Preliminary Interdependency Analysis: An Approach to Support Critical Infrastructure Risk Assessment

Robin E. Bloomfield[ab], Peter Popov[a], Kizito Salako[a], Vladimir Stankovic[a], David Wright[a]

[a]The Centre for Software Reliability, City, University of London, EC1V 0HB, London, UK

reb@csr.city.ac.uk, ptp@csr.city.ac.uk, k.o.salako@city.ac.uk, Vladimir.Stankovic.1@city.ac.uk, d.r.wright@city.ac.uk

[b]Adelard LLP, 24 Waterside, 44-48 Wharf Road, London N1 7UX, London, UK

reb@adelard.com

Abstract

We present a methodology, *Preliminary Interdependency Analysis* (PIA), for analysing interdependencies between critical infrastructure (CI). Consisting of two phases – qualitative analysis followed by quantitative analysis – an application of PIA progresses from a relatively quick elicitation of CI-interdependencies to the building of representative CI models, and the subsequent estimation of *any* resilience, risk or criticality measures an assessor might be interested in. By design, stages in the methodology are both flexible and iterative, resulting in interacting CI models that are scalable and may vary significantly in complexity and fidelity, depending on the needs and requirements of an assessor. For model parameterisation, one relies on a combination of field data, sensitivity analysis and expert judgement. Facilitated by dedicated software tool support, we illustrate PIA by applying it to a complex case-study of interacting Power (distribution and transmission) and Telecommunications networks in the Rome area. A number of studies are carried out, including: 1) an investigation of how "strength of dependence" between the CIs' components affects various measures of risk and uncertainty, 2) for resource allocation, an exploration of different, but related, notions of CI component importance, and 3) highlighting the impact of model fidelity on the estimated risk of cascades.

Keywords: Interdependency Analysis, Risk Assessment, Cascading failure, Critical Infrastructure Resilience

1. Introduction

It is recognised that one of the challenges in enhancing the protection of *Critical Infrastructures*[1] (CIs) against accidents, natural disasters, and acts of terrorism (including cyber terrorism) is establishing and maintaining an understanding of the interdependencies between infrastructures. Governmental agencies responsible for protecting national critical infrastructure need methods and tools to assess risks (including those related to interdependencies) and evaluate the alternatives available for mitigating these. The owners and operators of critical infrastructure need to know the likely impact, on their services, of disruptions from other infrastructures, so they can develop mitigations (e.g. in their emergency planning) and make considered investments in *resilience* [1].

Once one recognises the importance – in terms of risks – of interdependencies between critical infrastructures, one is then faced with the feasibility and cost of a risk-assessment [2-5], since critical infrastructures are typically large and very complex systems. Model-based risk-assessment can offer a feasible and cost-effective assessment approach for an assessor, *if* the assessor can gain enough confidence that her models are representative of the system's behaviour, capturing what she judges to be essential interdependencies. Faced with numerous choices about model structure, fidelity and parameters, our assessor can gain confidence in a model by a succession of model *refinements*, each refinement resulting from verifying and validating an earlier version of a model and making judgements about what changes to the model are needed for an improvement while, at the same time, *not* putting in more detail than she judges to be necessary for her needs. So, for instance, if an assessor has certain *risk-measures*[2] in mind (e.g. the distribution of loss in network-connectivity resulting from component failure or

[1] As defined by the U.S. Department of Homeland security (see https://www.dhs.gov/what-critical-infrastructure).

[2] In this paper, for ease of presentation and without-loss-of-generality, a *risk-measure* is a probability distribution of unwanted events arising from random changes in a CI's state.

the distribution of loss in supplied electrical power due to line-outages in a snow storm) which, to be computed, require the model to explicitly have dynamics of a certain kind (e.g. packet-routing algorithms or electrical power flow models), then these dynamics will need to be incorporated in a revision of the model.

Clearly, with so many choices to make, the task of model building and refinement can be a daunting one, with serious ramifications for the risk-assessment to be carried out. Any methodology/tools which support an assessor in this endeavour should afford the assessor the flexibility to 1) create models at any desired level of abstraction, 2) alter/add/remove stochastic and deterministic processes, and 3) define any risk-measure of interest. To this end, we propose *Preliminary Interdependency Analysis* (PIA) – a systematic method to support building, refining and analysing models of interdependent *Large Complex Critical Infrastructures* (LCCI). PIA starts off at a high-level of abstraction, supporting a cyclic, systematic thought process, directed towards identifying dependencies between components of CIs. Eventually, (hybrid) probabilistic models are deployed, once they have been judged to be appropriate for risk-assessment; these are used to conduct studies focussed on computing different measures of interests, e.g. the likelihood of cascade failure under a given set of assumptions, or the identification of the weakest link in the modelled system. And, if modelling with even greater detail is required, PIA can assist in this process too, e.g. by adding models of the consequences of LCCI operator actions, or by introducing various constraints on such actions, such as limiting the maintenance resources available in the case of a major disaster, or adding deterministic models specific to a particular LCCI (e.g. power flows for power systems).

The PIA method is applicable as both:

1. a lightweight method used to provide an initial identification of interdependencies and to scope the options for more detailed studies. The approach should be accessible to a range of stakeholders, particularly *Small-to-Medium Enterprises* (SMEs) in support of their business continuity planning
2. a more heavyweight method of studying, with an increasing level of detail, complex regional and nationwide CIs by combining probabilistic and deterministic models of the CIs.

There are numerous studies about CI interdependencies, including some which rely on complex dynamic models. As pointed out in a recent survey [6] summarising research on interdependencies in power systems for the last 5 years, many studies analyse interdependencies without detailing *how* these interdependencies were identified in the first place, giving the impression that the interdependencies are *all known* to the analyst. Systematic methods which can be followed to identify interdependencies are lacking in the literature. The authors of the survey, therefore, recommend that methods for interdependency identification be given high priority. We agree, and PIA provides significant support in this direction.

We illustrate the use of PIA on a realistic case study: a regional system of two CIs, namely the power grid and the telecommunication network around Rome, Italy (i.e. Rome case-study). In the study, we used a set of tools – the PIA Toolkit – which consists of two software applications we developed:

- Using the PIA Designer, a modeller can construct and parameterize a visual representation of interdependent CIs. The PIA Designer converts this visual model into a probabilistic model ready to be solved via Monte Carlo simulation. The Designer uses third party proprietary software called ASCE [7].
- The Execution Engine allows for Monte Carlo simulation using models created with the PIA Designer. The Execution engine uses Möbius [8], which we customised extensively to 1) allow for various forms of dependencies between the modelled elements, and 2) for integration of third party software in simulation (e.g. various deterministic flow models, typically used with the CIs).

The rest of the paper is organised as follows. Section 2 presents related research, while an overview of the PIA method – both its qualitative and quantitative aspects – is given in section 3. Section 4 details the mathematical family of models underlying quantitative PIA, including models of interdependent CIs and their dependent constituent components. In section 5 we describe the case study used to illustrate our approach. This is followed by a presentation of results obtained, and a discussion of their plausibility, in section 6. In section 7, we discuss our findings, and open issues for future research, while finally concluding the paper in section 8. Appendix A contains a detailed illustration of model development over various stages of PIA, using PIAs tool support in the aforementioned case-study.

2. Related Research

The authoritative paper by Rinaldi et al. [9] established interdependency related terminology and concentrates on high level dependencies between infrastructures. It was noticed, however, that such an approach, although useful at a conceptual level, is inappropriate for risk quantification as further elaboration is needed. Many authors, including ourselves, have since argued in favour of service-level models of a different flavour.

An overview of CI interdependency research is provided in our earlier study on interdependencies for UK agencies [10, 11]. A more recent survey is [12], in which a number of modelling and simulation approaches are grouped into six categories: 1) Empirical, 2) Agent-based, 3) Economic-based, 4) Complex-Network based, 5) System-dynamics based and 6) "Others", which covers all approaches not included in the previous categories. According to this classification, our work belongs to the "Others" category, partly because our work incorporates approaches from more than one category. We compare these approaches to PIA below.

PIA allows one to estimate risk using alternative, consistent models, thereby allowing risk-measures resulting from these models to be directly compared. We see this capability as a useful step in addressing the research gap identified at the end of section 4.1.2 in [12]. As an empirical modelling approach, PIA can be used for 1) identification of frequent and significant failure patterns, as well as 2) quantification of any risk-measures chosen by an assessor.

Agent-based models, consisting of dynamically interacting rule-based agents, are based on the idea that complex behaviour or phenomena emerge from many individual and relatively simple interactions of autonomous agents [13-15]. In terms of emergent model properties, there are similarities between PIA and agent-based modelling approaches. The deterministic rules that govern the behaviour of agents can be modelled in PIA as well, as the deterministic responses of components to a system's random changes in time. But, PIA extends this concept by introducing *stochastic association*s, which define deterministic rules governing *how* the uncertainty in the model depends upon the state of the system and its components.

