Towards a Resilience Metric Framework for Cyber-Physical Systems

Ivo Friedberg
Queen's University Belfast
AIT Austrian Institute of Technology
ifriedberg01@qub.ac.uk

Kieran McLaughlin
Queen's University Belfast
kieran.mclaughlin@qub.ac.uk

Paul Smith, Markus Wurzenberger
AIT Austrian Institute of Technology
first.last@ait.ac.at

Resilience is widely accepted by research communities, industry and politics as a desirable system property for cyber-physical systems. However, there exist no scalable and flexible metrics for resilience that are specific to cyber-physical systems (CPS) and consider the multi-dimensional nature of performance in these systems. In this work, we present first results towards a framework to design such a resilience metric. The key values of this framework are threefold: First, it allows to evaluate resilience with respect to different performance dimensions. Further, unnecessary complexities of the system can be deliberately left out of the metric to answer specific questions about the system. Finally, it supports the identification of causes for low resilience to improve the system design.

resilience, metric, CPS, cyber-physical system, recovery potential, absorbing potential

1. INTRODUCTION

To design and evaluate the resilience of cyber-physical systems, a scalable, flexible and computable metric for resilience is needed. A metric is scalable, if its complexity can be adapted to current needs, i.e., we can abstract aspects that are not of interest. Various performance measures can be used to evaluate a CPS. A flexible metric is one that can be used to analyse a system with respect to different performance measures, while interdependencies between performance measures and external challenges are considered. To design such a metric it is necessary to understand the design principles, domains and the interplay between these domains for CPS. The dependencies between performance measures are critical, as they can be used to identify the causes for decreased resilience and to motivate changes in the system. Current work does not provide such a metric.

Arghandeh et al. (2016) provide a detailed definition of resilience for the power system domain that is widely applicable for CPS in general. Their work explicitly excludes the design of a resilience metric. We use their work as a basis for our framework. A metric is provided by Linkov et al. (2013) that describes resilience in four dimensions on a policy level, which does not capture the runtime performance of a system. In work by Watson et al. (2015), the authors identify a strong correlation between resilience and the probability and impact of adverse incidents on performance. According to their work, a system is more resilient if the probability of adverse incidents and/or their impact is reduced. While the correlation between resilience and risk is also defined in other work (Arghandeh et al. 2016), it again ignores the temporal dimension of resilience. Another set of resilience metrics is given in Wei and Ji (2010) that observes the temporal aspect of system performance. However, their work defines performance as a one-dimensional property over time. This simplification limits the flexibility of the metric and it is therefore not possible to identify the system's shortcomings based on the metric results. The closest to a scalable and flexible metric is the work by Rieger (2014) and Eshghi et al. (2015). Their metric considers the CPS domains, analogous to Arghandeh et al. (2016). In Rieger (2014) resilience is considered with respect to control stability. The author uses control response and stability as performance measure. Similarly, Eshghi et al. (2015) models a system as a hierarchical set of controllers that are then analysed bottom-up to retrieve system resilience. It is questionable, however, how generally applicable control stability is as a performance measure and therefore how flexible this metric is.

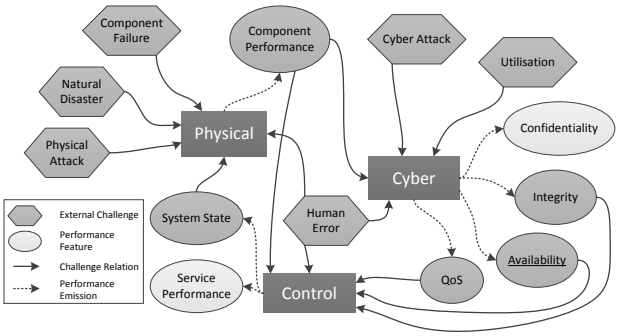


Figure 1: *Challenge-performance relationship diagram with respect to CPS domains.*

In this work, we present a novel resilience metric framework. This framework aims to guide the design of flexible and scalable resilience metrics through three key properties: First, it allows to evaluate resilience with respect to different performance dimensions (e.g., monetary loss as well as system availability). Second, unnecessary complexities of the system can be deliberately ignored or simplified in the metric, to answer specific questions about the system. For example, while the weather is a challenge to some performance measures in a power grid, it is probably not directly relevant for its resilience against cyber-attacks. Through the ability to ignore this unnecessary complexity of the system, the framework becomes more scalable. Finally, the framework supports the identification of causes for low resilience, which can be used to improve the system design.

