.

A look at the CFIUS investigation into Grindr reveals a software trustworthiness assessment based on multiple criteria, not just simple questions of legal ownership. CFIUS's investigation found that Beijing Kunlun Tech's ownership of Grindr posed a national security risk because of the sensitivity of Grindr's data and the Chinese company's control of Grindr's operations and that data. CFIUS

[23] Adapted in part from Kamran Kara-Pabani and Justin Sherman, "How a Norwegian Government Report Shows the Limits of CFIUS Data Reviews," *Lawfare*, May 3, 2021, https://www.lawfareblog.com/how-norwegian-government-report-shows-limits-cfius-data-reviews.

forced Beijing Kunlun Tech to sell the app back to U.S. owners and in doing so addressed the question of company ownership, which was assessed to be the reason for distrusting the app.

In addition, subsequent technical analyses of Grindr have found that the app widely shares data with dozens of third parties, from third-party software development kit (SDK) developers to real-time ad bidding networks, including with one SDK from Chinese technology company Tencent. If the U.S. government's concern was that sensitive Grindr user data might end up in the hands of a Chinese company, focusing on the company's direct ownership of the app did not account for other considerations that posed similar risks, such as the app's technical architecture. In other words, while the government focused on the axiomatic layers of corporate and government policy in its review, later examinations of the analytic layers of system performance exposed additional technical security concerns not covered by the other analysis.

The specific criteria that yielded CFIUS's order remain classified, but the overall decision and its surrounding context can be analyzed through the lens of technical, legal, and corporate governance trust criteria:

- First, CFIUS was seemingly concerned about Grindr's collection of sensitive personal data on U.S. citizens. Information including sexual orientation, HIV status, real-time location, and dating habits are key to the use of the app. Further, Grindr has a wide user base, with 27 million users as of 2017.[24] Grindr data has also been used against U.S. politicians before: A user of the app outed Randy Boehning, a Republican member of the North Dakota House of Representatives, after he voted against a gay rights bill.[25] The Trump administration was also greatly concerned about the possibility of the Chinese government combining commercial data it acquired from companies with data it stole in the hack of the U.S. Office of Personnel Management. Beijing Kunlun Tech, the logic went, ostensibly had or could exert technical control over Grindr's data collection and storing mechanisms through its ownership of the app.
- Second, on the legal side, the U.S. government has expressed concerns, particularly over the past few years, about the Chinese government's legal requirements for Chinese Communist Party access to Chinese companies' data. An oft-cited 2017 Chinese intelligence law

[24] Carl O'Donnell, Liana B. Baker, and Echo Wang, "Exclusive: Told U.S. Security at Risk, Chinese Firm Seeks to Sell Grinder Dating App," Reuters, March 26, 2019, https://www.reuters.com/article/us-grindr-m-a-exclusive/exclusive-us-pushes-chinese-owner-of-grindr-to-divest-the-dating-app-sources-idUSKCN1R809L.

[25] Peter Moskowitz, "Grindr User 'Outed North Dakota Politician in Retaliation for Anti-Gay Vote'," *Guardian,* April 28, 2015, https://www.theguardian.com/us-news/2015/apr/28/north-dakota-politician-randy-boehning-outed-grindr-nude-photos.

mandates that "'any organization … shall support, assist, and cooperate with state intelligence work.'"[26] As Beijing Kunlun Tech is incorporated in China, this would make the company subject to such laws.

- Third, CFIUS's logic about the risks of the Chinese government getting access to Grindr's data stemmed from the app's then-owner—as mentioned, linking that corporate ownership structure to Beijing Kunlun Tech's technical access (or effective technical access) to Grindr's data and Beijing Kunlun Tech's de jure and de facto obligations to comply with Chinese government data access requests. That Grindr was owned by a Chinese company both tied together technical and legal trust concerns and placed Beijing Kunlun Tech's acquisition of the app squarely within CFIUS's purview. Forcing a change in the Grindr ownership structure was CFIUS's way to mitigate these concerns about foreign state access to sensitive U.S. data. After the decision, Sens. Edward Markey and Richard Blumenthal said that CFIUS "should continue to draw a line in the sand for future foreign acquisition of sensitive personal data."[27]

This multipronged approach to trust, however—using technical, legal, and corporate governance criteria—also highlights the limitations of CFIUS's authorities in boosting trust in the Grindr app.

Indeed, the irony of the review was that while CFIUS prevented Chinese ownership of Grindr, it did not do anything—some might argue could not do anything—to prevent the company from then buying, selling, and licensing that data on the open market. In other words, Grindr's new corporate ownership structure, American though it is, does not prevent the app's data from going to China. Indeed, as becomes apparent when one examines the technical layer, Grindr shares data widely with third-party ad networks, data brokers, and software development kits, including an SDK made by Chinese giant Tencent. This was detailed in a January 2020 report by the Consumer Council of Norway.[28] In total, the Consumer Council observed Grindr communicating with fifty-three different unique domains and thirty-six different advertising companies, with eleven parties receiving the user's exact GPS location, four parties receiving the user's IP and/or MAC address, and seven parties receiving "personal information about the end user, such as age or gender."