In contrast with the "bottom-up" approach of Agent-based models, *System-dynamics* approaches take a "top-down" view [16-18] by focusing on the nonlinear behaviour of systems over time, using *stocks* and *flows*, internal *feedback loops* and time delays. This nonlinear behaviour is typically characterised by a set of differential equations capturing the behaviour of systems with *fixed network topologies* – some see this as a significant limitation [12]. PIA is fully compatible with these techniques, but in addition allows an assessor to analyse a system with uncertainty in network topology.

The quantitative analysis of risk typically requires the evolution of a CIs state be modelled as a stochastic process; the process is defined by a collection of joint probability distributions over a very large state-space. While, to some extent, there exist tools and formalisms to aid an assessor with this, such as PRISM [19-21], difficulties can arise if 1) the state-space is exceedingly large (e.g. too large to explicitly fit in computer memory), making infeasible the solution of such problems using transition-rate matrices, 2) the inter-event times for the process have no known mathematical closed-form. However, PIA, by using a combination of stochastic associations, deterministic state-transitions and the competing-risks algorithm [22, 23], affords a user the ability to both specify sophisticated joint distributions and simulate the resulting process. These resulting processes are *hybrids* of semi-Markov processes and embedded deterministic state-transitions. Furthermore, *any* inter-event time distribution that can be sampled from efficiently can be used in defining the process.

Complex-Network based approaches [24, 25], broadly grouped in [12] into topology-based and flow-based methods, model single CIs by networks (i.e. graphs with nodes and links) and describe interdependencies by inter-links, providing CI representations with detailed descriptions of their topologies and flow patterns. In terms of model fidelity, these approaches are pitched at a fairly high level of abstraction. Also, the use of *probabilistically independent* events is quite common with these approaches and simplifies the analysis of such models. PIA is, however, not restricted by level of abstraction, or the use of probabilistically independent events. On the contrary, we encourage PIA users to explore various levels of abstraction and alternative forms of stochastic dependence between the modelled entities. By such exploration, an assessor is better equipped to make an informed decision about model accuracy and usefulness.

In [26] an approach to modelling interdependencies is developed that considers both structural properties, using techniques employed in graph theory, as well as functional properties, to increase the fidelity and usefulness of the approach. The approach is applied to a complex case-study that includes rail transportation, power grid and telecommunication. In essence, this is used to study the effects of removing a single component from the respective network and how system performance varies as a result. The work, however, does not take into account the likelihoods of different components becoming inoperable; this is significantly different from our methodology. Also, the approach is clearly limited in its potential to see the effects of multiple elements being removed (e.g. when they fail simultaneously).

An interesting observation made in [27] is that services in some types of infrastructure, such as telecommunications or the electric grid, are provided and consumed instantly. Others, notably oil, gas and infrastructure built on physical resources, however, exhibit buffering characteristics. This aspect is not explicitly modelled in our approach, although taking this into account should not pose a problem - PIA offers ways of modelling complex stochastic behaviour and buffering (at certain level of abstraction) seems no more than deterministic delays between "cause" and "consequences". More detailed models of buffering can be added via a

custom-built deterministic model.

An approach similar to ours is presented in [28] in which the authors seem to refer to a scaled-down version of the case-study used in this paper, and concentrate on modelling the availability of the SCADA system. The key difference is that the focus there is on the topology of the specific system and on building a specialised *Stochastic Activity Networks* (SAN) model of availability, rather than presenting a generic method of studying CI interdependency scenarios. In addition, based on our experience, building SAN models from scratch for every new case-study 'does not scale' up – for scenarios of typical complexity it is time consuming, error prone and can be difficult to debug. We addressed this difficulty by developing a tool support, based on SAN formalism and ASCE tool, which complements PIA, and which is briefly summarised in the introduction and its use is demonstrated in the appendix. A large number of publications review the concept of risk when applied to interdependent CI. The survey in [29] discusses a number of approaches, and proposes an approach easily applicable in practice. These authors argue in favour of ranking the incidents according to their frequency and impact and demonstrate the application to a case-study of critical infrastructures in Oslo. The approach, however, seems simplistic, as the ball park figures used for ranking the incidents are not convincing. Instead of adopting a similar view – prescribing a particular way of defining component criticality – in this paper we demonstrate that criticality may vary significantly with the definition of criticality. A component seen as highly critical using one definition of criticality, e.g. the one used by those authors, may well turn out to be a low criticality component using a different definition of criticality, e.g. the likelihood of a component being a part of a large cascade. We discuss the practical implication of our observation.

It is also interesting to relate PIA to other methods used in safety and dependability analysis. Methods for safety analysis can be understood in terms of how they support the *discovery* of the right system model (even if implicitly) and the *exploration* of the implications of that model e.g. by exploring its state space. Techniques such as *fault-tree analysis* (FTA) or *failure-mode, effects and criticality analysis* (FMECA) largely concern exploring an existing model (e.g. one given by plant diagrams, a circuit board design). Conducting *Hazard Operability studies* (HAZOP) is an interesting technique, in that it combines the discovery of the appropriate model with an exploration of the implications of that model: in undertaking a security-informed HAZOP, the attacks or failure of the system may come from a much lower level of abstraction than originally chosen, or the connectivity in the system might be different from that assumed. PIA is similar in this regard, combining model discovery and an exploration of model consequences. PIA offers more, however. For, while some interdependencies *are* obvious once the model is scoped correctly, many other behaviours are complex, requiring a detailed simulation-based approach to explore not just the consequences of the model but also the impact of *uncertainty in model structure*.

3. Method: Preliminary Interdependency Analysis (PIA)

Preliminary Interdependency Analysis (PIA) is an analysis activity that seeks to understand the range of possible interdependencies and provide a justified basis for further modelling and analysis. Given a collection of CIs, the objectives of PIA are to develop and analyse, through a process of iterative refinement, an appropriate model for the infrastructures, and to document assumptions about resources, environmental impact, threats and other factors.

The context within which PIA models are developed, and analysed, is defined by a scenario and related requirements. Here, the narrative aspect of a scenario is enormously important, as it provides the basis for asking questions and discovering interdependencies; typically, this is the starting point for the use of more formal models.

Figure 1 illustrates how one might start constructing, say, a service model, and identifying interdependencies between the services. Each of the services is likely to consist of various components. In Figure 1 we show two services with their respective *Information and Communications Technology* (ICT), components, networks, and information assets. *Some* reasons why interdependencies may exist between services include:

- *functional dependencies*, i.e. a service consumes the output of another service as either input (e.g. oil is used as a raw material in a chemical plant and is subjected to transformation) or resource used (e.g. fuel for heating, or power for communication equipment);
- *similar components*, e.g. ICT components used in multiple services, via which common cause/mode failures may lead to simultaneous or related failures of the services (e.g. a virus may affect the computers running the same standard configuration of OS/applications used by different services providers);
- *common environment*. Stressful conditions in the environment are likely to increase the likelihood of failure or cause a failure of components in different services. Spatial dependencies are a typical example here, but one can easily envisage other forms too, e.g. services use the same cloud provider for their IT operation which will lead to simultaneous impact on both services if the cloud provider is stressed.

PIA allows for model refinement by revisiting earlier stages in the PIA process in the light of the outcomes of latter stages. For example, an initial application of PIA should result in a sufficiently concrete and clearly defined

model of CIs (and their dependencies). However, an analysis of the model could lead one to question the assumptions made and, as a consequence, the model should be revised and refined. As we shall see later, revisiting previous phases of the model development process is a key aspect of the PIA method and philosophy, overall.

