~~Security-by-design~~ Securing a compromised system

Awais Rashid[1][0000−0002−0109−1341], Sana Belguith[1][0000−0003−0069−8552], Matthew Bradbury[2][0000−0003−4661−000X], Sadie Creese[3][0000−0002−2414−9657], Ivan Flechais[3][0000−0002−3620−0843], and Neeraj Suri[2][0000−0003−1688−1167]

[1] University of Bristol, UK
[2] Lancaster University, UK
[3] University of Oxford, UK

Abstract. Digital infrastructures are seeing convergence and connectivity at unprecedented scale. This is true for both current critical national infrastructures and emerging future systems that are highly cyber-physical in nature with complex intersections between humans and technologies, e.g., smart cities, intelligent transportation, high-value manufacturing and Industry 4.0. Diverse legacy and non-legacy software systems underpinned by heterogeneous hardware compose on-the-fly to deliver services to millions of users with varying requirements and unpredictable actions. This complexity is compounded by intricate and complicated supply-chains with many digital assets and services outsourced to third parties. The reality is that, at any particular point in time, there will be untrusted, partially-trusted or compromised elements across the infrastructure. Given this reality, and the societal scale of digital infrastructures, delivering secure and resilient operations is a major challenge. We argue that this requires us to move beyond the paradigm of security-by-design and embrace the challenge of *securing-a-compromised-system*.

Keywords: Security · Convergence · Cyber physical systems

1 Introduction

The security of infrastructures, architectures, and mechanisms is built on assumptions. This includes speculations or approximations about context, usage, threat models or interactions with other systems. Even when correct, these assumptions may only hold at a particular point in time and are often shaped by additional assumptions about the system's lifespan and that the designed security approaches will mitigate against vulnerabilities over that lifespan. These assumptions do not survive contact with the reality of deployed systems.

In practice, a system involves a range of other sub-systems several of which are not in the purview of the developers or the organisation deploying the system—in many instances assets and services are outsourced to third parties with security breaches having major knock-on effects on a wide array of systems and users [6]. Even where these sub-systems are within the development or administrative control of the system owner, there are complex technology stacks with

a plethora of third party libraries, hardware, software components and diverse development practices – including often misplaced assumptions about threat models [7, 20]. Furthermore, threat actors evolve quickly in terms of their capabilities, motivations and tactics, techniques and procedures, for example using generative AI techniques to create malware [13]. Systems, particularly large-scale ones that underpin societal scale infrastructures, e.g., water, power, digital services for citizens, do not evolve as rapidly. It takes time and money to change a system and, where such change is enacted, for example, for a power system or railway infrastructures, it is a major investment of hundreds of millions or billions of pounds involving rearchitecting the system, upgrading hardware and software systems, testing for safety and uptime, and retraining of staff. In some cases, it is even not possible to upgrade legacy systems to state-of-the-art security mechanisms due to real-time requirements or the need to formally prove safety and dependability related properties.

Furthermore, digital infrastructures have complex interdependencies and intersections with human users who are an integral part of the work and information flows. Often, human interactions with the systems catalyse dynamic composition of services which create new interactions and dependencies across systems at runtime. Usability of security mechanisms is paramount [23] not only to ensure that security does not create significant overheads but also to mitigate against shadow security practices [18] by users. Security mechanisms typically aim to address specific threats or vulnerabilities. For example, the Digital Security by Design (DSbD) programme is making key advances to eliminate memory vulnerabilities at the hardware-level [9]. While this holds great promise, there remain risks of developer-induced vulnerabilities [28] or constraining assumptions as to who the threat actor is, e.g., one aiming to extract data from RAM after super cooling it [29].

The reality is that it is *impossible* to secure all aspects of a system by design. Measuring security – and its *goodness* – is an open problem and polling *goodness* of a system cannot perfectly determine if the system's behaviour is good. The best one can do is probabilistic [4]. The reality is that systems will become compromised or will always have untrusted, partially trusted, or compromised elements. Pragmatic considerations mean also that one cannot simply shutdown a whole transportation infrastructure because, say, a traffic signal is compromised, or disconnect large parts of the power grid because specific components are under attack. How we ensure that the system continues to operate within specified bounds of safety and resilience – albeit potentially at reduced capacity – is critical, as is the capability to limit impacts of partial breaches including cascading effects across interconnected infrastructures. *We, therefore, posit that research needs to move beyond the paradigm of security-by-design and embrace the challenge of securing-a-compromised-system.* This requires scientific advances in four key dimensions. We discuss these next to present a research agenda for the research community.