2. SYSTEM DESCRIPTION

To derive a mathematical model for CPS resilience, a common system understanding needs to be established. In the presented framework, resilience is considered with respect to system performance. System performance is comprised of different performance features that are influenced by each other and by external challenges. Each performance feature is the property of one domain of the overall CPS. We identify our system model based on a three step approach. The high level results for a generic CPS are shown in Fig. 1.

1. *Domain Identification:* CPS, while single complex systems, are comprised of different domains that interact to provide a service. In the same way as Arghandeh et al. (2016), we have identified three domains: the physical, the cyber and the control (cyber-physical in Arghandeh et al. (2016)) domains.

2. *Feature Extraction:* Each domain contributes a set of performance features. To produce these features, the domain leverages capabilities of related domains. Fig. 1 highlights a set of performance features for each domain. These features are generic and need to be refined for each concrete system and analysis scope, in order to retrieve a meaningful and measurable representation of that system.

3. *Impact Correlation:* Each domain is affected by challenges that can be internal or external to the system. External challenges originate outside the scope of the CPS under evaluation (e.g., weather conditions). Internal challenges originate inside the system, and manifest as a performance degradation of a performance feature. Thus, it is first necessary to identify external challenges. Then, for each domain, the challenges (both internal and external) that affect the performance of the respective domain are identified.

The representation in Fig. 1 describes the complexity of performance and provides a starting point towards a meaningful metric. For example, a distributed denial of service attack is a challenge to the *cyber* domain. Subsequently the Quality of Service (QoS) of the communication will decrease. This degraded performance is not necessarily the performance we want to evaluate. However, the limited throughput is a challenge to the *control* domain, as required feedback for control decisions can be delayed. On this level the original challenge, the attack, is abstracted by a performance feature from the *cyber* domain.

3. RESILIENCE METRIC FRAMEWORK

Due to the multi-dimensional nature of performance in CPS, there needs to be a clear decision with respect to what should be measured. Each performance feature represents one dimension of performance, and depends on other performance features, as well as external challenges. Given that each performance feature has a nominal performance p_N and is described by a function $p(t)$ with respect to time, we can measure resilience $\mathcal{R}$ as the area between the actual performance and the nominal performance. This relation is described by Eq. 1. It results in a single numerical value between 0 (no resilience) and 1 (perfect resilience).

$$\mathcal{R} : \mathbb{R}^+ \to [0;1] : t \mapsto 1 - \frac{1}{(t-t_0)p_N} \cdot \int_{t_0}^{t} p(\tau)\mathrm{d}\tau \quad (1)$$

This metric can be used directly for a running system, if $p(t)$ is measurable. However, on its own, it does not allow to draw any conclusions about the causes of the performance decrease. To enable a meaningful use of resilience, the metric needs to model the influences on performance to identify causes. According to Arghandeh et al. (2016),

resilience in a system is rooted in two potentials. The *absorbing potential* is the degree in which challenges can be handled without performance degradation. The *recovery potential* describes a system's ability to restore normal operation in the face of challenges. The change in performance can then be described with respect to these two potentials, as well as the severity of challenges. Equation 2 provides a general representation of this differential equation, which is solvable as an initial value problem (IVP), where $p(t_0) = p_0$.

$$\dot{p}(t) = [f(t, r, p(t), p_N) - g(t, \vec{a}, \vec{c}(t), p(t))] \cdot \Theta(p(t)) \tag{2}$$

Here, f represents the recovery potential that has a positive impact on performance. It depends on the time t, a recovery rate r that needs to be identified for each system and can be a scalar or a more complex function, the current performance $p(t)$ and the nominal performance p_N which can never be exceeded. On the other hand, g represents the negative impact of challenges on performance and is analogous to Arghandeh's absorbing potential. It depends on the time, a set of challenges $\vec{c}(t)$ and the current performance. Further, $\vec{a}$ describes normative factors for each challenge. They need to be defined for each challenge and makes them comparable. Further, they represent the impact of each challenge on performance. Finally, $\Theta(p(t))$ is a heaviside function that describes the performance threshold p_T under which the system is considered collapsed. Once collapsed, a system has a recovery potential of 0 and cannot restore performance. The function is defined by Eq. 3.

