**Proceedings of the ASME 2016 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference**
**IDETC/CIE 2016**
**August 21-24, 2016, Charlotte, North Carolina, USA**

**DETC2016-59426**

# SYSTEM RESILIENCE QUANTIFICATION FOR PROBABILISTIC DESIGN OF INTERNET-OF-THINGS ARCHITECTURE

**Yan Wang**
**School of Mechanical Engineering**
**Georgia Institute of Technology**
**Atlanta, GA 30332**

## ABSTRACT

*The objects in the Internet of Things (IoT) form a virtual space of information gathering and sharing through the networks. Designing IoT-compatible products that have the capabilities of data collection, processing, and communication requires open and resilient architecture with flexibility and adapability for dynamically evolving networks. Design for connectivity becomes an important subject in designing such products. To enable a resilience engineering approach for IoT systems design, quantitative measures of resilience are needed for analysis and optimization. In this paper, an approach for probabilistic design of IoT system architecture is proposed, where resilience is quantified with entropy and mutual information associated with the probabilities of detection, prediction, and communication among IoT-compatible products. Information fusion rules and sensitivities are also studied.*

## 1. INTRODUCTION

Internet of Things (IoT) refers to uniquely identifiable physical objects that form an Internet-like structure in cyber space [1]. The original idea of IoT was to extend the capability of radio-frequency identification (RFID) chips with internet connectivity. Later, the concept was generalized to any physical objects with data collection, processing, and communication capabilities. We can imagine that in the future any object we interact with in our daily lives would probably have the functions of data collection and exchange, be it thermostat, pen, car seat, or traffic light. These objects form the so-called cyber-physical systems. All objects in the physical environment of IoT also form a virtual space of information gathering and sharing. This information can affect every decision we make daily, such as which jacket to wear, which medicine to take, which commute route to follow, etc.

According to Gartner [2], "the IoT, which excludes personal computers, tablets, and smart phones, will grow to 26 billion units in 2020, representing an almost 30-fold increase from 0.9 billion in 2009. By 2020, component costs will have come down to the point that connectivity will become a standard feature, even for processors costing less than $1." Therefore, IoT is likely to affect each industry or consumer product we produce in the near future. Designing a product that is IoT-compatible will become a common subject for design engineers, regardless product types or industry sectors.

There are some new challenges in designing IoT-compatible products. The complexity of IoT-compatible products and cyber-physical systems has increased. Designing each product requires the consideration of hardware, software, as well as network connectivity, which is beyond the existing mechatronic systems, where hardware and software are simultaneously designed but with much lower complexity. IoT-compatible products are meant to be Internet-ready. Each product is an open system that can be re-configured and re-adapted into the evolution of the Internet itself. Therefore, the concept of open system design with robust and diverse connectivity becomes important. In addition, the functions cyber-physical systems are collected efforts from individuals. The confederated systems formed by IoT-compability products do not have centralized control and monitoring units. Ad hoc networks are formed by vastly different products. The reliabilities as well as working conditions of the individual products and components can be highly diverse. Good adaptability and resilience are important in designing the architecture of such systems. Yet, different from traditional communication networks, IoT networks do not just transfer information. Each node of the networks also creates information. IoT networks are also different from traditional sensor networks, where the main task of sensors is collecting information whereas the logical reasoning for decision making is still done at centralized computers. In IoT networks, the level of computational intelligence and reasoning capability of the nodes are much higher and a major portion of decisions are done locally at individual nodes.

In this paper, resilience of IoT architecture is studied. The term resilience has been loosely used by many and is semantically overloaded. There is a lack of standard definition of what resilience is and how to measure it quantitatively for analysis. The definitions are domain dependent. Generally speaking, resilience refers to the capability of a system that can regain its function or performance after temporary degradation or breakdown. Recently researcher started looking into formal quantification of resilience. Nevertheless, how to quantify function or performance of existing systems such as communication and transportation networks still remains at a very abstract level. There is also a need of developing quantitative approaches to design emerging systems such as IoT networks. Here, a probabilistic design approach to design IoT system architecture is proposed to enable resilience engineering of the systems. A formal metric to quantify the functionality and performance of IoT systems is also proposed, which is based on entropy and mutual information associated with the detection, prediction, and communication capabilities of nodes.

In the remainder of this paper, an overview of resilience research is provided in Section 2, which includes the quantitative studies of resilience and the applications in engineering and networks. It is seen that resilience is a common and interdisciplinary subject for complex system study across many domains. Yet, the effort of quantitative analysis for resilience engineering and system design is still very limited. In Section 3, a probabilistic model for IoT architecture design and the metrics to measure system performance are proposed for resilience engineering. Based on the metrics, a formal approach to design IoT architecture and optimization with sensitivity analysis is demonstrated in Section 4.

## 2. BACKGROUND

### 2.1 The Concepts of Resilience in Various Domains

The history of systematic resilience study can be retrieved back to early 1960s by ecologists, who were interested in ecosystem stability. The researchers look at an ecosystem from the perspective of multiple time and size scales. The system thus may be stabilized at more than one stable equilibrium. In contrast, resilience traditionally studied in engineering focuses on the system behavior near a stable equilibrium and studies the rate at which a system approaches the steady state following a perturbation. The studies are about how to improve the ability to resist the change and how to reduce the time of recovery. Although the concept of resilience has appeared in the literature of various domains, such as ecology, economics, materials science, computer engineering, and computer networks, it has not been uniformly defined and characterized.

The resilience perspective emerged in ecology more than four decades ago through the study of interacting population of predators and prey in an ecosystem [3,4,5,6]. Resilience is regarded as the capacity to absorb shocks and maintain dynamic stability in the constant transient states. The accepted definition of resilience in ecology is the capacity to persist within one or several stability domains. Resilience determines the persistence of relationships within an ecosystem and is a measure of the ability of these systems to absorb changes of state variables, driving variables, and parameters, and still persist [6]. The measure of resilience is the size of stability domains, or the amount of disturbance a system can take before its controls shift to another set of variables and relationships that dominate another stability region [7]. The concept of slow and fast variables at multiple time scales is observed in ecosystems. Because of the dynamic nature of the ecosystem, the terms "regimes" and "attractors" were proposed to replace "stable states" and "equilibria" [8]. The resilience of ecosystems emphasizes not only persistent and robustness upon disturbance, but also adaptive capacity to regenerate and renew in terms of recombination and self-reorganization. Ecosystem resilience has also been proposed to be a major index of environmental sustainability during economic growth. Economic activities are sustainable only if the life-support ecosystems on which they depend are resilient [9].

The resilience of regional economics is generally considered as the capability of returning to a pre-shock state, as defined and measured by employment, output, and other variables, after disturbances or adverse events such as economic crisis, recessions, and natural disasters [10,11]. Several notions of regional resilience have been proposed. For example, Foster [12] defined regional resilience as the ability of a region to anticipate, prepare for, respond to, and recover from a disturbance. Hill et al. [13] defined it as the ability of a region to recover successfully from shocks to its economy that either throw it off its growth path or have the potential to throw it off its growth path. Yet, there is no standard and precise definition and measurement. Unlike physical or ecological systems, a regional economy may never be in an equilibrium state. It can grow continuously. Therefore, regional economics resilience emphasizes on returning to the pre-shock path or state, regardless whether it was in equilibrium or not. The four dimensions of regional resilience are: *resistance* (the vulnerability or sensitivity of a regional economy to disturbances and disruptions), *recovery* (the speed and extent to return to the pre-shock state), *re-orientation* (the adaptation and re-alignment of regional economy and its impact to the region's output, jobs, and incomes), and *renewal* (the resumption of the growth path) [11].