---

[26] Murray Scot Tanner, "Beijing's New National Intelligence Law: From Defense to Offense," *Lawfare,* July 20, 2107, https://www.lawfareblog.com/beijings-new-national-intelligence-law-defense-offense.
[27] Carl O'Donnell, Liana B. Baker, and Echo Wang, "Exclusive: Told U.S. Security at Risk, Chinese Firm Seeks to Sell Grinder Dating App," Reuters, March 26, 2019, https://www.reuters.com/article/us-grindr-m-a-exclusive/exclusive-us-pushes-chinese-owner-of-grindr-to-divest-the-dating-app-sources-idUSKCN1R809L.
[28] Øyvind H. Kaldestad, "Out of Control," Forbrukerrådet, January 14, 2020, https://www.forbrukerradet.no/undersokelse/no-undersokelsekategori/report-out-of-control/.

The CFIUS-compelled sale of Grindr from Beijing Kunlun Tech to San Vicente Acquisition LLC may have mitigated against one potential area of concern vis-à-vis supply chain trust—the risk of a Chinese firm directly accessing sensitive U.S. citizen data at the Chinese government's behest—but CFIUS could not and did not address a range of other data-sharing practices that similarly cause sensitive U.S. citizen data to move throughout the internet ecosystem. Only a more integrated approach to trustworthiness could accomplish that.

COMBINING MEASURES OF TRUST

In the end, trustworthiness depends on a combination of hardware and software engineering, corporate governance, and state law and policy. None of these measures can establish trustworthiness in isolation. There are trade-offs among technical measures, organizational measures, and national laws and regulation. Different measures also come into conflict. There are technical means to circumvent national laws and regulation; see, for example, the interaction between law enforcement agencies in the United States and the United Kingdom as they cope with the end-to-end encryption offered by companies such as Apple. Under the right institutional circumstances, these same laws can cast doubt on the honest implementation of a protection like end-to-end encryption as, for example, with China's national intelligence law (which, though nominally neutral, appears to give the Chinese government plenary directive authority).[29]

Viewed at a high level, the fundamental dilemma is that trustworthiness must be asserted by people, not purely by technology, and thus can be undermined by deliberate human intervention or the perception of the possibility of it. Use of open standards and third-party auditing can help moderate this fear, but the fear will always remain to some degree, especially posed by national laws and regulation. Thus, corporate governance and how it contributes to a firm's reputation for competence and good governance are important factors in fostering trustworthiness, and they mediate assertions of trust made on the basis of technology alone. So too do the requirements of national laws and regulations. Our assessment of the strength of due process and rule of law in a state shape our belief that technical protections and laws may be fairly contested.

There is no magic wand to resolve these issues or to make them speak to each other on equal terms, and assessing factors like the rule of law in a state will remain a fuzzy area. In this section, however, we offer a means to characterize measures found in all three broad arenas discussed in this report—technical, corporate governance, and national laws and regulation—as a mechanism to put them into more direct conversation with one another.

[29] See the following translation: "National Intelligence Law of the P.R.C. (2017)," China Law Translate, June 27, 2017, https://www.chinalawtranslate.com/en/national-intelligence-law-of-the-p-r-c-2017/.

In effect, we seek to build a composite measure of trust. While we conclude, ultimately, that a unitary measurement of trust is impossible to achieve, the framework we outline here—a set of principles, if you will, for assessing trust—gives a well-intentioned evaluator a language with which to describe and explain the thought underlying trustworthiness decisions.

We start with two important overarching principles:

- *Transparency* is essential. This is true universally, whether it applies to technical measures, corporate governance, or national policy. If we answer the "transparency to whom" question correctly, then transparency allows the user of hardware or software to be informed about the risk of weakness in a supply chain. This sort of transparency involves visibility both into the technical assurance of a product and into the incentives of the producer, whether those are financial or the result of external governmental policy.
- Likewise, somewhere in the trustworthiness architecture lies the concept of *accountability* or *deterrence*. Degrading trustworthiness must have consequences. Those consequences can be legal (either through direct punishment or through indirect means like sanctions). They can be economic (through the loss of current income or through reputational harm that impacts future earnings). The deterrence can even be seen in quasi-economic or legal impacts through the uses of soft power. But in the absence of adverse consequences for creating untrustworthy products, there is little or no incentive (save pure altruism) to provide any assurance of trustworthiness.

With those two principles in mind, we offer the following as a composite sliding scale of trustworthiness. We recognize, as we must, that what follows are not absolute values. Nonetheless, the fact that there are exceptions to these general statements in specific cases does not detract, in our view, from the general utility of expressing these principles. The reader is cautioned that the various bullet points are not necessarily all of equal weight—indeed, the appropriate weighting is determined by the specifics of a case at hand.

The trustworthiness impact of *technical measures* is …

Strengthened

- by transparency to the customer or user
- by corporate requirements that require similar or mirrored practice from vendors
- by the use of open, nonproprietary standards and implementation that can be examined publicly

- by allowing outputs from these measures to be externally tested, audited, and validated
- by defense in depth, in which multiple protection barriers have to be breached for an attack to succeed

Weakened

- by the misalignment of the technology (or illegality) under national laws and regulations
- by being unsupported by other vendors or by using a nonstandard methodology (though, we acknowledge, this will also be true in cases of genuine technical progress)
- by being misapplied
- with evidence of poor technical capacity or weak implementation of other features or products
- by single points of trust and/or failure

The trustworthiness impact of *organizational measures* is …

Strengthened

- if they are rooted in common industry practice, at least where common industry practice is not subject to known criticism as eroding trustworthiness
- when they are transparent and auditable
- when they are aligned with national laws and regulations, at least to the extent those laws and regulations are the product of transparent, legitimate democratic processes
- if they are well established rather than recent

