DOI: http://dx.doi.org/10.14236/ewic/ICS2016.3

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 as a desirable system property for cyber-physical systems. However, there are no metrics that can be used to measure the resilience of cyber-physical systems (CPS) while the multi-dimensional nature of performance in these systems is considered. In this work, we present first results towards a resilience metric framework. The key contributions of this framework are threefold: First, it allows to evaluate resilience with respect to different performance indicators that are of interest. Second, complexities that are relevant to the performance indicators of interest, can be intentionally abstracted. Third and final, it supports the identification of reasons for good or bad resilience to improve system design.

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

1. INTRODUCTION

Resilience is widely accepted as a desirable system property for cyber-physical systems. To evaluate the resilience of cyber-physical systems (CPS) – namely, the behaviour of system performance over time in the presence of challenges – a computable metric is needed. System performance is not a one-dimensional property for CPS – numerous properties can be measured and distinct stakeholders are interested in different measures. Therefore, a metric for assessing CPS resilience needs to consider the multi-dimensional nature of system performance, and take into account the interplay of different measures.

To do this, a resilience metric needs to be embedded in a well-designed framework that represents the system in the metric for analysis. The complexity of the framework depends on two factors. First, on number and nature of the performance measures that are considered relevant. Second, on the complexity of the underlying system. The framework has to be flexible enough to provide a sensible system abstraction. It needs to allow the metric to be applied to different performance aspects without ignoring the multi-dimensional complexity of performance. To manage the system complexity, the framework needs to allow the analyst to scale the system representation by abstracting unnecessary aspects. In the example above, it might be irrelevant (or deliberately ignored) how earthquakes affect system performance.

In current work such a metric is not provided. Arghandeh et al. (2016) provide a detailed definition of resilience for the power system domain that is widely applicable for CPS in general and is used in this work. Their work explicitly excludes the design of a resilience metric. 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 ignores the temporal dimension of resilience. Another set of resilience metrics is given in Wei and Ji (2010) that observe 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 more difficult to identify the system's shortcomings based on the metric results.

Closest to an applicable metric is the work by Rieger (2014) and Eshghi et al. (2015). Similar to Arghandeh et al. (2016), their metric considers three domains in CPS: physical, cyber and cognitive domain. In Rieger (2014) resilience is considered with respect to control stability. The author uses control response and stability as performance

measures. Similarly, Eshghi et al. (2015) model a system as a hierarchical set of controllers that are then analysed bottom-up to retrieve system resilience. It is questionable, how generally applicable control stability is as a performance measure and therefore how flexible this metric is.

In this work, we present a novel resilience metric framework that has three key properties: First, it allows to evaluate resilience with respect to different performance indicators (e.g., monetary loss or system availability). Second, unnecessary aspects of the system can be deliberately abstracted, to manage system complexity. Finally, the framework helps to identify system aspects that increase or decrease resilience to further improve the system design.

2. SYSTEM DESCRIPTION

To derive a computable metric framework, a common system understanding needs to be established. In the presented metric, resilience is described as system performance over time. 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 accordance to Arghandeh et al. (2016), we identified a *cyber* and a *physical* domain. However, for the conjunctive domain the term *control* is chosen over cyber-physical as it is considered to be more expressive and accurate.

2. *Feature Extraction:* Each domain contributes a set of functions that can be measured through a set of performance features. Fig. 1 highlights a generic set of performance features for each domain. Further, each domain leverages functions of other domains which leads to interdependencies between domains.

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 manifest as a performance degradation of another performance features. 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.

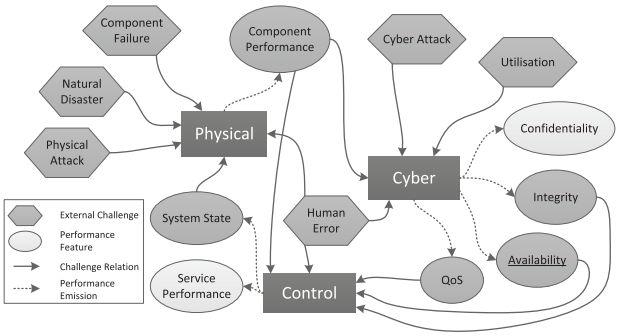


Figure 1: *Challenge-performance relationship diagram with respect to CPS domains.*

The representation in Fig. 1 describes the complexity of performance and provides a starting point towards a sensible framework. 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. Thus the original challenge, the attack, is abstracted by a performance feature from the *cyber* domain.

3. RESILIENCE METRIC FRAMEWORK

In the presented resilience metric framework, resilience is computed for each performance feature separately. Given that each performance feature has a nominal performance p_N – by this we mean the performance of the system in a challenge free environment – and is described by a function $p(t)$ ($0 \leq p(t) \leq p_N$) 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).

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

This is the metric and it 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 low resilience and is not superior to existing work. The underlying framework models each performance feature as one dimension of performance that depends on other performance features, as well as external challenges. In the framework the results from the metric can be rooted in the complete system. The framework can further be used to make predictions and estimates about

the resilience of the system without applying real challenges at runtime. 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 adapt in response to challenges to restore normal operation during the challenge is present or after the challenge is abated. Therefore, the framework models the system by the change in performance features based on these two potentials (see Eq. 2). This results in a differential equation for each performance feature that 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 of a degraded system. It depends on the time t, a recovery rate r which needs to be identified for each system and can be a constant or a function, the current performance $p(t)$ and the nominal performance p_N. On the other hand, g represents the impact that challenges have on performance. It is analogous to Arghandeh's absorbing potential and depends on the time, a set of challenges $\vec{c}(t)$ and the current performance. To make challenges comparable and to model the severity of each challenge on performance, $\vec{a}$ describes normative factors for each challenge. 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. This ensures that a collapsing system is considered not resilient by the metric. The function is defined by Eq. 3.