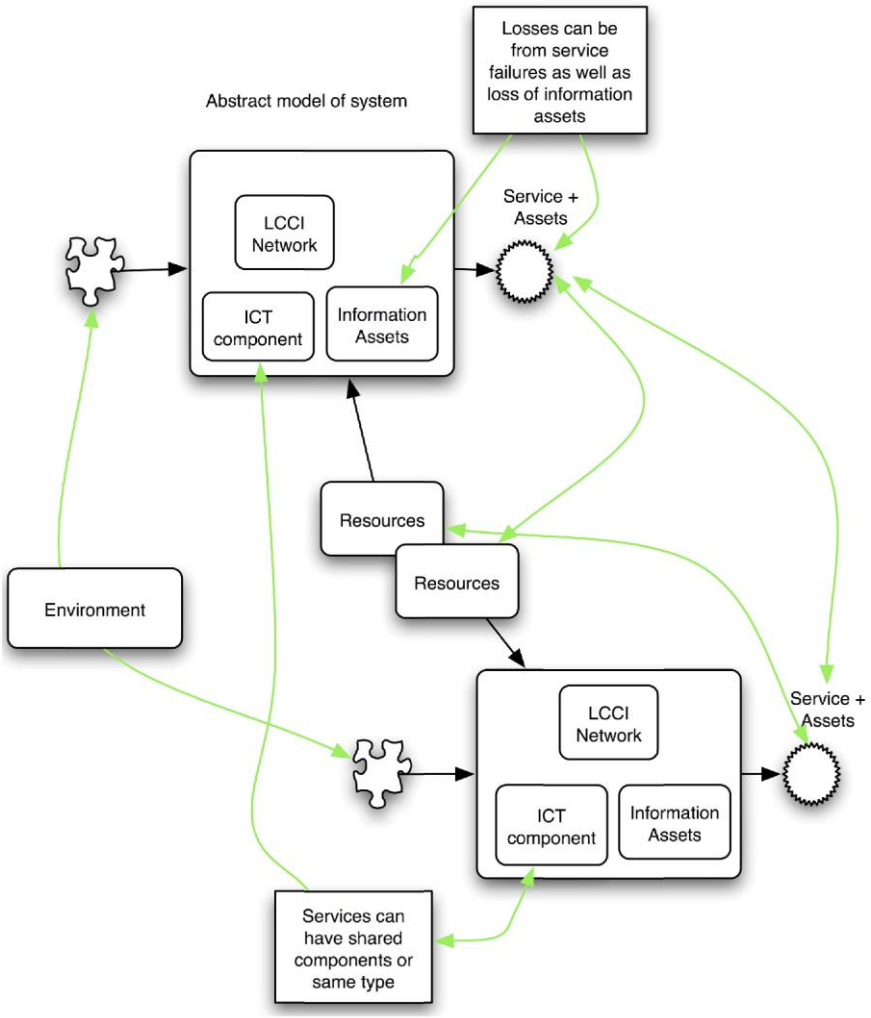

Figure 1: PIA sources of dependencies between resources

PIA is both qualitative and quantitative:

- Qualitative analysis. The modelling exercise begins with a definition of the boundaries of the system to be studied and its components. Starting off at a high level, the analyst may go through a cyclical process of refinements, but may also be focused on a particular service, so the level of detail may vary between the different parts of the overall model. The identification of dependencies (e.g. functional or due to shared environment) will start at this point.
- Quantitative analysis. The models created during the qualitative PIA are now used to construct an executable, i.e. a simulator of the model behaviour in the presence of failures/disruptions of the *modelled entities* – hereafter referred to as MEs – for the chosen model parameterisation. The model parameterisation may be based either on expert judgement or on analysis of incident data. Examples of such data analyses and fitting the available data to plausible probabilistic data models are discussed in [30].

3.1. PIA Model Architecture: Two Levels of Abstraction

PIA models broadly operate at two distinct levels of abstraction:

- *High Level Service Model (HLSM).* At this level the CIs are modelled as a set of interdependent services and the environment in which they operate. Here, the view is purposefully abstract, so that we can reason about dependencies between the services (e.g. data centre X depends on power plant Y), or between the services and the environment (e.g. the power system depends on the weather or the ICT system may be

affected by cyber-attacks). Now, each service is, itself, a complex collection of interacting, dependent components. Although the constituent elements of services are not explicitly modelled at this level (for such detail, see DSBM below), the inter-service dependencies *are*, and they are deduced from defined lower-level dependencies/relationships amongst the services' constituent entities (physical components, resources etc.). We refer to such "cross-service" associations between components as *coupling points*. The coupling points incoming to a service can be related to the resources that the service requires (e.g., a telecommunication service consumes "commodities" supplied by a power service). The resources consumed by a service can be obtained from the organisation's reserves (internal resources) or provided by another organisation (external resources). The outgoing coupling points, instead, define how the outputs from a service get consumed by other services (as either inputs or resources). Similarly, the dependence of the services on the environment can be defined in terms of coupling points between the environment and the services (e.g. "cyber-attack perimeter" can be defined in terms of points via which an adversary can attack an ICT service).

- *Detailed Service Behaviour Model* (DSBM). Here, implementation and behavioural details are provided for each individual service – this includes its underlying processes, constituent components and their relationships with one-another, relationships amongst components from different but dependent networks. For instance, a *Global System for Mobile* (GSM) telecommunication operator typically relies on a network of devices deployed to cover a particular area (e.g. masts, etc.). Via DSBM we can choose the level of detail used to model these networks. In the example above DSBM may range from a connectivity graph – representing how the components of the network are connected to each other – to a high fidelity model of the protocols used in the GSM network. We tend to think that DSBM models the networks owned (at least partially) and/or maintained by the respective service operator, i.e. an organisation. Although such a view is not necessary, it allows one to model several important aspects via DSBM. For instance, the level of investment and the culture within the organisation (e.g. strong emphasis on engineering vs. outsourcing the maintenance) will affect how well the network is maintained, which in turn will affect the frequency of outages and the speed of recovery. Thus, the speed/rate of recovery (a parameter that can be used in a DSBM) can be a useful proxy of the level of investment by the operator. In other words, through DSBM one can describe and study interesting scenarios which at first may seem outside the scope of PIA. An example of such a scenario would be a study of the impact of deregulation in a particular critical CI given the current or projected interdependencies with other CIs.

3.2. The PIA Process

Our experience with PIA [31-33] indicates that it can be applied in the following stages (for instance, see Appendix A for a detailed account of how we applied each of these stages. Figure 2 gives a pictorial overview):

1. *CI description and scenario context*. A CI description provides a concrete context and concept of operation. This is the first level of scoping for the analysis task; the CI description gives the first indications of analysis boundaries. DSBM entities are identified and recorded and the overall CI services defined. For example, the context of the case-study used to demonstrate PIA in this paper was provided by the real life flooding of a telecommunications node in the Rome area [28]. The flooding ultimately resulted in a loss of communication between two SCADA control centres in the local power network. In total, the incident involved components from 5 interdependent CI: 2 power networks and 3 telecommunication networks. To better understand the interdependencies involved in this incident, and interdependencies involved in incidents not yet seen, these 5 CI provide a natural scope for the study. In fact, the idea of scenario is still key here: once a well-scoped model of interacting CI *has* been built, one can seek further insight by asking questions of this model – questions phrased in new scenarios/contexts. For instance, in other work using PIA, we augment a well-known power network model (called NORDIC-32) – created for other contexts/scenarios – and use this augmented model in cyber-security research [31, 32].

2. *HLSM Model development*. A model of the services (resources, inputs, outputs, system states), the operational environment and system boundaries are developed, based on the CI description. Model boundary definitions are used at this stage to further restrict the scope of the analysis. An initial identification is made of dependencies between the services via the coupling points as defined in the section 3.1 above.
3. *DSBM model development*. DSBMs are defined by selecting the right level of abstraction for the services: all services are modelled as state-machines which, as a minimum, consist of "failed" and "Ok" states. In this case, their representation in the DSBM will require no further refinement. For those services, however, which are modelled in more detail, the state-machines modelling their behaviour may be significantly more complex. We start by defining, explicitly, their components and assets, and may resort

to using existing models of the underlying physical networks used by the services. These networks may include a number of components, which we call *modelled entities* (MEs). Similarly, the environment model is detailed to account for the particular threats to be included in the study. For instance, a state-based adversary model can be defined to include a number of modelled entities [31, 32]. A level of consistency is achieved between the service model (Stage 2) and DSBM: the coupling points appear in both views.