Weakened

- by a history of related malfeasance or failings
- if they are new or adopted in response to a crisis
- where poorly aligned with a company's business model
- where unsupported by complementary technical measures

The trustworthiness impact of an originating nation's *laws and regulations* is …

Strengthened

- when the law or regulation is applied to citizens or legal subjects as well as extraterritorially
- when the law or regulation is subject to transparency through public notice of some form (even in cases where the data is aggregated and delayed)
- when the law or regulation is subject to transparent and independent oversight
- when the law or regulation is subject to credible public debate and challenge
- when the law or regulation is applied uniformly to all citizens (instances)

Weakened

- when the law or regulation is implemented without due process or meaningful means of challenge
- when enforcement is subject to unobservable discretion of a single agency or authority
- where the law or regulation is developed or heavily influenced by a single entity or security agency
- when the law or regulation is not subject to public debate and industry feedback
- where there is no track record of decisions being made against government interests

CONCLUSION

People or organizations making trustworthiness judgments regarding given acquisitions or implementation decisions will undoubtedly find that the various factors we have outlined point often in different directions. Thus they will have to draw conclusions about how different trustworthiness measures trade off against one another. In the end, the challenge revolves around three interrelated questions:

- How does one build an artifact (a component, a device, or a system) that is trustworthy?
- How does one assess the trustworthiness of the artifact?
- How does one decide to treat an artifact as trustworthy?

These questions are distinct yet connected. Trade-offs in identifying acceptable answers to them may be driven by a variety of factors, including the risk tolerance of the user or organization, the use case for the product, and the environment in which the product can be used.

There are those hoping for a one-and-done, universally applicable algorithm to evaluate trustworthiness. Alas, there are no silver bullets for trustworthiness, and no substitute for careful, considered judgments made on a case-by-case basis. By distinguishing between products that bear many of the indicia of trust we have outlined and those that do not, we believe that the principles described in this report can help to frame trustworthiness assessments of the supply chain and provide a way forward for decision-makers.

APPENDIX A

Task Force Rapporteurs

Paul Rosenzweig (Chief Rapporteur) — Paul Rosenzweig is the founder of Red Branch Consulting PLLC, a homeland security consulting company. He is also a senior adviser to The Chertoff Group and formerly served as deputy assistant secretary for policy in the Department of Homeland Security. He is a professorial lecturer in law at George Washington University, an adviser to and former member of the American Bar Association (ABA) Standing Committee on Law and National Security, and a contributing editor of *Lawfare*. He is also a member of the ABA Cybersecurity Legal Task Force and a senior editor of the *Journal of National Security Law & Policy*.

Justin Sherman (Assistant Rapporteur) — Justin Sherman is a fellow at the Atlantic Council's Cyber Statecraft Initiative, a research fellow at the Tech, Law & Security Program at American University Washington College of Law, and a fellow at Duke University's Sanford School of Public Policy. He is also an independent cybersecurity, technology policy, and geopolitical risk consultant.

Task Force Chairman

Benjamin Wittes — Benjamin Wittes is a senior fellow in Governance Studies at the Brookings Institution. He co-founded and is the editor-in-chief of *Lawfare*, which is devoted to sober and serious discussion of "Hard National Security Choices." He is a contributing writer at the *Atlantic* and a law analyst at NBC News and MSNBC. Wittes is the author with Susan Hennessey of *Unmaking the Presidency: Donald Trump's War on the World's Most Powerful Office*. His previous books include *The Future of Violence: Robots and Germs, Hackers and Drones-Confronting A New Age of Threat* (2015), co-authored with Gabriella Blum; *Detention and Denial: The Case for Candor After Guantanamo* (2011); *Law and the Long War: The Future of Justice in the Age of Terror* (2008); *Confirmation Wars: Preserving Independent Courts in Angry Times* (2006); and Starr: A Reassessment (2002). He has edited three books: *Campaign 2012: Twelve Independent Ideas for Improving American Public Policy* (2012), *Constitution 3.0: Freedom and Technological Change* (2011), and *Legislating the War on Terror: An Agenda for Reform* (2009).

Task Force Members

Trey Herr — Trey Herr is the director of the Cyber Statecraft Initiative under the Scowcroft Center for Strategy and Security at the Atlantic Council. His team works on cybersecurity and geopolitics including cloud computing, the security of the internet, supply chain policy, cyber effects on the battlefield, and growing a more capable cybersecurity policy workforce. Previously, he was a senior security strategist with Microsoft handling cloud computing and

supply chain security policy as well as a fellow with the Belfer Cyber Security Project at Harvard Kennedy School and a nonresident fellow with the Hoover Institution at Stanford University. He holds a doctorate in political science and a bachelor of science degree in musical theatre and political science.

David Hoffman — David Hoffman is the Steed Family Professor of the Practice of Cybersecurity Policy at the Sanford School of Public Policy. He previously served as associate general counsel and global privacy officer for Intel Corporation. Hoffman currently chairs the Civil Liberties and Privacy Panel of the Director's Advisory Board for the U.S. National Security Agency. He also chairs the board of the Center for Cybersecurity Policy and Law and serves on the advisory boards for the Future of Privacy Forum and the Israel Tech Policy Institute. Hoffman also founded and chairs the board for the Triangle Privacy Research Hub, which highlights and fosters cybersecurity and privacy academic research done in the North Carolina Research Triangle.