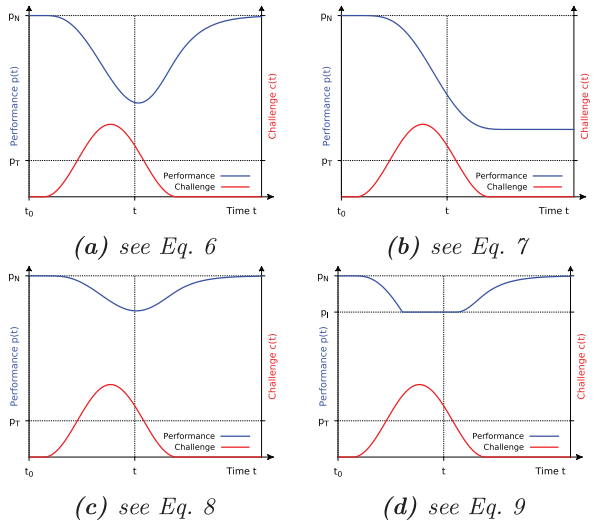
$$\Theta : \mathbb{R}^+ \to \{0, 1\} : t \mapsto \begin{cases} 1, & p(t) > p_T \\ 0, & p(t) \leq p_T \end{cases} \tag{3}$$

An example 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. In the case that a challenge is internal, the performance decrease of one performance feature is a challenge to other performance features. Equation 5 provides a generic formula for the conversion with x as the normalising factor to match the performance unit to the other challenge units.

$$\tau : \mathbb{R}^+ \to \mathbb{R}^+ : t \mapsto \frac{1}{x} \cdot (p_N - p(t)) \tag{5}$$



(a) see Eq. 6 *(b) see Eq. 7*

(c) see Eq. 8 *(d) see Eq. 9*

Figure 2: *Graphical representation of performance behaviour under different challenge types given in Eqs. 6 - 9 and a recovery rate r of 1 (see Eq. 4). Without loss of generality, we assume a smooth change in challenge levels.*

The greatest difficulty in the design of the framework is to derive the impact each challenge has on the various performance features. Here system complexity can be scaled by ignoring challenges deliberately. To guide this process, we identify four basic challenge types. Although these types cannot model every possible challenge, they provide an understanding of the most common challenges; they can be modeled by a single type or by a combination of multiple types. Equations 6 - 9 provide a mathematical definition of the types; a visualisation is provided in Fig. 2.

$$g(t, a, c(t)) = a \cdot c(t) \tag{6}$$

$$g(t, a, c(t)) = a \cdot \int_{t_0}^{t} c(\tau)\, d\tau \tag{7}$$

$$g(t, a, c(t)) = a \cdot \int_{t-\Delta t}^{t} c(\tau)\, d\tau \tag{8}$$

$$g(t, a, c(t), p(t)) = \begin{cases} a \cdot c(t), & p(t) > p_l \\ 0, & p(t) \leq p_l \end{cases} \tag{9}$$

To describe the different challenge types in more detail, an example from the smart grid domain will be provided for each type. In the first case (see Eq. 6 and Fig. 2), 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 7 and Fig. 2 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. 8 and Fig. 2) 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. 9 and Fig. 2. The impact of a challenge on performance cannot get higher than a certain threshold p_l (with $p_l > p_T$). 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 feedback measurements to miss their time window. However, once all measurements miss their windows, a further delay will not cause any further performance decrease.

Equations 2 to 9 provide the mathematical tools to design the framework for a specific system instance. Then, Eq. 1 can be used to measure, predict and estimate the system resilience with respect to all relevant performance indicators.

4. DISCUSSION AND CONCLUSION

The presented metric framework aims to provide a scalable system model and a flexible way to measure system resilience numerically. However, it does not come without challenges and limitations. First, the approach aims to describe a single system; something that is not easy to define in CPS. For example, how much of the Internet's infrastructure is part of a smart grid installation that leverages the Internet to transport smart metering data? Another challenge lies in the implementation of the framework and its evaluation. The type of information needed to establish the interdependencies between different performance features, is often not available for existing CPS installations. Therefore, this knowledge currently has to be established by the analyst manually by measuring system behaviour, or through the analysis of incident records. This is not only difficult and time consuming; the results are also hard to evaluate. However, the proposed solution surpasses previous work because it was designed to tackle these problems. The multi-dimensional performance concept makes the framework more generally applicable than work by Rieger (2014), which is focused on control response as performance measure. This allows a broader understanding of CPS in the context of resilience. Further, the complexity of the framework can be limited by a deliberate choice about the challenges to consider and those to ignore for the current analysis. It is one goal of future research to analyse the complexity of this task, and to provide guidance and support to analysts. Finally, it is planned to instantiate the metric framework in a concrete system from 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. et al. (2016). On the definition of cyber-physical resilience in power systems. *Renewable and Sustainable Energy Reviews*, 58. 1060–1069.

Eshghi, K. et al. (2015). Power system protection and resilient metrics. In *Resilience Week (RWS)*, 1–8.

Linkov, I. et al. (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*, 1–10.

Watson, J.-P. et al. (2015). Conceptual framework for developing resilience metrics for the electricity, oil, and gas sectors in the United States. Sandia National Laboratories. Tech. Rep.

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*, 15–22.