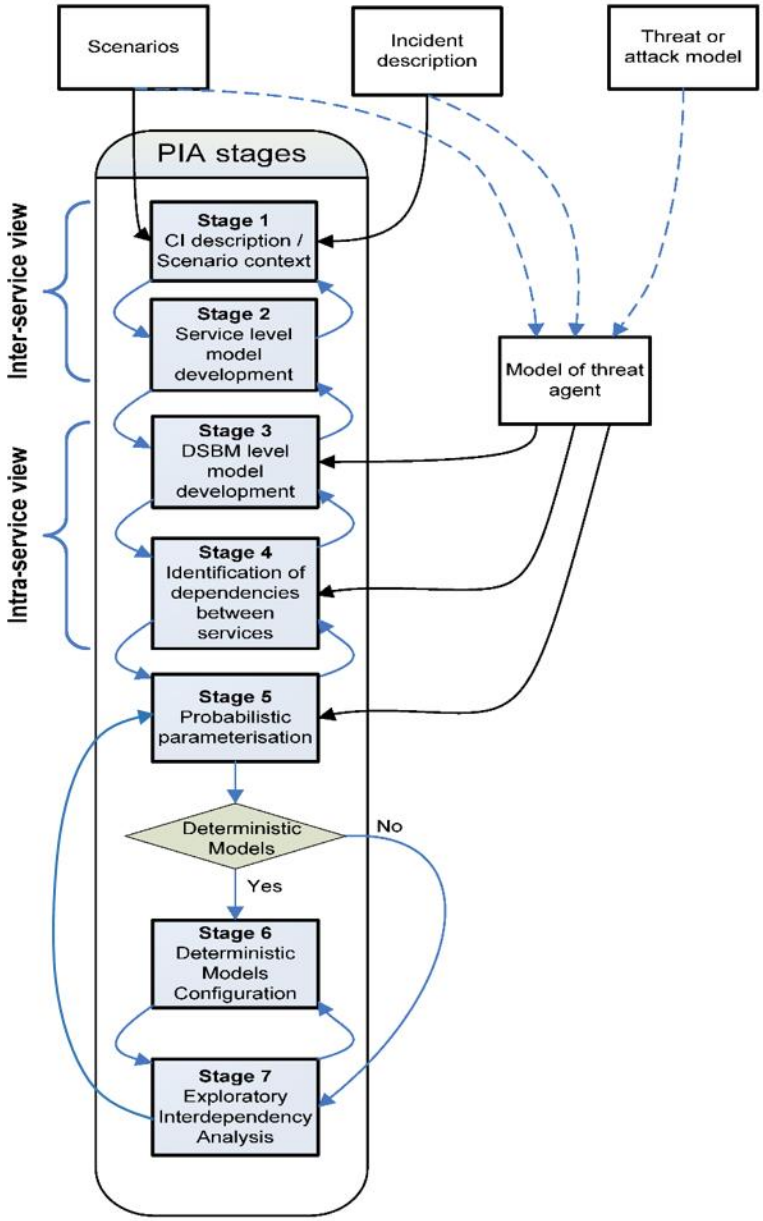

Figure 2: PIA method stages

4. *Initial dependency and interdependency identification*. While some of the service dependencies have already been identified and recorded in Stage 2 (via input/output/resource identification), at this stage the modeller looks for additional sources of dependence (e.g. common components/assets), which may make several services vulnerable to common faults or threats. These can be derived by examining the service-level model, taking into account other contextual information (e.g. scenarios, threat models, attacker profile). The captured dependencies are modelled as *stochastic associations* between the services or components thereof. This is detailed in section 4 below.
5. *Probabilistic-model development and parameterisation*. Since we are dealing with risk we take the view that, given the state space formed by the MEs, a stochastic process [34, 35] must be constructed that captures the state-changes and interactions of the MEs over time. Then, via simulation, risk is quantified as a suitable *random variable* – this is a real-valued function of the realisations of this stochastic process defined to estimate the risk of some unwanted events. We shall refer to both these random variables and

their related probability distributions as *risk-measures*. Now, in principle, *any*[3] risk-measure of interest may be defined by an assessor. And, depending on an assessor's requirements and preferences, she may define multiple risk-measures for the same model to address different aspects of risk (e.g. risks associated with component unavailability, the size of cascades when they occur, etc.) To this end, at this stage of PIA, a stochastic process is defined in terms of the *Stochastic Activity Networks* (SANs) formalism [36] and the theory of *Competing Risks* [37], as well as any risk-measures of interest. Examples of such definitions are given in sections 4 and 5.

6. *Deterministic models configuration* (*optionally adding deterministic models of behaviour*). Our earlier work established that purely probabilistic models can be inadequate in capturing essential system properties, otherwise captured by hybrid models [38]. So, a modeller might choose to extend the behaviour of a probabilistic PIA model by adding to it deterministic models of behaviour. Such a step is useful when seeking to extend the fidelity of the simulation beyond standard mechanisms possible with a purely probabilistic model. Furthermore, such extensions can be used to study the impact of the level of abstraction/fidelity on the modelling results, an important aspect of model validation. Examples of deterministic models include various flow models (e.g. AC power-flow, models of fluid flow, network traffic and transport flow models), state-estimation schemes (e.g. Newton-Gauss based methods used in power system analysis), network connectivity models, etc.
7. *Quantitative interdependency analysis*. A Monte Carlo simulation [39] is used to quantify the impact of the interdependencies on the behaviour of the system under study and draw conclusions about the interdependency-related risks.

During these stages we found that narrative information is very relevant and useful. It usually comes from the following sources:

- Scenarios: PIA is a scenario-driven approach. Once the system has been modelled, "what-if" questions will be used to explore vulnerabilities and failure cascade possibilities. Scenarios can be developed from a variety of assumptions or experiences. For instance, one can begin by asking a question as abstract as "what happens if there is a flood", or "if power plant X fails". Such questions form the basis for scenarios, which focus the analysis on particular conditions, exploring potential vulnerabilities.
- Incident description: PIA can be used to model an incident that has already occurred. This can be used as a baseline for generating and exploring variations of the same scenario or simply further exploring a system that has been compromised, or has failed, as the incident revealed unknown vulnerabilities and failures.
- Threat or attack model: here, we are considering modelling assumptions based on accidental failures or malicious attacks.
- Model of threat agent: The above (scenarios, incident description, threat or attack model) are elements that will shape the profile of a threat that is modelled in our system. This can be a source of natural disaster (e.g. flood) or a malicious agent (e.g. a terrorist).

4. Quantitative Dependency Modelling

In this section, we describe PIA's quantitative modelling approach (underlying stages 5-7 of the methodology) which captures various ways in which network components might be dependent on one another. Sources of dependence between network components *can* include the following:

1) *The Physical Network Topology*: Components are sometimes reliant on being physically connected to each other in order to receive a resource or perform some function. Therefore, via these physical connections, the failure of one component can impact the operation of another component. So, for example, medium-voltage-trunks are used to provide electrical power from the Power distribution network to Telecommunication network sites: damage to the trunk can result in service disruption from Telecommunications network. Also, physically co-located network components, such as those power network components situated at a power substation, might collectively be affected by the same disturbance/event, for instance local flooding or forest fires.

2) *Functional Relationships between components*: Network components can be related because they are "coupled" in the function they perform. So, for instance, when a powerline is tripped in the Power Transmission Network there is a redistribution of power flow across the network. This inevitably affects the quality of service provided at various points in the network, such as the local amount of load required at a given point. Another example can be seen in the Telecommunications network, where backup power generators supply multiple Telecoms components with power in the event of a power cut from the Power

[3] Limitations to this primarily stem from whether enough state-information can be included in a model and simulated.

distribution network. Potentially, if the generators become unavailable, multiple Telecoms components stop functioning.

3) *Stochastic Correlation*: The state changes of components may be observed to be correlated in terms of, both, what new states these components enter into, and, when these state changes occur. Here, we note that such correlation could be exhibited by scenarios such as those pointed out in bullet points 1 and 2 above. However, we also acknowledge other possibilities, such as components experiencing synchronised falls and peaks in the quality of service they provide due to phenomena known to occur in technical systems, e.g. *common-mode* failures, and common stress on network elements resulting from extreme weather, natural disasters or new computer viruses.

The model we describe here adequately captures each of these examples. Our exposition on how we model dependence is carried out in two steps: firstly, we give the definition of an individual isolated ME's behaviour (thereby effectively ignoring other MEs) and, secondly, we extend this behaviour to take into account how changes in the state of the ME affect, and are affected by, the state changes of other MEs. Formally, the resulting model is a generalisation[4] of a *Continuous–time Semi–Markov process* [34, 35] which, in order to simulate, we implement as a SAN in the Möbius modelling environment [8, 36].

We begin with an individual ME; it experiences a state change with probability, according to an appropriately defined *Competing Risks* [37] model. To illustrate, given that the ith ME can be in any one of the M_i possible states $\mathcal{S} := \{s_1^i, \dots, s_{M_i}^i\}$ at a given point in time, suppose that the ME enters into a state $s_{j_i}^i(t_0) \in \mathcal{S}$ at time t_0 (where $j_i = 1, \dots, M_i$). *Competing Risks* then requires that each state the ME could potentially enter into next has an associated probability distribution for how long it could take to enter into that state. To determine which state will actually be entered into next, each of these potential time lengths is randomly generated according to these distributions, and the minimum of these generated times is defined to be how long it will actually take for the ME to change state (this is the MEs *sojourn time* in the state $s_{j_i}^i(t_0)$). Since each of the generated times is associated with a unique next state, this minimum also determines what the next state will be[5].

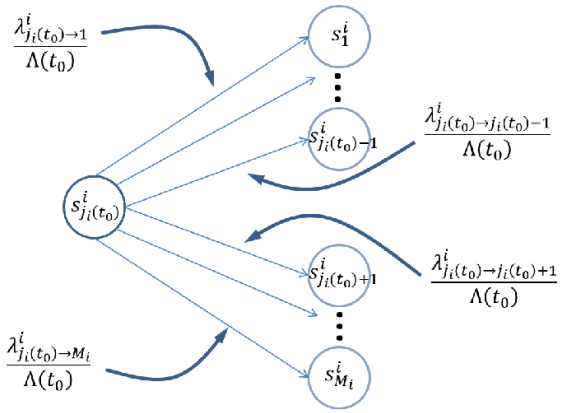

Figure 3: Using a competing risks algorithm, an ME transits, with probability, from a state $s_{j_i}^i(t_0)$ at time t_0 to some other state at some future point in time. In this example, each transition's associated probability of occurrence is a quotient (such as $\frac{\lambda^i_{j_i(t_0)\to 1}}{\Lambda(t_0)}$) of transition rates for exponential distributions.