$$\Theta : \mathbb{R}^+ \to \{0, 1\} : t \mapsto \begin{cases} 1, & p(t) > p_T \\ 0, & p(t) \leq p_T \end{cases} \tag{3}$$

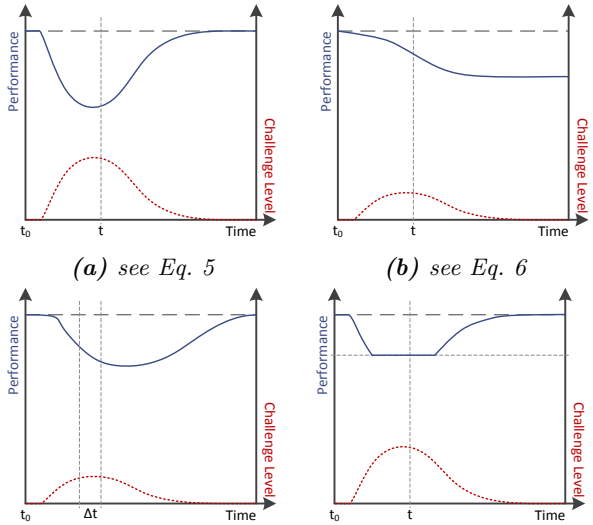
An example of a concrete equation for the recovery potential can be found in Eq. 4. It describes a recovery process that has the nominal performance as an equilibrium solution. The recovery speed depends on the system specific recovery rate r.

$$f(t, r, p(t), p_N) = r \cdot (1 - \frac{p(t)}{p_N}) \cdot p(t) \tag{4}$$

As shown by Fig. 1, each domain is exposed to a set of challenges. These challenges build $\vec{c}(t)$ and can be external or internal. To get a better understanding of their impact on performance (formally described by g), four basic types of challenges can be identified. Equations 5 - 8 describe these basic types mathematically; a visualisation and description is given by Fig. 2.

$$g(t, a, c(t)) = a \cdot c(t) \tag{5}$$

$$g(t, a, c(t)) = a \cdot \int_{t_0}^{t} c(\tau)\ d\tau \tag{6}$$



(a) *see Eq. 5* ***(b)*** *see Eq. 6*

(c) *see Eq. 7* ***(d)*** *see Eq. 8*

Figure 2: *Graphical representation of performance behaviour with respect to the four different challenge types given in Eq. 5 - 8. Without loss of generality, we assume a smooth change in challenge levels for better visibility. More specifically, challenges and also performance in CPS can change abruptly. These changes are possible with the use of the mathematical representation, even if not shown here.*

$$g(t, a, c(t)) = a \cdot \int_{t-\Delta t}^{t} c(\tau)\ d\tau \tag{7}$$

$$g(t, a, c(t), p(t)) = \begin{cases} a \cdot c(t), & p(t) > p_T \\ 0, & p(t) \leq p_T \end{cases} \tag{8}$$

To describe the different challenge types in more detail, an example from the smart grid domain will be provided for each type. Note, that a meaningful example cannot be given by a challenge alone. It needs to describe the relationship between a challenge and a specific performance feature that is affected by the challenge. In the first case (see Eq. 5 and Fig. 2a), the decrease of performance is immediate and proportional to the severity of the challenge. An example for this type is given by the relationship between the amount of power produced by a solar panel and the coverage of the sun. If the sunlight is blocked, the energy production decreases immediately. In the same way, the performance will increase again, as the amount of sunlight reaching the panel increases. Equation 6 and Fig. 2b describe the case where the impact of a challenge on performance stays present even after the challenge has abated. A challenge can manifest as a quick short burst, with a long lasting impact. For example, an operator is interested in the number of components in the distribution system that are operational to ensure that the N-1 criterion is not violated. Wind is a challenge to these components. With increasing windspeeds, a tree might fall on a power line and