2 Research Challenges

2.1 Predictability

Predictability is an inherent goal in security: knowing what can and will happen, what can be done to mitigate it and the extent to which any mitigation is effective. Predictability requires *measuring security* which is a hard problem in any system. It is compounded in digital infrastructures as complexity is paramount: mix of technology (legacy and non-legacy), uncertainty about threats and effectiveness of controls, emergent behaviour, interactions between security and other system goals, trustworthiness of people and organisations and divergence from rules (shadow practices).

A large body of work has focused on developing metrics. Reference sources such as NIST 800-55 [5] and ISO 27004 [15] adopt a catalogue approach: reference metrics classified into categories and documented with scenarios and examples. However, the contextualisation of metrics relies on arbitrary examples and use cases, limiting their expressiveness and hence their ability to address the complexity and inherent uncertainty. Others promote a more structured way of designing security measurements [12, 17, 11]. However, they presume that one knows a priori what is pertinent to measuring security and that instrumenting all elements is feasible—not the case given the dynamism and opaqueness in contemporary and future digital infrastructures.

Standards such as NIST SP 800-160 Volumes 1 and 2 [21, 22] offer guidance on engineering trustworthy secure and resilient systems. However, such standards are based on the premise that the problem, solution and trustworthiness contexts can be established *a priori* and that systems can be architected with a high degree of control over their components. These assumptions do not hold in large-scale infrastructures. There are systems about which one can collect relevant metrics (e.g., a sub-system into which deep instrumentation can be deployed) and for others one can not. Uncertainty also comes from what is unseen, e.g., shadow practice. So modelling the dependencies and deriving relevant metrics to understand the security implications of those dependencies is a major scientific challenge.

2.2 Composition

Composing security provision in any system is a hard problem. For instance, a longstanding principle is that of *secure distributed composition* which states that when multiple sub-systems or components are composed, the resulting system does not weaken the security policies enforced by its components. Security policy enforcement approaches typically take an organisation- or network-centric view of security, e.g., [10, 26]. These tend to be either obligation-driven or authorisation-driven [26]. In the former case, policies are enforced actions in response to particular events or stimuli within a system while, in the latter, they provide access control rules specifying whether a particular subject can legitimately access (or not) a particular object. Such approaches assume that the

system, whether distributed or not, is within a single administrative control and even where platform or geographical boundaries are crossed, this happens within the control of a single organisation or a federated security management framework [8]. This is not the case for digital infrastructures under discussion in this paper, which are globally interconnected open-ended networked environments.

The challenge is further compounded by the cyber-physical nature of many constituent systems where legacy hardware and software are abound and security assurances can vary widely—from poorly designed network protocol stacks to access control models that do not enforce privileges at suitable levels. Furthermore, such environments are not static. Devices, systems and services can dynamically (and, increasingly, automatically) compose based on context and locality. Human actors are integral to the dynamics, and often catalyse dynamic composition and delivery of services, e.g., through wearables that bridge multiple systems simultaneously. Consequently, security orchestration can be, at best, delivered through service-level agreements (SLAs). However, violation of such SLAs is often only detected post-hoc. Furthermore, in a large set of scenarios, e.g., those involving untrusted or partially-trusted third party systems, specification, agreement and enforcement of an SLA is impossible.

2.3 Continual Assurance

For well-structured systems (e.g., control systems, database/transactional systems) with clearly specified security requirements on a) interactions and dependencies across sub-systems, services and components, and b) the expected threats, research has developed a variety of sophisticated capabilities to monitor and analyse their security posture to assert (with varying levels of confidence and accuracy) the requisite levels of security assurances [1, 14, 19]. This is not the case for globally interconnected open-ended networked heterogeneous environments where a complete awareness of all dependencies and knowledge of all operational paths is not viable. This becomes even more challenging in an ultra-large scale environment where conjunctions of secure and unsecure, trusted and untrusted, and reliable and unreliable elements are present.

For instance, for complex and dynamically interconnected systems, the consequent lack of a) complete and stable system and security specifications including the threats, and b) complete and stable dependency and interface specifications, make provisioning of continual assurance a challenge. Such systems are typically heterogeneous couplings of structured, unstructured, synchronous and asynchronous elements and services. This precludes a single system model invariably considered in state-of-the-practice/art approaches [25].

2.4 Incident Response

Over the past 20 years significant progress has been made to mature and develop incident response and recovery capacity, whether delivered by in-house security operations centres (SOCs) or by third party managed service providers. This is supported by automation and tooling, often in the form of Security Information

and Event Management (SIEM) systems that provide real-time information to human operators in a SOC. However, selecting the best response and recovery actions remains a largely human task [2]. Orchestrating incident response on an infrastructure-scale requires research into the appropriate balance between human-machine decision-making.