The term resilience has been used in materials science for decades. A material with good resilience is similar to a spring. It reacts on compression, tension, or shearing forces elastically and rebounds to its original shape. The term appeared in the literature of textile material [14,15,16] and rubber [17,18,19] as early as in 1930s. The resilience of a material is generally regarded as the energy dissipation property of storing and releasing energy elastically, and can be characterized as the ratio of energy given up in recovery from deformation to the energy applied to produce the deformation, which is measured through the energy loss during repeated load and unload cycles [19].

With the continuing downscaling of CMOS technologies and reduction of power voltage, sporadic timing errors, device degradation, and external environment radiation may cause so-called single-event transient errors in computer chips and microelectronic systems. Designers of such computing systems use resilience to describe the systems' fault tolerance [20,21,22,23]. The main approaches to enhance error resilience include error checking for recovery, co-design of hardware and software, and application-aware hardware implementation. Hardware resilience can be achieved by applying machine learning algorithms to process data collected from fault-affected hardware and perform classification for inference and decision making [24, 25]. Statistical error compensation [26] can be applied to maximize the probability of correct prediction given hardware errors.

The reliability and resilience of cyberinfrastructure and cybersecurity have been the research focus for decades [27,28]. Resilience of computer network is regarded as the ability of the network to provide and maintain an acceptable level of service in the face of various faults and challenges to normal operation [29]. The considered factors for computer network resilience include fault tolerance due to accidents, failure, and human errors; disruption tolerance due to external environment such as weather, power outage, weak connectivity, and malicious attacks; and traffic tolerance because of legitimate flash crowd or denied of service attacks. Fault tolerance typically relies on redundancy if the failures of components are independent, whereas survivability depends on diversity for correlated failures.

To improve the reliability and safety of socio-technical systems with a proactive and systems engineering approach, resilience engineering is a term people coined to promote the concept of enabling the capability of anticipating and adapting to the potential accidents and system failures [30]. It is the intrinsic ability of a system to adjust its functioning prior to, during, or following changes and disturbances, so that it can sustain required operations under both expected and unexpected conditions. The emphasized capabilities are anticipation, learning, monitoring, and responding. It is concerned with exploiting insights on failures in complex systems, organizational contributors to risk, and human performance drivers in order to develop proactive engineering practices. In resilience engineering, failure is seen as the inability to perform adaptations to cope with the dynamic conditions in real world, rather than as breakdown or malfunction [31]. The scope of systems includes both physical and human components, as human error is one of the major sources of system failures. Domain experts' over-confidence could also impede the proper development of anticipation of unexpected severe situations [32]. The important issues of resilience engineering include the dynamics and stability of complex systems.

## 2.2 Quantification of Resilience

Most of the existing studies in resilience focus on the conceptual and qualitative level of system analysis. Although various definitions of resilience have been proposed [33,34], there are limited quantification methods to measure the resilience of systems for analysis and comparison. These methods calculate resilience metrics based on the curve of recovery. The curve of recovery shows the dynamic process that the function or performance of a system degrades during a shock and recovers afterwards. The typical concepts are illustrated in Figure 1, by which Francis and Bekera [34] used to define resilience factors. In the figure, $F_o$ is the original stable system performance level, $F_d$ is the performance level immediately post-disruption, $F_r^*$ is the performance level after an initial post-disruption equilibrium state has been achieved, $F_r$ is the performance at a new stable level after recovery efforts have been exhausted, $t_\delta$ is the slack time before recovery ensues, and $t_r$ is the time to final recovery. Other researchers used the curves with minor variations, for instance, without explicit consideration of the initial post-disruption equilibrium state $F_r^*$, or the new stable state $F_r$ being the same as the original stable state $F_o$.

Several resilience metrics have been proposed. Francis and Bekera [34] proposed a resilience measurement based on the ratios between the new stable states and the original state as

$$\rho = S_p \frac{F_r}{F_o}\frac{F_d}{F_o}$$

where $S_p$ is the speed recovery factor calculated from recovery times to new equilibrium. In this metric of resilience, $F_d/F_o$ captures the absorptive capacity of the system, and $F_r/F_o$ expresses the adaptive capability. Therefore, the more functionality is retained relative to the original capacity, the higher the resilience is.

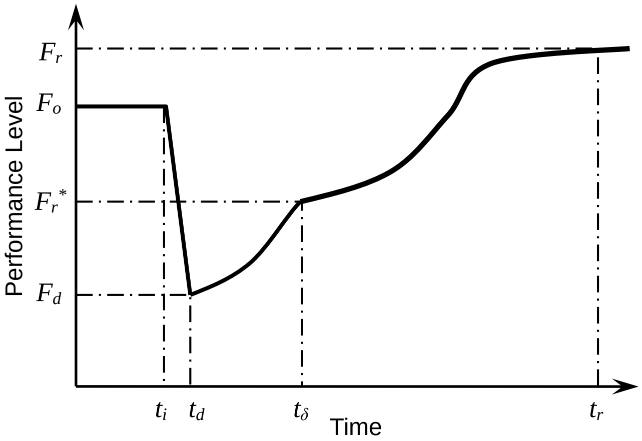


Figure 1: System performance curve used by Francis and Bekera [34].

Bruneau and Reinhorn [35,36] quantified resilience by

$$R_1 = \frac{1}{t_r - t_i}\int_{t_i}^{t_r} Q(t)dt$$

where $Q(t)$ is a dimensionless functionality function that has the value between 0 and 1, $t_i$ is the time when the adverse event occurs that causes the loss of functionality, and $t_r$ is the time of full recovery. That is, resilience is the area under the curve of performance divided by the time of duration, which is the average functionality. Among four factors of resilience that authors proposed, rapidity, robustness, resourcefulness, and redundancy, the first two are quantified. Rapidity is the slope of the functionality curve during recovery as $dQ(t)/dt$, whereas robustness is quantified as $1-L$ where $L$ is a random variable that represents the loss of functionality due to the adverse event.

Ouyang et al. [37] proposed a resilience metric based on the expected area under the performance curve as

$$R_2 = \frac{\int_{t_i}^{t_r} F(t)dt}{\int_{t_i}^{t_r} F^*(t)dt}$$

where $F$ is the performance curve as a stochastic variable, and $F^*$ is the target performance curve. The resistant, absorptive, and restorative capabilities are considered all together in the integral form.

To provide more granularity for different failure and recovery modes, Ayuub [38] proposed the metric

$$R_3 = \frac{T_d \left[\frac{\int_{t_i}^{t_d} F(t)dt}{\int_{t_i}^{t_d} F^*(t)dt}\right] + T_r \left[\frac{\int_{t_d}^{t_r} F(t)dt}{\int_{t_d}^{t_r} F^*(t)dt}\right]}{T_d + T_r}$$

where $T_d = t_d - t_i$ and $T_r = t_r - t_d$ are the disruption and recovery time periods respectively. This metric provides the additional measures of failure and recovery speeds.