Pictorially, an example of this process is shown in Figure 3 for the ith ME at time t_0. Here, each of the potential transitions from the initial state $s_{j_i}^i(t_0)$ to another state in $\mathcal{S}$ is represented as an arrow connecting a pair of states. In the particular case where each transition starting at time t_0 has an exponentially distributed potential length of time, as depicted in Figure 3, our model becomes equivalent to an appropriately defined *Continuous-time Markov Chain* [34, 35]. More generally, *any* class of sojourn-time distribution may be used in PIA, as long as it can be efficiently sampled from.

So, *Competing Risks* determine both the next state for the ME and how long it will take (from time t_0) for this state to be entered into. This procedure is then repeated each time the component enters into a new state, resulting

[4] A generalisation, in that the process has both *instantaneous* and *time-consuming* sojourns in system states.

[5] For our purposes, in order for the next state to be uniquely defined, we require that this minimum must be associated with only one of the potential next states. That is, we exclude the possibility that two or more of these generated time lengths can have the minimum value. This limitation is not serious in practice. Should the need to resolve it explicitly occur the next state can be chosen at random from those which produced the same shortest sojourn time.

in the ME evolving, over time, from state to state as depicted in Figure 4.

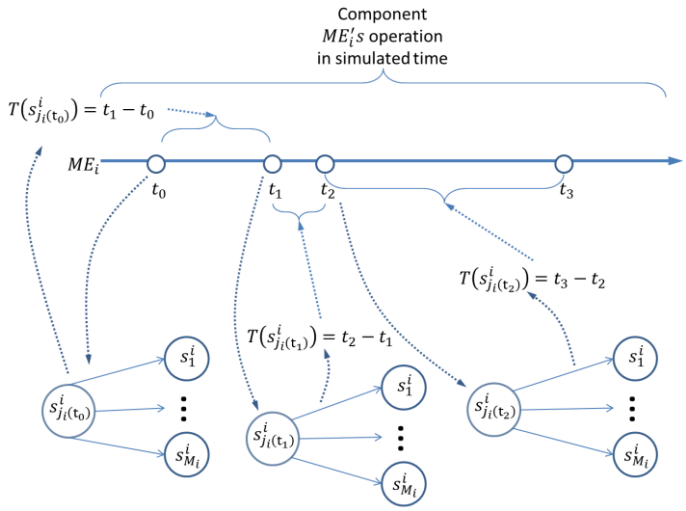

Figure 4: The evolution of the ith modelled entity's state in time is accomplished by applying the competing risks algorithm to the entity at some time t_0, resulting in the ME changing state from $s^i_{j_i}(t_0)$ to state $s^i_{j_i}(t_1)$ at time t_1 (that is, the ME remains in the state $s^i_{j_i}(t_0)$ for a time $T(s^i_{j_i}(t_0)) = t_1 - t_0$), at which point the competing risks algorithm is applied again to determine the next state change.

The dynamical ME model outlined thus far can be extended to capture interdependencies between MEs. Intuitively, the evolution of an ME's state at a given point in time is dependent on the current state of the ME, how it got into that state, the current state of other MEs and their respective evolutionary histories. In particular, a dependent ME – hereafter called a "child" – probabilistically changes state in a way that depends on another ME – hereafter called a "parent". For brevity, we shall use "*Stochastic Association*" to refer to the triplet of a given ME, the parents of the ME, and a definition of how these parent MEs' states and history determine the probabilistic law (i.e. determine a member of a family of probability distributions) that governs the stochastic behaviour of the child ME[6]. In this way, each ME can potentially have several parents and several children, and the MEs are made probabilistically dependent on each other (so, an ME can be both parent and child of another ME). This allows one to model rather complex failure and recovery behaviours, for instance. A simple example of a network consisting of MEs with stochastic associations defined between them is depicted in Figure 5. In the figure, the MEs' states evolve as time flows from left to right. Solid circles indicate points in simulated time when state changes occur and, as a consequence, next states and/or new sojourn-times are computed as required for some MEs. Dashed circles indicate as yet unrealised potential state change events. So, ME_1 enters into state $s^1_{j_1(t_1)}$ at time t_1 and, therefore, a new next state and sojourn-time (of duration $t''' - t_1$) are computed for ME_1 at that time. However, in determining these, the state of its parent, ME_2, is used. That is, $T(s^1_{j_1(t_1)}, s^2_{j_2(t_1)})$ – the sojourn-time for ME_1 – is a function of both ME_1 and ME_2's states at time t_1. And, when ME_3 changes state at time t_2 it, along with its child ME_2, requires new next-state and sojourn-time computations, resulting in the model experiencing its next state change event at time *t*. Note, however, that ME_1 is not affected since ME_2 did not experience a state change, in accordance with the stochastic association defined between ME_1 and ME_2. During simulation, this particular behaviour is achieved by using the SAN mechanism of "*reactivation*" [36].

The specification of a collection of stochastic associations is equivalent to defining, for each point in simulated time, a joint probability distribution over the Model's state-space, where this state-space is the so-called *cartesian product* of the individual ME state-spaces. With this in mind, defining stochastic associations is quite convenient when compared with the alternative of explicitly defining the related joint distribution. For models with a sufficiently large number of MEs, the explicit specification of such a joint distribution – say, as a suitable *markov chain* – would be both daunting and tedious, and may require unfeasibly large state-transition diagrams. And, even if accomplished without errors, it is likely that such an explicit representation will not be usable when attempting to simulate the model: such a model, even of modest size in terms of the number of modelled elements,

[6] A generalisation of this would be to define, explicitly, stochastic associations between groups of MEs, e.g. instead of modelling the effect of each parent node on its children independently, one could define the effect of a collection of parents on a collection of children. Doing this, however, might significantly increase the number of model parameters.

could have a state-space that is too large to fit in any available computer memory[7].

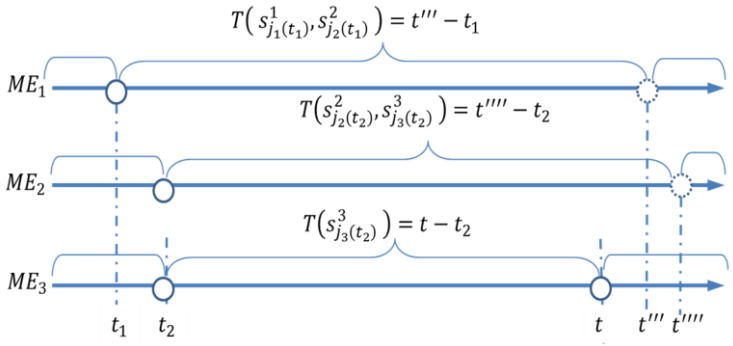

Figure 5: A Network of modelled entities subject to stochastic associations. ME_1 is a child of ME_2, itself a child of ME_3. Consequently, upon entering the state $s^1_{j_1(t_1)}$ at time t_1, the random variable that is the time until ME_1 experiences its next state change is defined by both the state of ME_1 and the state of its parent, ME_2, at time t_1. We emphasize this dependence by writing the random variable as $T\left(s^1_{j_1(t_1)}, s^2_{j_2(t_1)}\right)$. The dependence of ME_2's sojourn-time on ME_3's state is similarly illustrated.

Each MEs evolution is, therefore, determined by marginalising the aforementioned joint distributions. In general, different forms of these marginal probability distributions can be specified, depending on what is suitable in a given modelling situation. For instance, for our experiments, MEs such as *fibre-optic cables* had the distribution of their next state completely defined by conditionally independent, exponentially distributed random variables, conditional on the states and history of the MEs parents (in accordance with the general depictions in Figure 3 and Figure 5). More detail on this is given in section 5.

The collective evolution of all MEs is itself governed by a *Competing Risks* algorithm. At each point in simulated time, each ME has a potential sojourn-time at which it could experience its next state change. So, whenever an ME changes state, the time at which the next state change occurs in the model is determined by the minimum of these potential sojourn-times across all MEs. Recall, whenever a parent enters into a new state, the potential sojourn-time before the parent next experiences a state change is computed. In addition, the potential sojourn-times before each of the parent's children next experience state changes are also recomputed, in a manner dependent on the new state of the parent. All of these sojourn-times are compared with the sojourn-times associated with the other MEs, in order to determine the time of the next state change in the model. For example, in Figure 5 where, upon ME_3 entering into state $s^3_{j_3(t_2)}$ at time t_2, the potential sojourn-times for both the parent ME_3 and its child ME_2 are computed and compared with the potential sojourn-time (previously computed at time t_1) for ME_1. The minimum of these sojourn-times is associated with ME_3, implying that the next state change will be experienced by ME_3 at time t.