Existing standards such as ISO/IEC 27035-2:2023 [16] offer guidelines on how to plan, prepare and learn lessons from any incidents, both in terms of system defences and the incident response approach. Given the high-level nature of such guidance, operationalisation happens through *playbooks*, acting as recipes on steps and actions to take during incident response. However, playbooks remain very much a manual setup, often taking the format of natural language texts or flow charts—typically in printed format placed in SOCs. Recent works have argued for more systematic model-based representations of playbooks [24], and have highlighted the lack of a) usability studies of playbooks, and b) specificity even for highly rated playbooks for completeness and correctness by experts [27].

In the infrastructures under discussion, each constituent system will have its own playbook unlikely to be formalised into any structured or systematic common model [24]. Orchestrating a globally coordinated incident response on this scale is, therefore, a major research challenge. It is made even more challenging by the dynamism—systems composing with the infrastructure or leaving. Furthermore, constituent systems' playbooks will change in response to incidents over time. So one cannot start from the assumption that the playbooks are convergent or will remain so over time. The complexity is further compounded because contextual information is a challenge in SIEMs as SOC workers are not involved in the design choices, configurations and operation of specific organisational assets from where telemetry is fed into the SOC. Where contextual information is communicated, this happens informally and thus remains tacit and not formally documented [3].

3 In Conclusion

Advancing the paradigm of *securing-a-compromised-system* will require a *systems* approach that addresses the aforementioned four dimensions. We need new ways to elicit, specify, and validate security assurances for service composition in the presence of uncertainty, dynamism, and human behaviour. New mechanisms to compose and orchestrate security provision across diverse and heterogeneous evolving infrastructures with legacy and non-legacy elements will be critical in this regard. Alongside, it is paramount that the research community develops ways to reason about the security state at runtime in order to provide continuity of oversight and trust in the presence of partially trusted, under attack, vulnerable, or compromised elements. Last, but by no means least, it is essential that we address how we may orchestrate incident response that accounts for heterogeneous incident response practices in constituent systems and provides situational awareness at the necessary pace and resolution for optimal human-machine decision-making.

Acknowledgments. This research is supported by the Engineering and Physical Sciences Research Council grant SCULI: Securing Convergent Ultra-large Scale Infrastructures [EP/Z531315/1].

References

1. Ayodeji, A., Mohamed, M., Li, L., Di Buono, A., Pierce, I., Ahmed, H.: Cyber security in the nuclear industry: A closer look at digital control systems, networks and human factors. Progress in Nuclear Energy **161**, 104738 (2023). https://doi.org/10.1016/j.pnucene.2023.104738
2. Bada, M., Creese, S., Goldsmith, M., Mitchell, C., Phillips, E.: Computer security incident response teams (CSIRTs): An overview. The Global Cyber Security Capacity Centre (2014)
3. Bhatt, S.N., Manadhata, P.K., Zomlot, L.: The Operational Role of Security Information and Event Management Systems. IEEE Secur. Priv. **12**(5), 35–41 (2014). https://doi.org/10.1109/MSP.2014.103
4. Bradbury, M., Jhumka, A., Watson, T.: Trust Trackers for Computation Offloading in Edge-Based IoT Networks. In: 40th IEEE Conference on Computer Communications, INFOCOM 2021, Vancouver, BC, Canada, May 10-13, 2021. pp. 1–10. IEEE (2021). https://doi.org/10.1109/INFOCOM42981.2021.9488844
5. Chew, E., Swanson, M., Stine, K., Bartol, N., Brown, A., Robinson, W.: Performance Measurement Guide for Information Security (Jul 2008). https://doi.org/10.6028/NIST.SP.800-55r1, NIST SP 800-55 Rev. 1
6. Chowdhury, P.D., Renaud, K.V., Rashid, A.: When Data Breaches Happen, Where Does the Buck Stop ... and Where Should it Stop? In: Proceedings of New Security Paradigms Workshop (NSPW) (2024). https://doi.org/10.1145/3703465.3703474
7. Chowdhury, P.D., Sameen, M., Blessing, J., Boucher, N., Gardiner, J., Burrows, T., Anderson, R.J., Rashid, A.: Threat Models over Space and Time: A Case Study of E2EE Messaging Applications. Software: Practice & Experience **54**, 2316–2335 (2024)
8. Decat, M., Lagaisse, B., Joosen, W.: Middleware for efficient and confidentiality-aware federation of access control policies. Journal of Internet Services and Applications **5**(1) (2014). https://doi.org/10.1186/1869-0238-5-1
9. Digital Security by Design (DSbD) (2024), `https://www.dsbd.tech/`, Accessed On: 2024-11-15
10. Hadjiantonis, A.M., Charalambides, M., Pavlou, G.: A Policy-Based Approach for Managing Ubiquitous Networks in Urban Spaces. In: IEEE International Conference on Communications. pp. 2089–2096 (2007). https://doi.org/10.1109/ICC.2007.346
11. Hayden, L.: IT Security Metrics : A Practical Framework for Measuring Security & Protecting Data. McGraw Hill (2011)
12. Herrmann, D.S.: Complete guide to security and privacy metrics: measuring regulatory compliance, operational resilience, and ROI. CRC Press (2007)
13. HP Wolf Security: Threat Insights Report: September 2024 (Sep 2024), `https://threatresearch.ext.hp.com/hp-wolf-security-threat-insights-report-september-2024/`, Accessed On: 2024-11-15
14. Hudic, A., Smith, P., Weippl, E.R.: Security assurance assessment methodology for hybrid clouds. Computers & Security **70**, 723–743 (2017). https://doi.org/10.1016/j.cose.2017.03.009