### 2.3 Resilience of Communication Networks

The most relevant domain to IoT system resilience is the resilience of telecommunication networks such as Internet, wireless networks, and vehicular networks [29,39]. Resilience can be qualitatively measured in a state space formed by service parameters and operational state. The quantitative approaches measure system resilience by message delivery failure probabilities due to packet loss [40], payload error [41], or delay [ 42 ] during transmission. For topological analysis, the communication failures are quantified based on the connectivity in the Erdös-Rényi random graph [43]. Simulation models [44] have also been developed. The performance and resilience of networks are measured by packet delivery ratio [44], route diversity [45], node valence and connectivity [46,47], or quality of service [48].

Different from the above efforts which focus only on communication, the probabilistic model proposed here is to model both communication and reasoning capabilities of IoT systems, which is described in the following section.

## 3. PROBABILISTIC MODEL OF IOT ARCHITECTURE

In this paper, the architecture of IoT is modeled as a graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, in which $\mathcal{V} = \{v_i\}$ is a set of $N$ nodes represent IoT-compatible products, and $\mathcal{E} = \{(v_i, v_j)\}$ is a set of edges that indicate the information flow between nodes. An adjacency matrix $\boldsymbol{A} \in \mathbb{I}^{N\times N}$ is used to model the topology and its elements defined as

$$A_{ij} = \begin{cases} 1, & (v_i, v_j) \in \mathcal{E} \\ 0, & otherwise \end{cases}$$

In the probabilistic model, the correlations among nodes are represented with the correlation probability matrix $\boldsymbol{C} \in [0,1]^{N\times N}$ and its elements are conditional probabilities $C_{ij} = P(x_j|x_i)$ with random state variables $x$'s associated with the nodes. Therefore the edges in the probabilistic graph model are directed.

### 3.1 Probabilistic Model

In an IoT system, suppose that each node has its own sensing, computation, and reasoning capabilities. The *prediction probability* that node $v_i$ detects the true state of world $\theta$ is

$$P(x_i = \theta) = p_i \tag{1}$$

where $x_i$ is the state variable. The information dependency between nodes is modeled with *reliance probability*

$$P\left(x_j = \theta \middle| x_i = \theta\right) = p_{ij} \tag{2}$$

which is the probability that node $v_j$ predicts the true state of world given that node $v_i$ predicts correctly. Similarly, we also have

$$P\left(x_j = \theta \middle| x_i \neq \theta\right) = q_{ij} \tag{3}$$

The *entropy* corresponding to the prediction probability of the *i*th node is

$$H(x_i) = -p_i \log p_i - (1 - p_i)\log(1 - p_i) \tag{4}$$

and the ones to reliance probabilities are

$$\begin{gathered} H\left(x_{ij}\right) = -p_{ij} \log p_{ij} - \left(1 - p_{ij}\right)\log\left(1 - p_{ij}\right) \\ H\left(x_{ij}^C\right) = -q_{ij} \log q_{ij} - \left(1 - q_{ij}\right)\log\left(1 - q_{ij}\right) \end{gathered} \tag{5}$$

Additionally, the *conditional entropies* that quantify the information inter-dependency between state variables $x_i$'s are defined as

$$\begin{aligned} H\left(x_j \middle| x_i\right) &= -\sum_{x_i}\sum_{x_j} P\left(x_j \middle| x_i\right)P(x_i)\log P\left(x_j \middle| x_i\right) \\ &= -p_{ij}p_i \log p_{ij} - (1 - p_{ij})p_i \log(1 - p_{ij}) \\ &\quad -q_{ij}(1 - p_i)\log q_{ij} - (1 - q_{ij})(1 - p_i)\log(1 - q_{ij}) \end{aligned} \tag{6}$$

The *mutual information* between state variables $x_i$ and $x_j$ is defined as

$$M\left(x_i, x_j\right) = H\left(x_j\right) - H\left(x_j \middle| x_i\right) = H(x_i) - H\left(x_i \middle| x_j\right) \tag{7}$$

which measures the extent that knowing one variable influences the knowledge about the other. It is zero if the two variables are independent. Mutual information thus can give an estimate of how much information exchange occurs among nodes in an IoT system. In a normal situation, the system is functioning at a stable level of information exchange. When the system is disrupted with connections broken down, the amount of information exchange will reduce. Therefore, mutual information is proposed here to measure the performance of an IoT system, described in the next section.

### 3.2 Performance Measure of IoT System

A metric that measures the performance of IoT systems should have the following properties. First, the metric should be deterministic and monotone so that one-to-one correspondence between systems and measures can be established. Mutual information of two random variables $x$ and $y$ is non-negative. It is zero when the two variables are totally uncorrelated. It reaches maximum when the two are the same variable. That is, $0 \leq M(x,y) \leq M(x,x)$ . In addition, mutual information is a symmetric metric and $M(x,y) = M(y,x)$.

Second, the metric should be dimensionality independent so that the performances of systems can be compared regardless their sizes. Calculating the average value of pairwise mutual information is necessary so that the measure is independent of the number of nodes. In addition, mutual information of random variables with discrete probability distributions also depends on the number of possible values for the random state variables, i.e. the size of state space or the probability mass functions associated with the state variables. A dimensionless measure for probabilistic design should incorporate the degrees of freedom for the system and the sizes of the state space.

Third, the metric should be sensitive to the change of systems when used for resilience measurement. The function and reliability of a system are sensitively dependent on those of subsystems and components. The metric should also be sensitive enough to reflect the changes at the component level.

Based on the above requirements, the proposed performance metric for an IoT system with $N$ nodes and $D$-nary state variables is

$$F = \frac{1}{DN^2}\sum_{i=1}^{N}\sum_{j=1}^{N} M\left(x_i, x_j\right) \qquad (8)$$

which is the average pairwise mutual information of the system. In the current setting of probabilistic design, $D$=2 (i.e. $x_i = \theta$ and $x_i \neq \theta$).

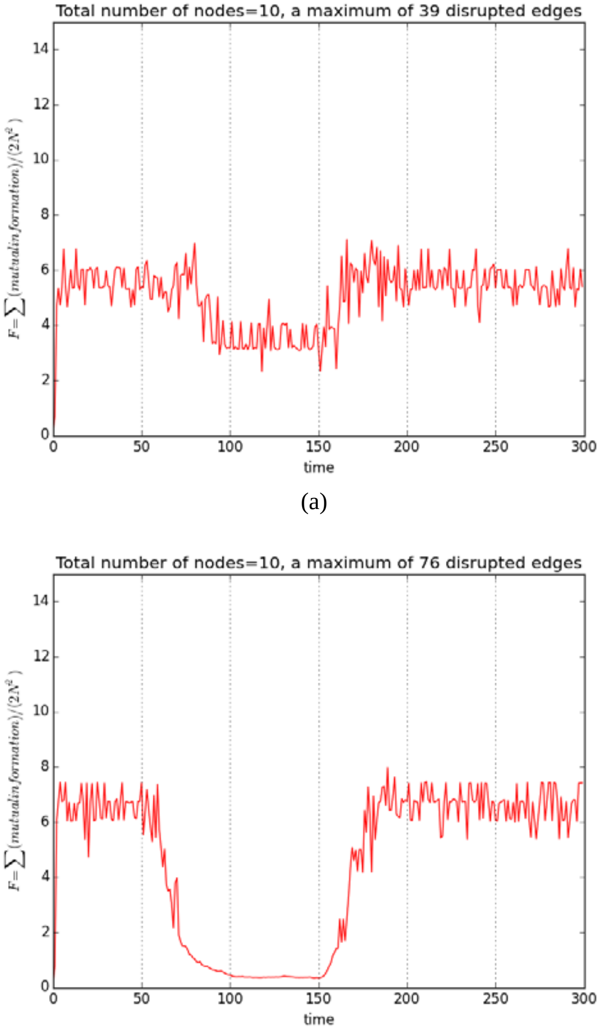
To demonstrate and evaluate the applicability of the proposed mutual information based performance metric to resilience measurement, a simulation study is conducted. In this study, the prediction and reliance probabilities for an IoT network are first randomly generated. Then samples of the random state variables are generated based on the prediction and reliance probabilities. Within each iteration, for each state variable $x_i$, its value as either true or false prediction is sampled based on prediction probability $p_i$ in Eq.(1). The prediction of $x_j$ is then updated to a sample that is drawn based on reliance probability either $p_{ij}$ in Eq.(2) or $q_{ij}$ in Eq.(3), depending on the value of $x_i$. The update of prediction is based on the following best-case rule of information fusion