Herb Lin — Herb Lin is senior research scholar for cyber policy and security at the Center for International Security and Cooperation and Hank J. Holland Fellow in Cyber Policy and Security at the Hoover Institution, both at Stanford University. His research interests relate broadly to policy-related dimensions of cybersecurity and cyberspace, and he is particularly interested in the use of offensive operations in cyberspace as instruments of national policy and in the security dimensions of information warfare and influence operations on national security. In addition to his positions at Stanford University, Lin is chief scientist, emeritus for the Computer Science and Telecommunications Board, National Research Council (NRC) of the National Academies, where he served from 1990 through 2014 as study director of major projects on public policy and information technology; adjunct senior research scholar and senior fellow in cybersecurity (not in residence) at the Saltzman Institute for War and Peace Studies in the School for International and Public Affairs at Columbia University; and a member of the Science and Security Board of the Bulletin of Atomic Scientists. In 2016, he served on President Obama's Commission on Enhancing National Cybersecurity. Prior to his NRC service, Lin was a professional staff member and staff scientist for the House Armed Services Committee (1986–1990), where his portfolio included defense policy and arms control issues. He received his doctorate in physics from MIT.

Bart Preneel — Bart Preneel is professor in applied cryptography at the Department of Electrical Engineering of the University of Leuven (Belgium), where he heads the COSIC research group. He was elected to membership in the Academia Europaea and is a fellow and past president of the IACR (International Association for Cryptologic Research). He is a co-founder and board member of the start-up nextAuth and a board member of Approach Belgium. He serves as an advisory board member of the European Union Agency for Cybersecurity (ENISA).

Samm Sacks — Samm Sacks is a senior fellow at Yale Law School's Paul Tsai China Center. Her research examines China's information and communications technology policies, with a focus on China's cybersecurity legal system, the U.S.-China technology relationship, and the geopolitics

of data privacy and cross-border data flows. Previously, Sacks launched the industrial cyber business for Siemens in China, Japan, and South Korea. Prior to this, she led China technology sector analysis at the political risk consultancy Eurasia Group and worked as an analyst and Chinese linguist with the national security community. She is a frequent contributor to the media, and her articles have appeared in outlets including the *Atlantic*, *Foreign Affairs*, and *Slate*. She has testified multiple times before Congress on China's technology and cyber policies. Sacks is also a cyber policy fellow at New America and a former Fulbright scholar in Beijing. She holds a master of arts degree in international relations from Yale University and a bachelor of arts degree in Chinese literature from Brown University.

Fred B. Schneider — Fred B. Schneider is Samuel B. Eckert Professor of Computer Science and former department chair at Cornell University. He was elected to membership in the U.S. National Academy of Engineering, its Norwegian counterpart, and the American Academy of Arts and Sciences. Schneider is founding chair of the National Academies Forum on Cyber Resilience, and he co-chaired Microsoft's external cybersecurity advisory board (TCAAB) for its eleven-year life. He has also served as a member of the Defense Science Board and the Naval Studies Board.

Daniel Weitzner — Daniel Weitzner is the founding director of the MIT Internet Policy Research Initiative (IPRI), a multidisciplinary effort producing technically grounded research to address critical internet public policy questions such as privacy, information accountability, cybersecurity, surveillance, network management, and trustworthy artificial intelligence systems. He is also the 3Com Founders Principal Research Scientist at the MIT Computer Science and Artificial Intelligence Research Lab (CSAIL).

APPENDIX B

ANNOTATED BIBLIOGRAPHY[30]

This annotated bibliography identified sources related to four categories:

- Political and legal criteria
- Corporate governance criteria
- Technical criteria for hardware
- Technical criteria for software

Political and Legal Criteria

CSIS Working Group on Trust and Security in 5G Networks, *Criteria for Security and Trust in Telecommunications Networks and Services* (Washington, DC: Center for Strategic & International Studies, May 2020).

> This report of the CSIS Working Group on Trust and Security in 5G Networks was requested by the State Department. Criteria are designed to complement the Prague Proposal and the European Union's 5G Toolbox, and they rely primarily on publicly available information. The criteria are broken into categories as follows: ten political and governance criteria (for example, suppliers are more trustworthy if headquartered in democracies with an independent judiciary and the rule of law); seven business practices assessment criteria (for example, suppliers are more trustworthy if they are transparently owned and publicly traded); ten cybersecurity risk mitigation criteria (for example, the supplier has passed independent, credible third-party tests; the technology uses open and consensus-based standards; the supplier has a record of patching systems in reasonable time); and four government actions to increase confidence in the choice of a supplier (for example, selection of diverse suppliers, government and private-sector ability to regularly conduct vulnerability tests and risk assessments).

Levite, Ariel, *ICT Supply Chain Integrity: Principles for Governmental and Corporate Policies*, (Washington, DC: Carnegie Endowment for International Peace, October 2019). [also Corporate Governance]

[30] A version of this annotated bibliography was originally published as Paul Rosenzweig and Claire Vishik, "Trusted Hardware and Software: An Annotated Bibliography," *Lawfare*, October 1, 2020, https://www.lawfareblog.com/trusted-hardware-and-software-annotated-bibliography.