Modelling MEs as being probabilistically dependent may be justified by identifying functional relationships between the components. Examples of these abound. For instance, in a power transmission network, those components that facilitate the supervision and control of power flow across the network (such as *Remote Measurement Units* or *Intelligent Electronic Devices*[8]) rely on ICT for their operation: lack of control may lead to an increased likelihood of the controlled components failing. Or, in a Telecommunications network, there is a reliance of ICT components (such as routers and add-drop multiplexers) on a stable source of power provided by other components: loss of power results in the components being inoperable.

Note, however, that the modelling requirement here is one of *probabilistic dependence* between the MEs, and *not necessarily* one of an observed causal relationship. This is convenient from a modelling perspective, since it is quite possible that in practice there might not be an immediately observable causal link between certain real world entities, and yet such entities may still be observed to exhibit some form of correlation between their state changes. While, upon extensive investigation, such correlation may be determined as being due to some common-cause, a determination of this kind is not needed to justify modelling these entities as dependent MEs.

[7] For illustration, a model with a 1000 MEs, where each ME has an associated state-space size of 2, will result in a model state-space with 2^{1000} states – a number that is easily greater than 10^{82} which is the estimated number of atoms in the observable universe.

[8] Here we use the terminology established by IEC 61850.

Depending on the time-scales over which typical events of interest occur in the model, an extreme case of a stochastic association would be an instantaneous (i.e. a particular form of deterministic) state change of a child ME as a result of a parent ME's state change. This can be modelled either stochastically (with instantaneous transitions) or by using deterministic models. An example of the latter might be using a power-flow model with MEs which are elements of a power grid. In this case, the random failure of some MEs, e.g. a power generator, may deterministically lead to insufficient supply of energy and, as a result, to load shedding, i.e. from the random failure of an ME, the state change of some lines will follow deterministically (with a duration that is either instantaneous or has some delay). Another example from a power grid might be the random failure of some power line (e.g. due to electrical shorting in extreme weather) resulting in an immediate and deterministic redistribution of power flow (according to a power-flow model) and the overloading of other power lines. As a consequence, these overloaded lines might become disconnected. Of course, when these components change state, the probabilities of when and how their stochastic children will next experience change are altered as well. In this way, there is a continual "dance" between stochastic and deterministic events in our models.

To finish this section we note, in passing, that an important aspect of model-building is model parameterisation; a problem which we address in sections 5, 6.3 and Appendix A.

5. A Multi-infrastructure Case Study: the Rome System

The PIA method was applied to a complex case study of a regional system of two critical infrastructures in the Rome area: the power grid and the telecommunications. Please see Appendix A for details of how the PIA process was applied. The power grid infrastructure includes two services: the high voltage transmission (150kV) and the medium voltage distribution (20kV), each with their own networks; the telecommunication infrastructure includes the 3 services with their respective networks: the fibre-optics backbone, the fixed lines service/network and the GSM mobile service/network and their interconnection.

The case-study was originally developed within the European Integrated Project IRRIIS [40] with the help of the actual network operators of the two modelled infrastructures, and this was further developed in the United Kingdom Technology Strategy Board (TSB) funded PIA-FARA project [41]. The model includes some 829 MEs (119 in the power infrastructure and 710 in the telecommunications infrastructure) and closely represents the topology of the real infrastructures in the area of Rome, Italy [40]. Details of how the PIA method was applied to the Rome case study are provided in the Appendix A.

Each ME is modelled as being in one of two possible states – *Failed* or *Ok* – at any given point in time[9]. As we indicated in section 4, a stochastic association requires the suitable definition of a collection of marginal probability distributions that determine a child MEs next state during simulation. In our experiments, we chose most of the MEs to transit from state to state according to exponentially distributed sojourn-times between state changes. The rates used in the computation of these times are determined as follows: suppose λ is the failure rate of an ME when all of its parents are in an *Ok* state. If, instead, n parents are in a *Failed* state, then the failure rate becomes $\alpha^n\lambda$, where $\alpha \geq 1$. The scaling factor α is a model parameter that indicates the "strength" of the stochastic dependence between the MEs[10]; varying the value of this parameter results in changes to the probabilistic behaviour of the MEs, in particular, and the model as a whole. This idea has some similarities to the approach to modelling dependent components presented in [42]. The use of α to model stochastic dependence in this way is merely one example of the many kinds of stochastic association PIA affords us – an example that is, of course, less challenging to calibrate than one with significantly many more parameters. We encourage users of PIA to come up with stochastic associations suitable for their particular needs.

The use of α allows us to introduce nonlinearity in the models of failure (consistent with observations made in previous studies e.g. [43]). Indeed, a small number of failed modelling elements may lead to a dramatic increase in the propensity of many other elements to fail, hence increasing the chances of large outages.

The values of α may differ across the "parent-child" element pairs in the model. Eliciting the α values is difficult; possibly infeasible. Therefore, we dealt with this problem by systematically applying sensitivity analysis under the assumption that the strength of dependence is the same for all "parent-child" pairs. We conducted 5 simulation campaigns using the α values 1, 10, 100, 150, 250 and 500, which are only a sample from the

[9] This state space is sufficient for our preliminary analyses as this is the simplest set of ME states that may be used to capture phenomena such as cascading failure and system recovery.

[10] Various generalisations of this concept are possible, such as requiring possibly unique scaling factors for each pair of MEs or between unique groups of MEs. Doing so can increase the richness of the model behaviour, but at the possible expense of making parameterising the model more challenging.

plausible range of α values studied. We chose not to investigate the model behaviour for values of α outside this range because, for values of 500 and above, the number of simultaneous component failures was too high and, clearly, unrealistic.

The parameterisation of the model (i.e. both the failure and repair rates, as well as the characteristics of each ME) was provided by the industrial partners in the IRRIIS project [40]: data on the power grid was provided by engineers from Siemens, the telecom data was provided by engineers from Telecom Italia. When data was needed we made informed estimates and checked their plausibility with these subject matter experts[11].

As explained earlier in stage 6 of section 3.2, in addition to the probabilistic behaviour of the MEs, deterministic effects may be included in the model. In the Rome study, we used dc-load flow computations to determine the redistribution of power when components fail in the power network. Also, in the event of an outage of the main power supply, the Telco network uses backup power supply units, such as batteries or generators.

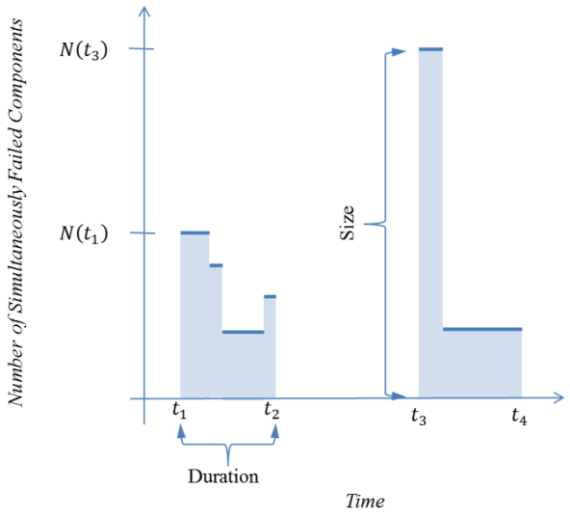

Figure 6 : This illustrates the evolution of two cascades. Each cascade has a duration and size that are random variables; only by observing a cascade to its conclusion can one know what its size and duration is. Each of these random variables can be used to define *risk-measures* related to cascades: that is, they define probability distributions (e.g. of cascade-size or cascade-duration) associated with the cost of a cascade's occurrence.

The model was used to simulate the CIs operating over an 11 year and 4 month period[12]. During operation the MEs that make up the CIs experienced failures and repairs. To illustrate the sort of analyses that is possible, we chose to focus on the occurrence of failure cascades, how long these last for (i.e. their durations), and the maximum number of simultaneously failed MEs involved (i.e. their sizes) as shown in Figure 6. A different focus merely requires the definition of some other risk-measure of one's choosing. The figure illustrates two cascades separated by a period of normal operation of all MEs. $N(t)$ is the number of simultaneously failed components at time t. A cascade is defined as any continuous period of time for which at least one ME is in a failed state[13]. Two measures of interest related to cascades are cascade-size and cascade-duration, both depicted here. The cascade-size is the maximum number of simultaneously failed MEs during a cascade (e.g. the sizes of the depicted cascades are $N(t_1)$ and $N(t_3)$) while the cascade-duration is the length of time for which a cascade occurs (e.g. the durations of the depicted cascades are $t_2 - t_1$ and $t_4 - t_3$).