$$P\left(x_j = \theta\right) = 1 - \prod_{i=1}^{N}\left(1 - P(x_j = \theta | x_i)\right) \qquad (9)$$

where any correct prediction as a result of the information cue from any connected node leads to a success. The sampling iterations continue until enough numbers of samples for all nodes are drawn for one time step. The prediction probabilities for all nodes are then updated based on the frequencies of correct predictions from the samples. The mutual information for each pair is calculated and the system performance in Eq.(8) is estimated. With the updated prediction probabilities, the system moves on to the next time step, and the same sampling and update procedures continue until the predetermined time limit is reached.

During the simulation, the system disruption and recovery occur at certain time steps, which are modeled with the changes of reliance probabilities. When the disruption occurs, the reliance probabilities (both $p_{ij}$ and $q_{ij}$) of some randomly selected pairs are set to be zeros. At the recovery stage, these disconnected pairs are reconnected with the previous reliance probabilities recovered.

The performance measures from the simulation of a system with 10 nodes is shown in Figure 2. For each iteration, 500 samples are drawn. The disruption starts at time step 50 and ends at time step 100, during which a number of connections are randomly selected as disrupted edges at each time step. By the time step of 100, the total number of disrupted connections is 39 for the case in Figure 2(a) and is 76 for the case in Figure 2(b). The recovery period starts from time step 150 and ends at time step 200. The system is fully recovered by time step 250 and reaches the new equilibrium.  It is seen that the proposed performance metric can sensitively detect disruptions from its trend. The volatility is mostly due to the relatively small number of nodes and sample sizes.

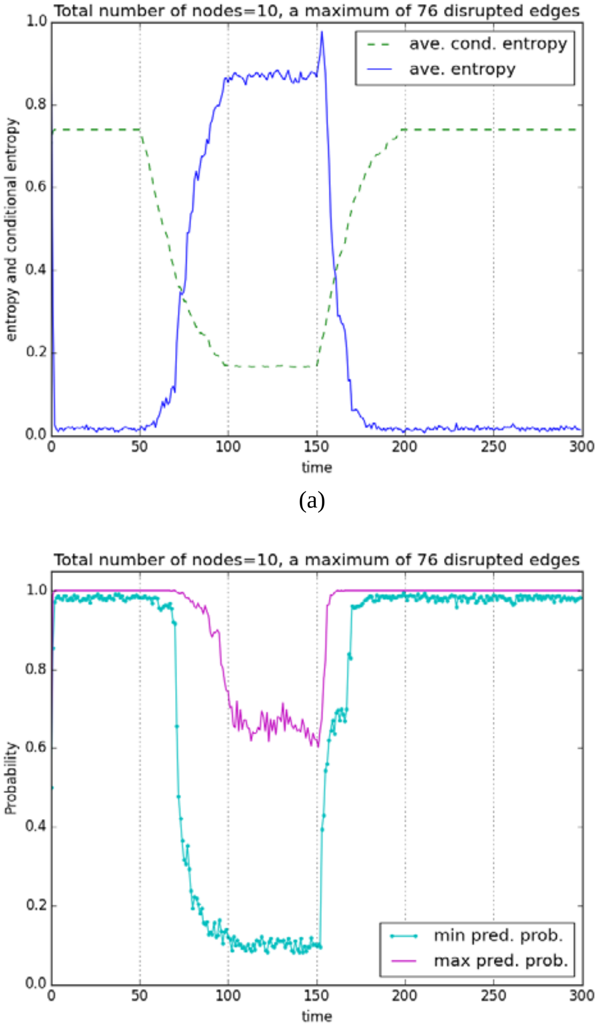


(a)

(b)

Figure 2: Performance measure in Eq.(8) for a simulated IoT system with 10 nodes. (a) The maximum number of disconnected edges is 39. (b) The maximum number of disconnected edges is 76.

The dynamics of entropies and probabilities in the system in Figure 2(b) is shown in Figure 3. The average values of conditional entropies calculated from Eq.(6) and the average values of entropies calculated from the prediction probabilities in Eq.(4) are shown in Figure 3(a). During the disruption, the conditional entropies decrease, while the entropies associated with the prediction probabilities increase. The entropies have small values during the normal working period, because the prediction probabilities are relatively high. This is illustrated in Figure 3(b) where the maximum and minimum values of prediction probabilities among the 10 nodes are compared. The highest prediction probability is one. During the disruption, the differences between the prediction probabilities significantly increase. In other words, disruption affect the prediction capabilities of some nodes, and their prediction probability drop. This in turn affects other nodes. It is seen the highest value of prediction probability among the nodes is not one any more.



(a)

(b)

Figure 3: The entropies and prediction probabilities of the simulated system in Figure 2(b) where the maximum number disconnected edges is 76. (a) The average conditional entropy calculated from Eq.(6) and the average entropy calculated from prediction probability in Eq.(4). (b) The minimum and maximum values of prediction probabilities among 10 nodes.

The number of nodes affects the overall performance and reliability of the system. Figure 4 shows the simulation results when the number of nodes increases to 30 and the total number of connections is 870. It is seen in Figure 4(a) that the system performs fairly robustly when the maximum number of disrupted connections is 49. The mutual information increases slightly instead of decrease during the disruption. This is because mutual information includes two components, entropy and conditional entropy, according to Eq.(7). During the disruption period, the conditional entropies associated with those disrupted edges reduce to zeros, whereas the prediction probabilities thus entropies of the relevant nodes are not affected. As a result, the mutual information increases. This phenomenon is also observed in Figure 4(b) where the maximum number of disrupted connections is 828. Shortly after the disruption starts at time step 50, the average mutual information increases. Again, this is due to the reduction of conditional entropies while entropies associated with prediction probabilities remain unchanged, which is verified by plotting the average entropies and conditional entropies in Figure 5(a) and the maximum and minimum prediction probabilities in Figure 5(b). As the number of disconnected edges keeps increasing, prediction probabilities are affected. Mutual information starts decreasing until the maximum number of 828 disconnections is reached at time step 100. The system is stabilized in the next 50 time steps until recovery starts. During recovery, mutual information returns to the level prior to disruption reversely. After time step 200, the system is fully recovered.