ground that line. The performance measure (number of operational components) is reduced through the challenge, even if the windspeed decreases afterwards. The third type (see Eq. 7 and Fig. 2c) describes challenge performance relationships where the impact depends on the recent challenge history. In contrast to type two, the impact will fade away over time, however, in contrast to type one it is not only dependent on the current challenge level. An example is a controller that minimizes the harmonics in the distributed power signal. It is challenged by an integrity attack on the feedback value. As a consequence, the controller will introduce harmonics to the system. Once the attack is mitigated and the controller regains state awareness it will work to minimize the introduced harmonics again. However, this will take time depending on the degree of harmonics in the system after the attack. The final case is described by Eq. 8 and Fig. 2d. The impact of a challenge on performance cannot get higher than a certain threshold. At some point, even if the challenge level further increases there is no more impact on performance. As an example, we can assume a real-time controller that imposes strict time constraints on feedback measurements. An increase in the time delay on the communication network will cause some missed windows by feedback at first. However, once all windows get missed, a further delay will not cause any further performance decrease.

From these four types, some conclusions can be drawn. Each challenge can be modeled by a single type or by a combination of multiple types. Then, g is the sum of negative impacts from each challenge on performance. In the case that a challenge is internal, the performance decrease of one performance feature is a challenge to other performance features. Equation 9 provides a generic formula to convert a performance decrease into an increase in challenge level $p(t)$ with respect to the nominal performance p_N. Here, x describes the normalising factor to match the performance unit to the other challenge units.

$$\tau : \mathbb{R}^+ \to \mathbb{R}^+ : t \mapsto \frac{1}{x} \cdot (p_N - p(t)) \tag{9}$$

4. CONCLUSION

This paper presents a scalable and flexible framework to design a numerical resilience metric for CPS; something that is needed to draw meaningful conclusions about a system from its resilience. The framework design is based on a recent definition of resilience by Arghandeh et al. (2016) and is an improvement over future work by Rieger (2014) and Eshghi et al. (2015). It supports the design of a resilience metric that can evaluate a system with respect to different dimensions of performance that we call performance features. By doing so, a more flexible metric can be built because the scope of the evaluation goal can be better represented by the metric. A metric designed with this framework is not only describing a specific performance feature. It also models the dependency between the measured performance and external as well as internal challenges. That way the metrics can be used to reason about the causes for decreased resilience and subsequently to improve system design. Finally, the framework encourages the metric designer to develop a detailed system understanding. With this understanding, the designed metrics can be more flexible, as it allows informed and deliberate decisions about simplifications to the metrics. This is important, as models of CPS should be meaningful abstractions to handle the system's complexity. In future work, we will use this framework to develop a resilience metric for a concrete system in the smart grid domain.

Acknowledgements
This work was partly funded by the EU FP7 SPARKS project (Contract No. 608224) and the EPSRC CAPRICA (Contract No. EP/M002837/1) project.

REFERENCES

Arghandeh, R., von Meier, A., Mehrmanesh, L., and Mili, L. (2016). On the definition of cyber-physical resilience in power systems. *Renewable and Sustainable Energy Reviews*, 58:1060–1069.

Eshghi, K., Johnson, B. K., and Rieger, C. G. (2015). Power system protection and resilient metrics. In *Resilience Week (RWS), 2015*, pages 1–8.

Linkov, I., Eisenberg, D. A., Plourde, K., Seager, T. P., Allen, J., and Kott, A. (2013). Resilience metrics for cyber systems. *Environment Systems and Decisions*, 33(4):471–476.

Rieger, C. G. (2014). Resilient control systems Practical metrics basis for defining mission impact. In *Resilient Control Systems (ISRCS), 2014 7th International Symposium on*, pages 1–10.

Watson, J.-P., Guttromson, R., Silva-Monroy, C., Jeffers, R., Jones, K., Ellison, J., Rath, C., Gearhart, J., Jones, D., Corbet, T., Hanley, C., and Walker, L. T. (2015). Conceptual framework for developing resilience metrics for the electricity, oil, and gas sectors in the united states. Technical report, Sandia National Laboratories.

Wei, D. and Ji, K. (2010). Resilient industrial control system (RICS): Concepts, formulation, metrics, and insights. In *Resilient Control Systems (ISRCS), 2010 3rd International Symposium on*, pages 15–22.