[11] Examples of unavailable data that required estimates validated by expert judgement include the failure and repair rates of various powerlines connecting telecommunication nodes to secondary power sources (diesel generators and batteries)

[12] This is a duration equal to 100 000 hours (approximately 11 years and 4 months).

[13] While this definition of cascade *does* include trivial cases, such as the failure and recovery of a single ME, its real usefulness lies in the fact that it also encompasses so many interesting dynamic sequences of the failure and recovery of groups of MEs. For instance, a cascade can consist of a sequence that begins with 5 ME failures, followed by 4 recoveries (so, at this point, only 1 component is "failed"), followed by 1 new failure, followed by…and so on, till there are no failed MEs. This definition also has the advantage that it appears to be readily applicable to describing "cascading" phenomena both within and across very different CI, partly because it does not rely on causality to explain these failure and recovery sequences. Of course, where PIA is concerned, this is merely *one* choice of risk-measure – *any* more suitable risk-measure should be used when required.

6. Results: Illustrative insights from the case-study

Using the *PIA toolkit* described in section 1 above and the stages detailed in Appendix A, the Rome scenario model was developed and used to explore systemic risks[14] of cascades.

6.1. Network Resilience

Typically, power networks are operated with the *resilience* requirement that they should tolerate single faults[15] [44, 45]. We observe this behaviour with the modelled power network (see Figure 7), with most of the observed failures in the network being single, isolated failures that do not result in cascades. The risk-measure used here is the distribution of cascade-sizes.

There appears to be a critical number of MEs (approximately 36) beyond which almost all cascade sizes are *spectacular* – the collapse of the entire power network. This property of the model is a consequence of both the level of functional redundancy in the model and the model parameterisation, both chosen by the IRRIIS consortium [40] to simulate the power network operating at close to its operational capacity. However, this model behaviour is consistent with the findings of [46] who report a nonlinear relationship between how much of the transmission lines' capacity is used and the resultant distribution of cascade size in their model: when transmission networks utilise above 80% of the lines' capacity, the risk of large cascades is significantly greater than when less than 80% is used.

Note, from the distribution of cascade-sizes associated with $\alpha = 1$ in Figure 7, that while the probability of spectacular cascades (e.g. approximately 10^{-2} for cascades of size 119) is significantly smaller than the probability of single failures (approximately 0.95) it is, however, of the same order of magnitude as the probability of significantly smaller cascades (e.g. cascades of size 10). This hints at the possibility that the failure of only a few MEs is sufficient for spectacular cascades; behaviour that is evidenced by the circumstances under which spectacular blackouts have occurred in practice [47].

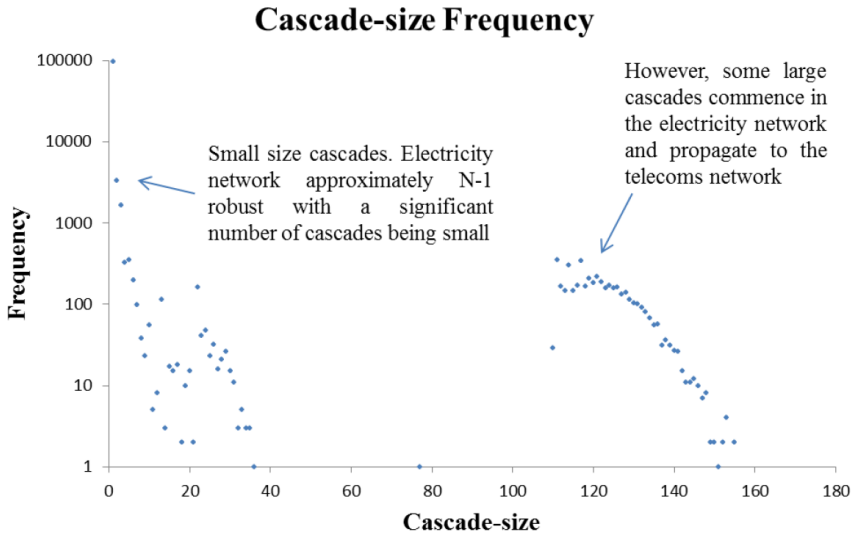

Figure 7: This risk-measure is the distribution of cascade-sizes across the combined network consisting of both the Power Critical Infrastructure and the Telecommunications Critical Infrastructure. The value of the stochastic dependence scaling factor is $\alpha = 1$.

Maintaining a network at a desirable level of resilience often involves decisions about how best to allocate finite resources. It seems reasonable that some ranking of the MEs in order of importance might be a useful first step[16] in making such a decision. Using appropriate risk-measures, there are a number of alternative approaches to such

[14] By this, we mean the risk imposed by interdependencies in a system, where the failure of a single entity or a group of entities can cause cascading failure.

[15] In power systems this is known as the "N-1 criterion".

[16] Further investigation may be required. For instance, given a statistical identification of a subset of MEs as being critical and worthy of attention, further investigation might reveal important causal relationships between these MEs, suggesting that limited resources can be targeted to an even smaller subset of these MEs.

a ranking. For instance, should components be ranked according to their reliability or, instead, should they ranked according to how likely they are to be involved in a large cascade? And, are alternative rankings related in some identifiable way? Perhaps there is a sense in which some measures are indicative of other measures, but are significantly easier to compute. We chose to study the following three measures, each capturing a different sense of "neighbourhood" or "locality":

1) *component (un)availability*: the unavailability of each ME, which focuses on the behaviour of individual MEs alone,
2) *component connectivity*: the number of immediate neighbours an ME has and, in this sense, is slightly less local than ME unavailability, and
3) *cascade membership*: the probability that an ME is contained in a cascade of at least size[17] 117, a relatively global measure that can take into account MEs located across the entire network.

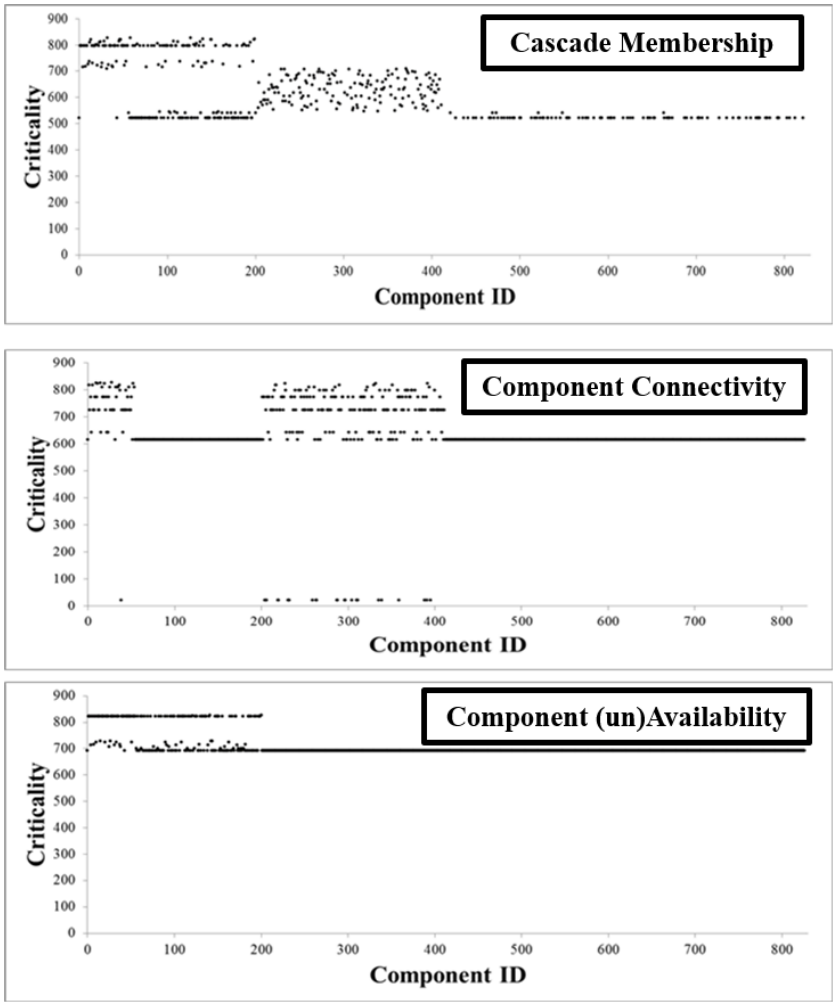

Figure 8: A comparison of three notions of criticality (with $\alpha = 1$). Each criticality notion results in a ranking of the 829 modelled components. Many of the components in each plot share the same criticality value[18], resulting in a wide vertical "gap", in each plot, within which no components assume any criticality values. The size and location of this "gap" differs across the criticality notions.

We rank the components according to their associated values for the chosen measures. Such a ranking gives a notion of criticality: from such a ranking we can identify which MEs are the most important and deserving of our attention.

The results of our investigation are shown in Figure 8. This consists of 3 plots – one for each criticality type. The

[17] In particular, for when the power network was operating close to its operational limits, 117 is the size of the largest cascade consisting of power network components alone.