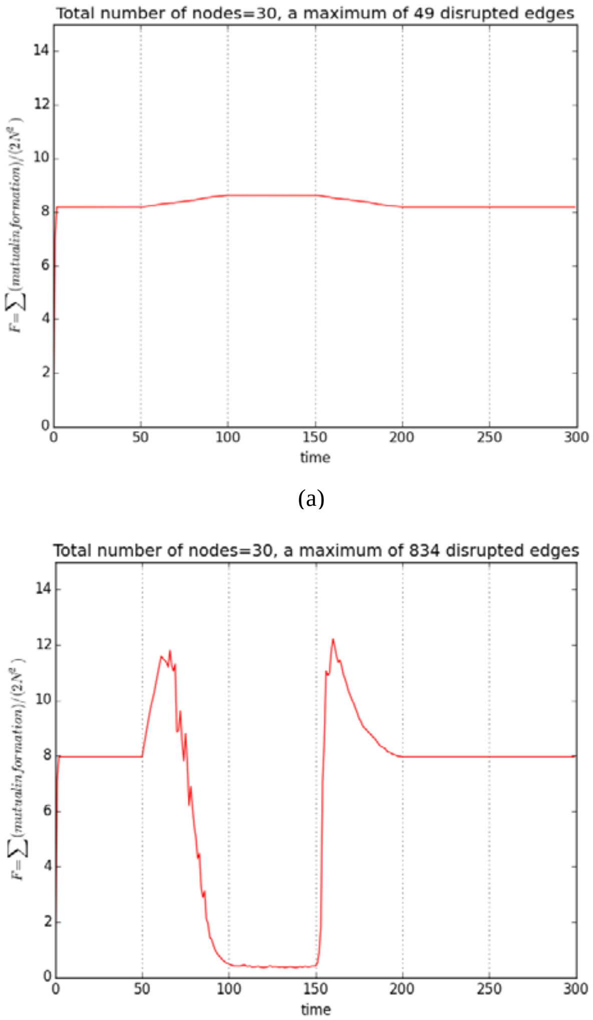
Notice that the average entropies are zeros at the normal working condition for the large network of 30 nodes in Figure 5(a). This is because the prediction probabilities of all nodes are ones before disruption, shown in Figure 5(b). The network is fully connected at the beginning because all pair-wise reliance probabilities are randomly generated. The predictions by all nodes are accurate. The predictions become not reliable after the number of disconnected edges reach certain level after disruption has started. Some of the prediction probabilities reduce. As a result, the average entropy increases. The prediction capabilities of the nodes quickly recover after some of the connections resume. Intuitively the system should become more resilient to disruption when the number of nodes increases. It is confirmed by the simulation results. The examples show that the mutual entropy based performance measure is sensitive to the system topological change. It provides detailed information about the changes of prediction and reliance probabilities. The metric allows us to quantify the resilience of IoT systems described with the probabilistic model. This performance measure is applied in the further study of system resilience and probabilistic design of the system architecture.

## 4. PROBABILISTIC DESIGN OF IOT ARCHITECTURE

With the performance metric quantitatively defined, system design and optimization can be performed. The overall goal of the design of IoT system architecture is to find optimum network topology such that the system performance is maximized.

It is seen that the reliability of prediction is related to the number of nodes in the system and connections that are available during disruption. Larger systems with more nodes and more connections tend to be more robust and give correct predictions than smaller systems. Therefore the design decision variables need to include the number of nodes, the respective prediction probabilities, and pair-wise reliance probabilities. Note that the topology of networks in our probabilistic model is quantified by. reliance probabilities instead of binary connectivity. In addition, the performance of prediction is also related to the information fusion rules, based on which the prediction probabilities are updated. Design decisions also include the selection of the rules.

In this section, several information fusion rules for reasoning and decision making are described. The sensitivities of system performance with respect to the prediction and reliance probabilities are also analyzed. Sensitivity analysis of design variables provides some insight of search domains in design optimization.



(a)

(b)

Figure 4: Performance measure of a simulated IoT system with 30 nodes. (a) The maximum number of disconnected edges is 49. (b) The maximum number of disconnected edges is 834.

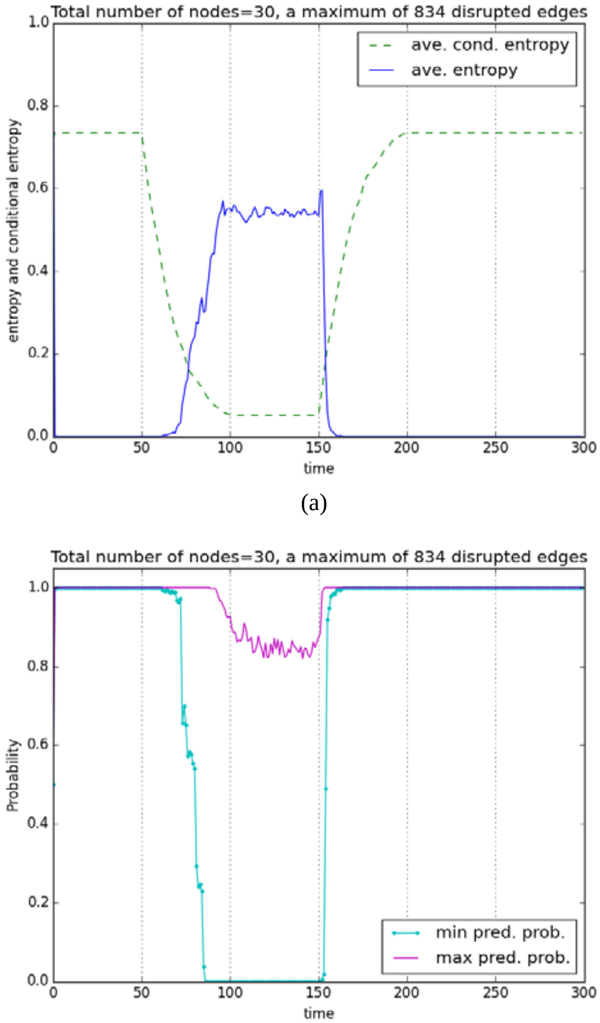
### 4.1 Rules of Information Fusion

The prediction probabilities are also sensitively dependent on the rules of information fusion during prediction update. When receiving different cues from topologically correlated neighbors, a node needs to update its prediction probability to reflect the true state of the world. Several rules can be devised in addition to the best-case rule in Eq. (9). They are listed as follows.