This paper proposes several measures for governments and corporations to undertake in order to increase trust in the integrity of information, communications, and operational technology supply chains. For example, it calls on governments to refrain from systemic interventions in supply chains and to establish interagency processes to consider the equities of potential interventions. The proposed corporate obligations include not supporting systemic interventions in their supply chain; protecting products and services throughout their life cycles; and accommodating reasonable, lawful requests for information. The paper goes on to outline how these obligations could be transformed into a binding normative framework through formal agreements or other arrangements that incentivize compliance. Finally, the paper explores approaches to verifying government and corporate compliance with the proposed obligations. Some of the proposals in the paper are not pragmatic (for example, the processes suggested for the resolution of issues), and technology controls are not well covered, but the paper provides a broad overview of the international aspects of the problem.

Moran, Theodore H., *CFIUS and National Security: Challenges for the United States, Opportunities for the European Union* (Washington, DC: Peterson Institute for International Economics, 2017).

This paper analyzes the Committee on Foreign Investment in the United States (CFIUS) approach to evaluating the risk of foreign investments. CFIUS takes a narrow approach to identify threats within sectors rather than issue blanket bans, and it focuses on security rather than economic effects (positive or negative). The paper points out that Trump's strategy of increasing protectionism opens the door to reciprocal retaliation. It discusses historical politicization of the CFIUS process and its impact on investment reviews. It also discusses early-day references to "national security" without definition and to foreign "control" with only a vague definition. In addition, the paper suggests a "three threat" framework to approach CFIUS reviews: the possible leakage of sensitive tech to a foreign company or government in ways that could harm U.S. interests; the ability of foreign acquisitors to delay, deny, or place conditions on outputs from newly acquired producers; and the potential that acquisition could allow a foreign company or its government to penetrate systems for monitoring, surveillance, or planting of malware. Then it discusses several examples that fit into this framework. A critical recommendation is for CFIUS to preclude foreign acquisitions from certain countries across entire sectors rather than evaluate national security threats within sectors. Lessons from the CFIUS strategy can be applied to the evaluation of imported technology, in terms of both security and political considerations.

Corporate Governance Criteria

Boyson, Sandor, Thomas Corsi, Hart Rossman, and Matthew Dorin, *Assessing SCRM Capabilities and Perspectives of the IT Vendor Community: Toward a Cyber-Supply Chain Code of Practice* (College Park: University of Maryland, Robert H. Smith School of Business, 2011).

> The project surveyed the cyber–supply chain risk management (SCRM) capabilities of 131 firms, using a questionnaire designed by the authors based on contributions from a variety of public- and private-sector agencies. The study found that companies of all sizes undermanage cyber-SCRM, an especially dangerous trend as companies are increasingly complicating their own supply chain risk profiles by working across one or more product/service boundaries (software, hardware, telecom/data networking, etc.). The study also found that companies of all sizes can be given incentives to improve cyber-SCRM management.

Charney, Scott, and Eric T. Werner, *Cyber Supply Chain Risk Management: Toward a Global Vision of Transparency and Trust* (Redmond, WA: Microsoft, 2011). [also Political and Legal]

> This paper argues in favor of a risk management approach to ensuring trustworthy hardware and software that addresses national security imperatives without threatening the vitality of the global information and communications technology (ICT) sector. In particular, the authors argue, efforts to enhance trust in ICT supply chains must embrace four principles: First, they must be risk based and utilize collaboratively developed standards. Second, they should promote transparency by both vendors and governments. Third, they must be flexible, allowing for suppliers to implement different types of controls and mitigations based on the technology they provide. Finally, participants (especially governments) must acknowledge that closing markets based on supply chain concerns will lead to reciprocal behaviors, threatening the global ICT sector.

Fan, Yiyi, and Mark Stevenson, "A Review of Supply Chain Risk Management: Definition, Theory, and Research Agenda," *International Journal of Physical Distribution & Logistics Management,* 48, no. 3 (April 2018).

> This literature review systematically examines the range of scholarship in supply chain risk management (SCRM) to develop a new comprehensive definition of SCRM, present current research on the four stages of SCRM (risk identification, assessment, treatment, and monitoring), and, in particular, understand the use of theory in SCRM. The review identifies ten gaps in the SCRM literature. Notable findings include the following: There are few attempts at studying the SCRM process holistically, a holistic framework is necessary for categorizing risks, risk monitoring is understudied, the field must develop SCRM strategies that provide guidance for practitioners, theories must be used more appropriately to deepen understanding of SCRM, the literature pays little attention to SCRM in developing-country contexts, and more research is needed from the supplier perspective.

Technical Criteria for Hardware

Defense Science Board, *Defense Science Board Task Force on Cyber Supply Chain* (Washington, DC: U.S. Department of Defense, April 2017).

This report was issued by the task force charged with assessing organizations, missions, and authorities related to microelectronics and components used in Defense Department weapons systems. The report notes the rising complexity of microelectronics and that the Defense Department "has become a far less influential buyer in a vast, globalized supplier base." It breaks down the supply chain into the Defense Department acquisition supply chain, the Defense Department sustainment supply chain, and the global commercial supply chain. It also notes flaws in tracking threats to those supply chains, including inadequate Defense Department tracking of inventory obsolescence and vulnerabilities, foreign ownership and competition that reduces the department's influence, and the lack of formal mitigation processes. Key research recommendations are to develop formal language to describe the scope of means for defense "given some assumed attack classes and capabilities for attackers" and to devise algorithms to automatically assess those possible defenses. Recommendations for the department divide into three assurance types: axiomatic (purchase from a wide set of suppliers), synthetic (use tamper-proof packing and unforgeable marking), and analytic (record provenance and assign trust based on that, and collect sample measurements of the system to test how a single instance runs). Additional research, especially in the area of cryptography-based integrity, is recommended.