15. ISO/IEC JTC 1/SC 27: Information technology — Security techniques — Information security management — Monitoring, measurement, analysis and evaluation. Standard, ISO/IEC (2016), https://www.iso.org/standard/64120.html, ISO/IEC 27004:2016, Edition 2
16. ISO/IEC JTC 1/SC 27: Information technology — Information security incident management — Part 2: Guidelines to plan and prepare for incident response. Standard, ISO/IEC (2023), https://www.iso.org/standard/78974.html, ISO/IEC 27035-2:2023
17. Jaquith, A.: Security Metrics: Replacing Fear, Uncertainty, and Doubt. Addison-Wesley Professional, 1st edn. (Mar 2007)
18. Kirlappos, I., Parkin, S.E., Sasse, M.A.: "Shadow security" as a tool for the learning organization. SIGCAS Comput. Soc. **45**(1), 29–37 (2015). https://doi.org/10.1145/2738210.2738216
19. Macaulay, T., Singer, B.L.: Cybersecurity for industrial control systems: SCADA, DCS, PLC, HMI, and SIS. CRC Press (2011)
20. Rashid, A.: Developer-Centred Security. Springer (2021). https://doi.org/10.1007/978-3-642-27739-9_1578-1
21. Ross, R., Pillitteri, V., Graubart, R., Bodeau, D., Mcquaid, R.: Developing Cyber-Resilient Systems: A Systems Security Engineering Approach (Dec 2021), https://doi.org/10.6028/NIST.SP.800-160v2r1, NIST Special Publication 800-160, Volume 2, Revision 1
22. Ross, R., Winstead, M., McEvilley, M.: Engineering Trustworthy Secure Systems (Nov 2022), https://doi.org/10.6028/NIST.SP.800-160v1r1, NIST Special Publication (SP) NIST SP 800-160v1r1
23. Sasse, M.A., Rashid, A.: The Cyber Security Body of Knowledge v1.1.0, 2021, chap. Human Factors. University of Bristol (2021), https://www.cybok.org/, KA Version 1.0.1
24. Shaked, A., Cherdantseva, Y., Burnap, P.: Model-Based Incident Response Playbooks. In: ARES 2022: The 17th International Conference on Availability, Reliability and Security, Vienna, Austria, August 23–26, 2022. pp. 26:1–26:7. ACM (2022). https://doi.org/10.1145/3538969.3538976
25. Shukla, A., Katt, B., Nweke, L.O., Yeng, P.K., Weldehawaryat, G.K.: System security assurance: A systematic literature review. Computer Science Review **45**, 100496 (Aug 2022). https://doi.org/10.1016/j.cosrev.2022.100496
26. Sloman, M.: Policy Driven Management for Distributed Systems. Journal of Network and Systems Management **2**(4), 333–360 (1994). https://doi.org/10.1007/BF02283186
27. Stevens, R., Votipka, D., Dykstra, J., Tomlinson, F., Quartararo, E., Ahern, C., Mazurek, M.L.: How Ready is Your Ready? Assessing the Usability of Incident Response Playbook Frameworks. In: CHI '22: CHI Conference on Human Factors in Computing Systems, New Orleans, LA, USA, 29 April 2022 - 5 May 2022. pp. 589:1–589:18. ACM (2022). https://doi.org/10.1145/3491102.3517559
28. Ullah, S., Rashid, A.: Porting to Morello: An In-depth Study on Compiler Behaviors, CERT Guideline Violations, and Security Implications. In: 9th IEEE European Symposium on Security and Privacy, EuroS&P 2024, Vienna, Austria, July 8-12, 2024. pp. 381–397. IEEE (2024). https://doi.org/10.1109/EuroSP60621.2024.00028
29. Wu, Y., Skipper, G., Cui, A.: Cryo-Mechanical RAM Content Extraction Against Modern Embedded Systems. In: 2023 IEEE Security and Privacy Workshops (SPW), San Francisco, CA, USA, May 25, 2023. pp. 273–284. IEEE (2023). https://doi.org/10.1109/SPW59333.2023.00030