- Best-case (optimistic)

$$P(x_j) = 1 - \prod\nolimits_{i=1}^{M} \left(1 - P(x_j|x_i)\right)$$

If any of the *M* correlated nodes provides a positive cue, the prediction of the node is positive. Some variations of the rule include when the cases of negatively correlated nodes are also considered, as



(a)

(b)

Figure 5: The entropies and prediction probabilities of the simulated system in Figure 4(b) where the maximum number disconnected edges is 834. (a) The average conditional entropy and the average entropy. (b) The minimum and maximum values of prediction probabilities among 30 nodes.

$$P(x_j) = 1 - \prod\nolimits_{i=1}^{M} \left(1 - P(x_j|x_i)\right)\left(1 - P(x_j|x_i^C)\right)$$

as well as when the node's own observation is excluded, as

$$P(x_j) = 1 - \prod\nolimits_{i=1, i\neq j}^{M} \left(1 - P(x_j|x_i)\right)$$

- Worst-case (pessimistic)

$$P(x_j) = \prod\nolimits_{i=1}^{M} P(x_j|x_i)$$

The prediction of the node is positive only if all of the $M$ correlated nodes provide positive cues. Similarly, there could be some variations of the rule, such as

$$P(x_j) = \prod\nolimits_{i=1, i\neq j}^{M} P(x_j|x_i)$$

- Bayesian

$$P'(x_j) \propto P(x_j)\left(P(x_j)\right)^r \left(1 - P(x_j)\right)^{M-r}$$

The prediction of the node is updated to $P'$ from prior prediction $P$ given the cues that the $M$ correlated nodes provide, among which $r$ of them provide a positive cue.

The simulation results based on the Bayesian fusion rule is shown in Figure 6, where the update of prediction probabilities is gradual and much slower than the update based on the other two rules.

Some other rules can be defined for information fusion, such as product-sum, weighted average, evidence-based, etc. Those empirical rules are less restrictive than the above three conventional ones.

### 4.2 Sensitivities of Probabilities

The closed-form local sensitivities of conditional entropies with respect to prediction and reliance probabilities can be obtained as

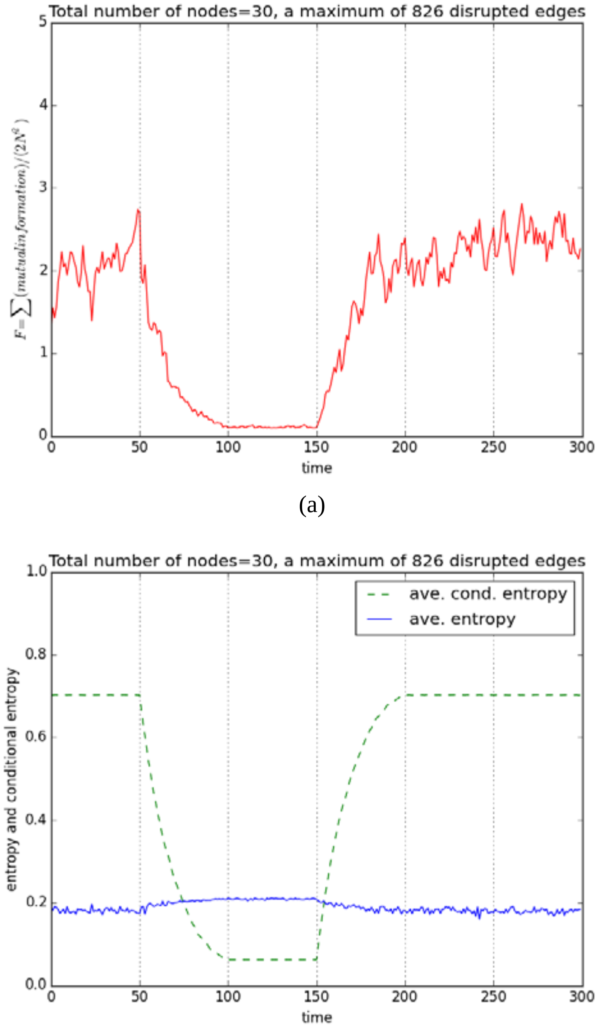
$$\frac{\partial H(x_j|x_i)}{\partial p_{ij}} = p_i \log \frac{1 - p_{ij}}{p_{ij}} \tag{10}$$

$$\frac{\partial H(x_j|x_i)}{\partial q_{ij}} = (1 - p_i) \log \frac{1 - q_{ij}}{q_{ij}} \tag{11}$$

$$\frac{\partial H(x_j|x_i)}{\partial p_i} = p_{ij} \log \frac{1 - p_{ij}}{p_{ij}} + q_{ij} \log \frac{q_{ij}}{1 - q_{ij}} + \log \frac{1 - q_{ij}}{1 - p_{ij}} \tag{12}$$

It is seen in Eqs.(10) and (11) that the first derivatives of conditional entropy with respect to reliance probabilities are monotonically positive when $p_{ij} < 0.5$ and $q_{ij} < 0.5$. For small probabilities, increasing their values would increase the conditional entropies. They become negative when $p_{ij} > 0.5$ and $q_{ij} > 0.5$, and the trend is the opposite.

The first derivatives of conditional entropies with respect to prediction probabilities are not monotonic, as seen in Eq.(12). They are functions of reliance probabilities, which have (0.5,0.5) as a saddle point, as shown in Figure 7. When $q_{ij} < 0.5$ and $q_{ij} < p_{ij} < 1 - q_{ij}$, or $q_{ij} > 0.5$ and $1 - q_{ij} < p_{ij} < q_{ij}$, the sensitivities are in the positive domain.



(a)

(b)

Figure 6: Simulation results based on the Bayesian fusion rule for a system of 30 nodes with a maximum of 826 disrupted connections. (a) Average mutual information performance measure. (b) Average conditional entropy and entropy.

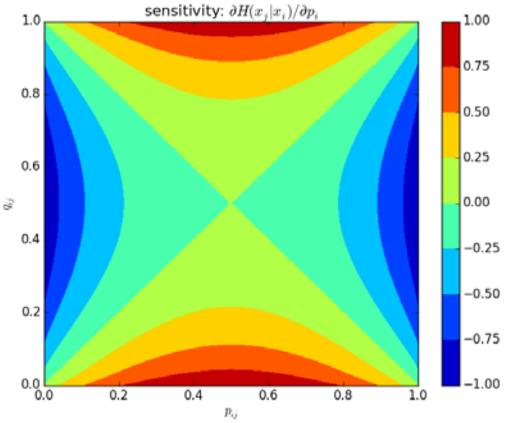
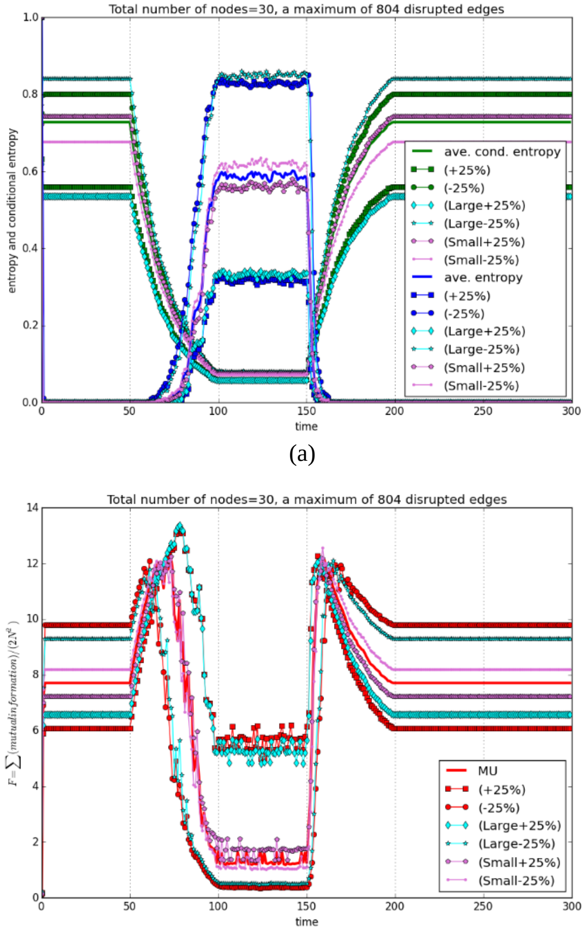


Figure 7: Sensitivity of conditional entropy with respect to prediction probability



(a)

(b)

Figure 8: Sensitivity analysis based on the best-case fusion rule by increasing and reducing all reliance probabilities by 25% (+/−25%), increasing and reducing only those large probabilities that are greater than 0.5 by 25% (Large+/−25%), and increasing and reducing only those small probabilities that are less than 0.5 by 25% (Small+/−25%). (a) Average conditional entropies and entropies. (b) Average mutual information.