Nissen, Chris, John Gronager, Robert Metzger, and Harvey Rishikof, *Deliver Uncompromised: A Strategy for Supply Chain Security and Resilience in Response to the Changing Character of War* (McLean, VA: MITRE Corporation, August 2018). [also Technical Criteria for Software]

This report concludes that the Defense Department and the intelligence community are "generally aware" of supply chain threats but don't adequately share knowledge and coordinate approaches. A compliance-focused approach prioritizes meeting minimums. Instead, eight lines of effort at the enterprise level (for the Defense Department and its contractors) are needed for the department to "deliver uncompromised": elevate, educate, coordinate, reform, monitor, protect, incentivize, and assure. This approach is coupled with fifteen courses of action, including elevating security as a primary metric in Defense acquisition and sustainment; forming a whole-of-government National Supply Chain Intelligence Center; identifying and empowering the chain of command for supply chain security and integrity accountable to the deputy secretary of defense; centralizing the Supply Chain Risk Management-Threat Assessment Center (SCRM-TAC) with an industrial security/counterintelligence mission owner under the Defense Security Service (DSS) and extending DSS authority; establishing independently implemented, automated, continuous monitoring of Defense Innovation Board software; extending National Defense Authorization Act Section 841 authorities for "Never Contract with the Enemy"; and instituting industry-standard information technology practices in all software developments (including possibly a software bill of materials).

Sacks, Samm, and Manyi Kathy Li, *How Chinese Cybersecurity Standards Impact Doing Business in China* (Washington, DC: Center for Strategic & International Studies, August 2018). [also Technical Criteria for Software]

> This brief outlines the system of security laws and regulations adopted by China to control the importation of foreign technology, including the Multi-Level Protection Scheme and a new Cybersecurity Law governing critical information infrastructure. Rather than establishing one set of clear legal requirements, the Chinese system is made up of a number of layers with gray areas regarding jurisdiction and compliance requirements, seemingly designed to allow officials to apply the law as they see fit. Most provisions are characterized as "recommended" but are in fact required when used in state procurement requirements. The Chinese government deliberately uses vague language in standards to avoid highlighting problematic issues externally (for example, in the World Trade Organization), while retaining for itself maximum flexibility for internal application of the law. The brief includes an appendix of more than 300 translated Chinese cybersecurity legal provisions.

Sethumadhavan, Simha, Adam Waksman, Matthew Suozzo, Yipeng Huang, and Julianna Eum, "Trustworthy Hardware From Untrusted Components," *Communications of the ACM*, 58, no. 9 (September 2015), 60–71.

> This paper outlines a three-system approach to security aimed at making any attack as expensive as possible. In the first system, the design is checked for backdoors. In the second system, inputs to hardware circuits are altered to prevent any triggers from activating the backdoor. In the third system, on-chip monitoring detects if a backdoor has turned on, allowing for it to be disabled or gracefully shut down. In both practical and theoretical domains, this paper makes a great number of unique points.

Technical Criteria for Software

Cabinet Office, *Supplier Assurance Framework: Good Practice Guide* (London: United Kingdom Cabinet Office, May 2018).

> This report offers a straightforward, proportionate, and transparent government approach to supplier information assurance when operating at OFFICIAL and OFFICIAL-SENSITIVE levels. (Contracts at SECRET and above are covered by List X criteria.) Common Criteria for Assessing Risk (CCfAR) is intended to be a continually updated risk assessment of suppliers that contains twenty criteria, nine of which are defined as critical and eleven as significant, and is used to broadly group suppliers into high, medium, and low risk based on a scoring sheet. Criteria span the sensitivity of stored data, methods for storing data, the number of transfers of data in the contractor's supply chain, and the possibility of certifying the contractor's information security controls. The report provides frameworks for integrating

> CCfAR into assessments of existing contracts and procurements of new ones. It also notes that the Security Policy Framework requires U.K. government agencies to reassess contracts every year for compliance with existing rules and to report risks to the Cabinet Office.

HCSEC Oversight Board, *Huawei Cyber Security Evaluation Centre (HCSEC) Oversight Board Annual Report* (Banbury, United Kingdom: HCSEC Oversight Board, March 2019). [also Technical Criteria for Hardware]

> The Huawei Cyber Security Evaluation Centre (HCSEC) Oversight Board's annual report provides details relating to the board's two-part mandate: to report on the HCSEC's assessment of Huawei's U.K. products as relevant to U.K. national security, and to evaluate the competence and independence of the HCSEC in relation to that mission. The board provides only limited assurance that the long-term security risks posed by Huawei equipment currently installed in the U.K. can be managed. Moreover, the board is not confident in Huawei's ability to meaningfully improve its software engineering and cybersecurity processes, which are the sources of vulnerabilities in Huawei equipment. The board finds the HCSEC both competent and independent based on independent audit.
>
> The HCSEC received four products from Huawei to test binary equivalence, validation of which was still ongoing at the time of the report but whose process exposed wider flaws in the build process. The HCSEC also looked at configuration management, operating system use, and life cycle management, and it performed a software engineering analysis comparing subsequent major versions of Huawei products to look for major improvements (an analysis that found major defects in new versions).
>
> This model of oversight is highly controversial, viewed as ineffective by some, and in the process of reevaluation by the U.K. government.

Herr, Trey, June Lee, William Loomis, and Stewart Scott, *Breaking Trust: Shades of Crisis across an Insecure Software Supply Chain* (Washington, DC: Atlantic Council, July 2020).

> This report examines the exploitation of technology supply chains by nation-states that has far-reaching impacts on the public, government, and industry. It also addresses what steps industry, the Defense Department, and other government entities can take to mitigate persistent vulnerabilities in software supply chains. The report analyzes more than 110 cases of software supply chain attacks and disclosures, showcasing critical vulnerabilities and trends in state and non-state attacks, and offering actionable recommendations for securing trust in the software supply chain.

Jackson, Daniel, Martyn Thomas, and Lynette I. Millett, eds., *Software for Dependable Systems: Sufficient Evidence?* (Washington, DC: National Academies Press, 2007).

> This report is directed to the question of whether or not direct observation of a system can provide better assurance of its trustworthiness than the credentials of its production method. It recognizes that the complexity of software systems, as well as the discontinuous way they behave, renders them extremely difficult to analyze. In the end, the report outlines a proposed approach for what it calls "explicit dependability claims" that advance the idea of a system for making software dependable in a cost-effective manner.

National Academy of Sciences, *Summary of Workshop on Software Certification and Dependability* (Washington, DC: National Academies Press, 2004).

> This report summarizes a workshop on software certification and dependability convened by the National Academy of Sciences. While wide ranging, the workshop identified at least two particularly salient conclusions. First, although following particular processes cannot alone guarantee certifiably dependable software, comprehensive engineering processes are nevertheless important to achieving this goal. And second, the process of certification may add value in a collateral fashion because attention must be paid to issues that might not receive it otherwise; given that software and its uses and contexts change over time, any value that certification has decays over time as well.

Neumann, Peter G., *Fundamental Trustworthiness Principles* (Menlo Park, CA: SRI International, March 2017). [also Technical Criteria for Hardware]

> This paper enumerates principles for designing hardware and software architecture and for assuring that the systems can satisfy mission needs. It then assesses whether and how the current Capability Hardware Enhanced RISC Instructions (CHERI) successfully embrace those principles to create functional and trustworthy software. The paper describes two sets of principles: Saltzer-Schroeder-Kaashoek Security Principles and the author's own principles for system development. In general, the two sets share a focus on developing sound architecture by simplifying where possible and adhering to the principle of least privilege. They also reject the notion that secrecy of design enhances security, advocating for open design methods. The author, however, cautions against blindly applying these principles, noting that individual principles can be in tension with each other in certain circumstances. The paper notes that CHERI architecture uses most of these principles constructively.

APPENDIX C

DEFINING A TRUSTED SUPPLY CHAIN[31]

Trust in a particular digital component or artifact is, in the end, dependent on trust in the underlying supply chain that produced the component. In this appendix we offer a more detailed description of the risks inherent in a digital supply chain and a definition of what that means for trustworthiness.

With a supply chain attack, there is a potentially long delay between the introduction of a vulnerability and its exploitation. In addition, infiltrating a supplier generally requires a well-resourced adversary and interaction with that supplier. So compared to the alternatives, preparations for a supply chain attack take longer and have a higher risk of discovery. The risks of discovery can be reduced, however, if inserted vulnerabilities resemble ordinary flaws and, thus, the malicious intent is disguised.

There are nevertheless still good reasons to undertake a supply chain attack through any of the various possible means. First, for systems that will not be accessible after deployment, introducing a vulnerability during system development or manufacture might be the only means of compromise. Second, supply chain attacks can, in some forms, exhibit a certain economy of scale, since vulnerabilities are installed into all new instances of a system and they can also, conversely, be highly targeted at a single digital system or small group of systems. Third, vulnerabilities exploited by supply chain attacks can be well concealed, offering the attacker reliable access when needed.

Characteristics of a supply chain are sometimes part of such justifications. As we discuss in this report, the same bases for justifying trust in a digital system also shed light on schemes being proposed to detect or forestall supply chain attacks.[32]

Most supply chains are actually not chains. For establishing trust in a digital artifact, supply chains are better described by trees. Figure 1 shows a supply chain depicted as a tree. The *sink* of the tree at the far right represents the artifact of interest; it has no outgoing edges. The other nodes in the tree also represent artifacts, where a directed edge from a node *n* to a node *m* signifies that the artifact that *n* denotes is used in producing the artifact that *m* denotes. Updates to artifacts thus also follow the paths depicted by the edges in a supply chain tree. For our purposes, artifacts will correspond to

---

[31] Portions of this appendix were originally published as Fred B. Schneider and Justin Sherman, "Bases for Trust in a Supply Chain," *Lawfare*, February 1, 2021, https://www.lawfareblog.com/bases-trust-supply-chain.

[32] Fred B. Schneider, "Putting Trust in Security Engineering," *Communications of the ACM,* 61 (May 2018), 37–39, http://www.cs.cornell.edu/fbs/publications/CACM.viewpoint.PuttingTrust.pdf.

digital systems and their components, ranging from wafers to networked information systems. Associated with each node could be the manufacturer and other attributes used for an axiomatic basis for trust.