Understanding the local sensitivity of conditional entropies is useful for local adjustment of probabilities especially when the system's prediction probabilities are not sensitive to the changes of reliance probabilities. Further increasing the reliance probabilities that are greater than 0.5 or decreasing the ones that are less than 0.5 for those uninterrupted nodes will reduce the conditional entropies. It could suggest that adjustment of reliance probabilities can be focused more on those ones with larger values.

The sensitivity analysis is verified by the results of sensitivity analysis shown in Figure 8. The sensitivity analysis is done by varying the levels of reliance probabilities. Six different situations are tested, including increasing and reducing all reliance probabilities by 25%, increasing and reducing only those large probabilities that are greater than 0.5 by 25%, and increasing and reducing only those small probabilities that are less than 0.5 by 25%. When a modified probability value exceeds 1, it is set to be the maximum value of 1. It is seen in Figure 8(a) that both conditional entropies and entropies are more sensitive to the large reliance probabilities than to the small ones. The changes applied to large probabilities are more effective than the changes to all probabilities for conditional entropies, as previously predicted. Similarly, changing large reliance probabilities gives the similar results of changing all of the probabilities.

Therefore, improving those relatively reliable connections or sources of information with large reliance probabilities is more effective to optimize the system performance than simultaneously considering all connections in a system.

The sensitivity of the system is also dependent on the information fusion rules. When the Bayesian rule is applied, the system is not sensitive to the changes of reliance probabilities any more. As shown in Figure 9, the variation of the average mutual information as a result of different reliance probabilities is small.

According to the quantitative definitions of resilience in Section 2.2, the systems with the Bayesian rule are more robust with respect to the changes of reliance probabilities, however less resilient with respect to disruption, than the ones with the best-case rule. In the above sensitivity studies, common random numbers are used in the comparison among different systems. This is to reduce variance introduced in the simulation.

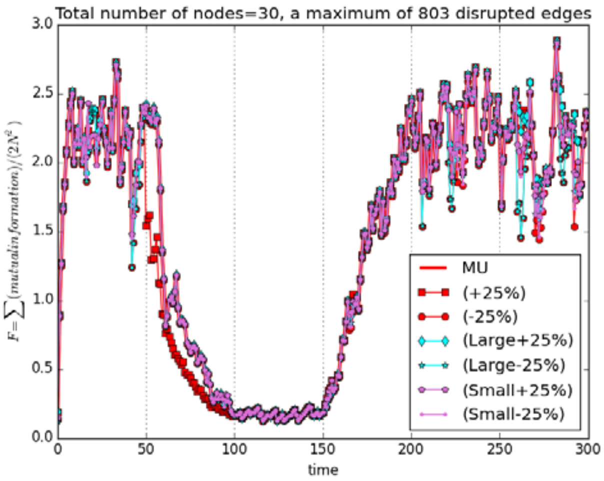


Figure 9: Sensitivity analysis based on the Bayesian rule

## 5. DISCUSSIONS AND CONCLUSION

In this paper, a probabilistic deign framework for designing IoT system architecture is proposed. In IoT networks, each node corresponds to an IoT-compatible product. The processes of communication during information exchange between nodes and reasoning at individual nodes are characterized with reliance and prediction probabilities respectively. The resilience of the system is quantified with the proposed performance metrics of entropy and mutual information. These metrics measure how communication and reasoning capabilities are affected during network disruption. The metrics are shown to be sensitive to the changes of network topology.

Several information fusion rules are also defined so that the probabilities associated with a node are updated based on the received information from neighboring nodes during reasoning. The system performance is also sensitively dependent on the fusion rules. During the design process of IoT systems, information aggregation rules also need to be optimized based on the expected dynamic performance.

The sensitivity studies also show that the system performance is influenced more by the tightly coupled nodes, where reliance probabilities are high, than those loosely coupled ones. The optimization of systems is more effective if efforts are focused on these connections with high reliance probabilities, if the available resource is limited for improvement.

The proposed metrics perform reasonably well with the simple reasoning scheme. As future extensions, the proposed metrics need to be further tested with some other information fusion rules. The optimization methods also need to be explored based on the sensitivity analysis results from this work.

## REFERENCES

[1] K. Ashton, “That ‘Internet of Things’ Thing,” *RFiD Journal*, vol.22, pp.97-114, 2009.

[2] Gartner (Dec. 12, 2013) “Gartner Says the Internet of Things Installed Base Will Grow to 26 Billion Units By 2020” http://www.gartner.com/newsroom/id/2636073

[3] Holling, C. S. (1961). Principles of insect predation. *Annual Review of Entomology*, *6*(1), 163-182.

[4] Rosenzweig, M. L., & MacArthur, R. H. (1963). Graphical representation and stability conditions of predator-prey interactions. *American Naturalist*, *97*(895), 209-223.

[5] Lewontin, R.C. (1969). The meaning of stability. In: Diversity and Stability of Ecological Systems. Brookhaven Symposia in Biology No 22. Brookhaven, New York.

[6] Holling, C. S. (1973). Resilience and stability of ecological systems. *Annual Review of Ecology and Systematics*, *4*,1-23.

[7] Folke, C. (2006). Resilience: The emergence of a perspective for social–ecological systems analyses. *Global Environmental Change*, *16*(3), 253-267.

[8] Scheffer, M., & Carpenter, S. R. (2003). Catastrophic regime shifts in ecosystems: linking theory to observation. *Trends in ecology & evolution*, *18*(12), 648-656.

[9] Arrow, K., Bolin, B., Costanza, R., Dasgupta, P., Folke, C., Holling, C. S., Jansson, B.-O., Levin, S., Mäler, K.-G., Perrings, C., & Pimentel, D. (1995). Economic growth, carrying capacity, and the environment. *Science*, *268*, 520-521.

[10] Christopherson, S., Michie, J., & Tyler, P. (2010). Regional resilience: theoretical and empirical perspectives. *Cambridge Journal of Regions, Economy and Society*, *3*(1), 3-10.

[11] Martin, R. (2012). Regional economic resilience, hysteresis and recessionary shocks. *Journal of Economic Geography*, *12*(1), 1-32.

[12] Foster, K. A. (2007). A Case Study Approach to Understanding Regional Resilience. Institute of Urban and Regional Development, University of California, Berkeley, Report No. 2007-08.

[13] Hill, E., Wial, H., & Wolman, H. (2008). Exploring regional economic resilience. Institute of Urban and Regional Development, University of California, Berkeley, Report No. 2008-04.

[14] Schiefer, H. F. (1933). The Compressometer An Instrument for Evaluating the Thickness, Compressibility and Compressional Resilience of Textiles and Similar Materials. *Textile Research Journal*, *3*(10), 505-513.

[15] Mark, H. (1946). Some remarks about resilience of textile materials. *Textile Research Journal*, *16*(8), 361-368.

[16] Hoffman, R. M. (1948). A Generalized Concept of Resilience. *Textile Research Journal*, *18*(3), 141-148.

[17] Fielding, J. H. (1937). Impact resilience in testing channel black. *Rubber Chemistry and Technology*, *10*(4), 807-819.

[18] Turner, L. B., Haworth, J. P., Smith, W. C., & Zapp, R. L. (1943). Carbon Black in Butyl Rubber. *Industrial & Engineering Chemistry*, *35*(9), 958-963.

[19] Dillon, J. H., Prettyman, I. B., & Hall, G. L. (1944). Hysteretic and Elastic Properties of Rubberlike Materials Under Dynamic Shear Stresses. *Journal of Applied Physics*, *15*(4), 309-323.

[20] Liu, J. W., Shih, W. K., Lin, K. J., Bettati, R., & Chung, J. Y. (1994). Imprecise computations. *Proceedings of the IEEE*, *82*(1), 83-94.

[21] Hegde, R., & Shanbhag, N. R. (1999). Energy-efficient signal processing via algorithmic noise-tolerance. In *Proceedings of the 1999 international symposium on Low power electronics and design, August 16-17, 1999, San Diego, CA, USA*, pp. 30-35.

[22] Cho, H., Leem, L., & Mitra, S. (2012). ERSA: Error Resilient System Architecture for Probabilistic Applications. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, *31*(4), 546-558.

[23] Chippa, V. K., Mohapatra, D., Raghunathan, A., Roy, K., & Chakradhar, S. T. (2010). Scalable effort hardware design: exploiting algorithmic resilience for energy efficiency. In *Proceedings of the 47th Design Automation Conference* (DAC’10), *June 13-18, 2010, Anaheim, California, USA*, pp. 555-560.

[24] Verma, N., Lee, K. H., Jang, K. J., & Shoeb, A. (2012, March). Enabling system-level platform resilience through embedded data-driven inference capabilities in electronic devices. In *2012 IEEE International Conference on*

*Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5285-5288.

[25] Wang, Z., Schapire, R. E., & Verma, N. (2015). Error Adaptive Classifier Boosting (EACB): Leveraging Data-Driven Training Towards Hardware Resilience for Signal Inference. *IEEE Transactions on Circuits and Systems I: Regular Papers*, *62*(4), 1136-1145.

[26] Abdallah, R., & Shanbhag, N. R. (2013). Error-resilient systems via statistical signal processing. In *Proc. 2013 IEEE Workshop on Signal Processing Systems (SiPS)*, pp. 312-317.

[27] Schneider, F. B. (Ed.). (1999). *Trust in cyberspace*. National Academies Press.

[28] Lin, H. S., & Goodman, S. E. (Eds.). (2007). *Toward a safer and more secure cyberspace*. National Academies Press.

[29] Sterbenz, J. P., Hutchison, D., Çetinkaya, E. K., Jabbar, A., Rohrer, J. P., Schöller, M., & Smith, P. (2010). Resilience and survivability in communication networks: Strategies, principles, and survey of disciplines. *Computer Networks*, *54*(8), 1245-1265.

[30] Hollnagel, E., Woods, D. D., & Leveson, N. (Eds.). (2007). *Resilience Engineering: Concepts and Precepts*. Ashgate Publishing, Burlington, VT.

[31] Madni, A. M., & Jackson, S. (2009). Towards a conceptual framework for resilience engineering. *IEEE Systems Journal*, **3**(2), 181-191.

[32] Hollnagel, E., & Fujita, Y. (2013). The Fukushima disaster – Systematic failures as the lack of resilience. *Nuclear Engineering and Technology*, **45**(1), 13-20.

[33] Zhou, H., Wan, J., & Jia, H. (2010). Resilience to natural hazards: a geographic perspective. *Natural Hazards*, **53**(1), 21-41.

[34] Francis, R., & Bekera, B. (2014). A metric and frameworks for resilience analysis of engineered and infrastructure systems. *Reliability Engineering & System Safety*, **121**, 90-103.

[35] Bruneau, M., & Reinhorn, A. (2007). Exploring the concept of seismic resilience for acute care facilities. *Earthquake Spectra*, **23**(1), 41-62.

[36] Cimellaro, G. P., Reinhorn, A. M., & Bruneau, M. (2010). Framework for analytical quantification of disaster resilience. *Engineering Structures*, *32*(11), 3639-3649.

[37] Ouyang, M., Dueñas-Osorio, L., & Min, X. (2012). A three-stage resilience analysis framework for urban infrastructure systems. *Structural Safety*, **36/37**, 23-31.

[38] Ayyub, B. M. (2015). Practical Resilience Metrics for Planning, Design, and Decision Making. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, *1*(3), 04015008.

[39] Khabbaz, M. J., Assi, C. M., & Fawaz, W. F. (2012). Disruption-tolerant networking: A comprehensive survey on recent developments and persisting challenges. *IEEE Communications Surveys & Tutorials*, *14*(2), 607-640.

[40] Miu, A., Balakrishnan, H., & Koksal, C. E. (2005, August). Improving loss resilience with multi-radio diversity in wireless networks. In *Proceedings of the 11th annual International Conference on Mobile Computing and Networking*, pp. 16-30.

[41] Lei, J. J., & Kwon, G. I. (2010). Reliable Data Transmission Based on Erasure-resilient Code in Wireless Sensor Networks. *TIIS Transactions on Internet & Information Systems*, *4*(1), 62-77.

[42] Huang, Y., Gao, Y., Nahrstedt, K., & He, W. (2009, June). Optimizing file retrieval in delay-tolerant content distribution community. In *Proc. 29th IEEE International Conference on Distributed Computing Systems (ICDCS'09).* pp. 308-316.

[43] Cohen, R., Erez, K., Ben-Avraham, D., & Havlin, S. (2000). Resilience of the Internet to random breakdowns. *Physical review letters*, *85*(21), 4626.

[44] Çetinkaya, E. K., Broyles, D., Dandekar, A., Srinivasan, S., & Sterbenz, J. P. (2013). Modelling communication network challenges for future internet resilience, survivability, and disruption tolerance: A simulation-based approach. *Telecommunication Systems*, *52*(2), 751-766.

[45] Rohrer, J. P., Jabbar, A., & Sterbenz, J. P. (2014). Path diversification for future internet end-to-end resilience and survivability. *Telecommunication Systems*, *56*(1), 49-67.

[46] Paul, G., Sreenivasan, S., & Stanley, H. E. (2005). Resilience of complex networks to random breakdown. *Physical Review E*, *72*(5), 056130.

[47] Sun, F., & Shayman, M. A. (2007). On pairwise connectivity of wireless multihop networks. *International Journal of Security and Networks*, *2*(1-2), 37-49.

[48] Shirazi, F., Diaz, C., & Wright, J. (2015, October). Towards measuring resilience in anonymous communication networks. In *Proc. the 14th ACM Workshop on Privacy in the Electronic Society*, pp. 95-99.