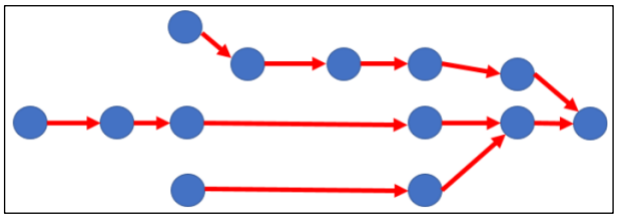

FIGURE 1. EXAMPLE OF A SUPPLY CHAIN

To establish trust in the artifact that the sink models, it might be tempting to focus on that artifact and ignore the rest of the supply chain. That view, however, is shortsighted:

- The use of synthesized bases is precluded, since they require knowledge of and trust in the artifacts being combined to form the sink.
- Analytic bases that employ testing might not be feasible, because access is required to the artifact's internal components for submitting inputs and for observing outputs. Those internal components correspond to other nodes in the tree.

Accessing the other artifacts in the supply chain overcomes these limitations. In theory, it should suffice to identify a *cut*—a set of nodes that intersect all paths to the sink from the leaves of the supply chain tree. The green nodes in Figure 2 are an example of a cut for the tree in Figure 1. To establish trust in the sink when there has been a cut, it suffices to establish (a) trust for nodes forming the cut and (b) the needed transitivity of trust by having a basis to trust each successor under the assumption that its immediate predecessors can be trusted.

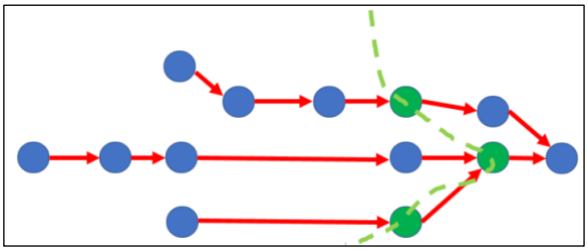

FIGURE 2. EXAMPLE OF A CUT IN A SUPPLY CHAIN TREE

In theory, obligation (a), trust for nodes forming the cut, would be discharged by using one or more of the bases for trust discussed above. In practice, for most digital artifacts, beliefs derived in this way will be incomplete and, as noted above, the trust bases do benefit from access to predecessor nodes. The obvious remedy is to work backward through the supply chain, establishing trust in predecessors of the nodes forming the cut. Just running tests on an artifact could, for example, miss

problematic behaviors; running tests on the artifact's predecessors in the supply chain might reveal some of those problems. However, for various reasons it might not be possible to learn the full pedigree for an artifact in a supply chain. That means trust in an artifact corresponding to a node at the cut—and therefore trust in the artifact at the sink—is necessarily going to be incomplete.

Obligation (b), transitivity of trust, involves the moral equivalent of tamper-proof packaging, so that users can detect when an artifact they trust becomes compromised as it progresses through the supply chain. For semiconductor chips, this guarantee can be enforced by physical packaging. For software, hash functions or other cryptographic means along with deterministic compilation and builds can be used to compute checks that would indicate when bits have been changed; hardware support (for example, the Trusted Platform Module and Software Guard Extensions) for so-called *trusted computing* is helpful here.[33] Self-test capabilities also can be effective for defending against certain attacks that compromise tamper-proof packaging of hardware or software.

Both obligations can benefit from deterrence through accountability, even when some of the institutions involved in producing the components are not subject to the regulations (and prosecution) that is creating the deterrence. If trust in *n* is being justified by deterrence through accountability, then the producer of *n* has reason to enforce requirements on its suppliers. So, a supply chain creates a *reverse cascade* that enforces requirements on suppliers even if those suppliers are not within the jurisdiction of the institution implementing the accountability or the deterrence.[34]

Finally, cryptographic techniques such as threshold cryptography and secure multiparty computation (MPC) allow one to share information over multiple parties and compute over this shared data in a distributed way. The protocol can be designed such that compromising one or more nodes in terms of confidentiality, integrity, or both does not compromise the outcome. One can even push this approach to the limit and achieve confidentiality if at least one of the components is trustworthy. In terms of supply chain security, this means that one can build a system from components of independent manufacturers and achieve trustworthiness in terms of integrity and confidentiality even if the majority (or sometimes fewer) of the components are trustworthy. MPC comes at a high cost in terms of communications overhead; latency can be reduced by

---

[33] Paul England, Butler Lampson, John Manferdelli, Marcus Peinado, and Bryan Willman, "A Trusted Open Platform," *Computer*, 36 (July 2003), 55–62, https://ieeexplore.ieee.org/abstract/document/1212691.

[34] Nathaniel Kim, Trey Herr, and Bruce Schneier, *The Reverse Cascade: Enforcing Security on the Global IoT Supply Chain* (Atlantic Council Scowcroft Center for Strategy and Security, Cyber Statecraft Initiative, June 2020), https://www.atlanticcouncil.org/wp-content/uploads/2020/06/Reverse-Cascade-Report-v3.1.pdf.

precomputation. Figure 3 represents this visually: Each of the nodes in Figure 1 can be replaced by composite nodes from different manufacturers with fault tolerance or MPC techniques; this would reduce the dependence on a single manufacturer at the cost of a high complexity.

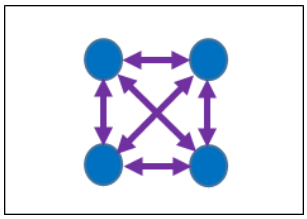

FIGURE 3. COMPOSITE NODE