[18] For example, the MEs have the same number of neighbours or are involved in the same number of cascades.

MEs are assigned unique IDs (all 829 of them) and, for each notion of criticality, the MEs are ranked according to their computed criticalities. Using these IDs, we depict the MEs criticality rankings, where each ME has three associated criticality values plotted – one in each of the respective plots. The unique IDs of the MEs are shown on each horizontal axis, with each vertical axis showing the possible criticality values. The most critical MEs have the highest criticality values.

We see significant disagreement between these three notions of criticality, with some MEs being important in terms of being very likely to be involved in large cascades, but unimportant in that they have relatively small downtime – that is, they are noticeably more available – than many other MEs.

This observation seems important as it emphasises how the ranking can significantly depend upon the chosen risk-measure. Our own experience with critical infrastructure analysis suggests that researchers and practitioners often use a given risk-measure to identify critical nodes without demonstrating an awareness that the ranking obtained thereby could depend significantly on the chosen measure. As a result, in many of these cases, the selection of a suitable measure appears to be a matter of convenience, lacking in proper justification for its use other than "this measure has been used elsewhere". Here, we show that the choice of measure does matter. By exploring alternative measures, an assessor can gain an appreciation of how the ranking changes and, thereby, choose those measures which are judged to best capture the objectives sought via suitable rankings. For instance, it seems quite clear that investing to improve components' individual availability (see "component unavailability" in Figure 8) is not necessarily a wise investment, if the actual effect sought is resilience improvement of the entire critical infrastructure.

6.2. *The Impact of Model Fidelity on Modelled Risk*

We studied how changing a model's level of detail and sophistication affects the risk of cascades occurring in the model by comparing two scenarios: one with relatively detailed power and telecommunication network models (referred to as the *full model*), and another in which the fidelity of the telecommunication network was unchanged but only those power network MEs which were directly physically coupled to telecommunication MEs were modelled (referred to as a *partial model*). In both scenarios a value of $\alpha = 1$ was used. Weibull distributions were used as the failure distributions for each power network ME in the partial model, the parameters of which were obtained by first estimating the marginal failure distributions for the respective MEs in the full model, and then fitting these distributions to Weibull distributions. A comparison of the cascade-size distributions for the telecommunication network in both experiments is shown in Figure 9.

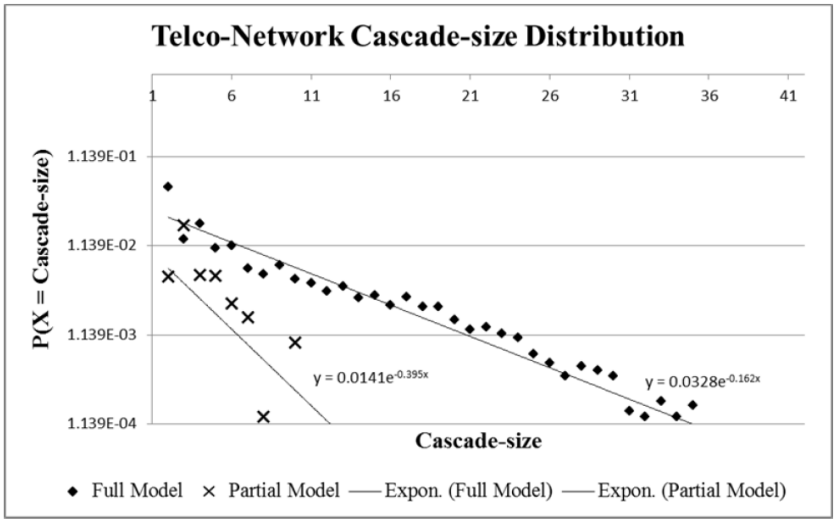

Figure 9: A comparison of the distribution of cascade-sizes in the Telecommunications Network under two levels of abstraction: 1) when the telecommunications network is coupled with a detailed model of the power network (full model), and 2) when only a simple model of the power network is used with everything else remaining the same (partial model). The value of the scaling factor for both models is $\alpha = 1$.

In particular, the partial model lacks correlated failure of power network MEs; a property which exists in the full model as a result of overloading power lines. That is, the partial model has independently failing power MEs. This means that geographically separated telecommunication sites experience power blackouts independently. We observed that the cascade-size distribution in the partial model underestimates the probability of large cascades,

with no cascades greater than size 11 occurring[19]. So, details about *how* cascades arise in the power network are clearly important for risk estimation in the Telco network. This result highlights a challenge for any interdependency analysis that seeks to estimate risk in a given CI *without* sufficiently characterising the joint uncertainty in the service delivered *at* the coupling points. In such cases, a priority must be the development and application of statistical techniques for inferring such uncertainty from limited (coupling-point) observational data.

6.3. *Stochastic Dependence Strength*

The strength of stochastic dependence, α, is a useful modelling device for introducing nonlinearity in system behaviour. Here, we give details of the sensitivity analysis we conducted – studying the behaviour of the model over plausible ranges of α values, using different measures. So far, the reported results of our analyses were all based on experiments which set $\alpha = 1$. Here, we show how varying the value of α affects three risk-measures of interest in the full model:

1. The distribution of cascade size;
2. The distribution of cascade duration;
3. The distribution of the "loss" due to cascade. For our purposes, we define the "loss" due to a cascade as the product of the cascade size and cascade duration.

It turns out that the model changes resulting from varying the stochastic strength show different patterns as the parameter α takes different values.

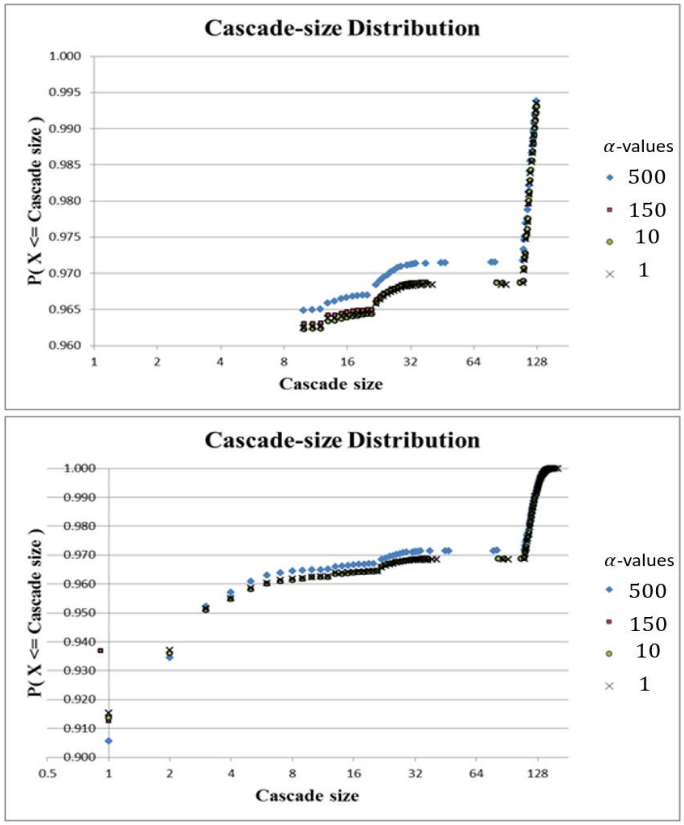

Figure 10: Cascade-size distributions for α values 1, 10, 150 and 500. The upper 95% of each distribution is depicted above, and the entire distributions are depicted below.

Changes in α result in changes in the distribution of cascade size (as defined earlier in Figure 6), but only over a very small range of probabilities (see Figure 10): the only differences occur above the 95 percentile for each of the distributions. This is because in each experiment the frequency of single component failures (about 100,000) is so much larger than the frequency of the larger cascades found in the tail of the cascade size distribution.

[19] This is because the full model, taking into account both the Telco and Power MEs, exhibits a power law trend for its cascade-size distribution, $P(\text{Cascade size} = x) \approx x^{-2.3}$, with larger cascades being more frequent than is predicted by the exponentially distributed cascades in the partial model.

Upon examining Figure 10 in this small range of probabilities we see that a stochastic ordering[20] exists between the respective distributions for $\alpha = 500$ and, say, $\alpha = 1$. There is no ordering, however, between the $\alpha = 150$ and $\alpha = 250$ distributions ("250" is not included in the plot). And, in fact, within the range of α-values from 1 to 150, the observed distributions of cascade size are very similar. So, somewhere in the range of α-values $[150, 500]$ there is a noticeable global[21] change in the model behaviour.

In addition, the different cascade size *probability mass function*s (pmf) resulting from different α values all possess the same general shape, as shown in Figure 11. Depicted are 6 different cascade size pmfs resulting from experiments conducted using 6 different values of α; there is little difference between them. For the Rome study, this similarity is a consequence of the interplay between the following two types of coupling:

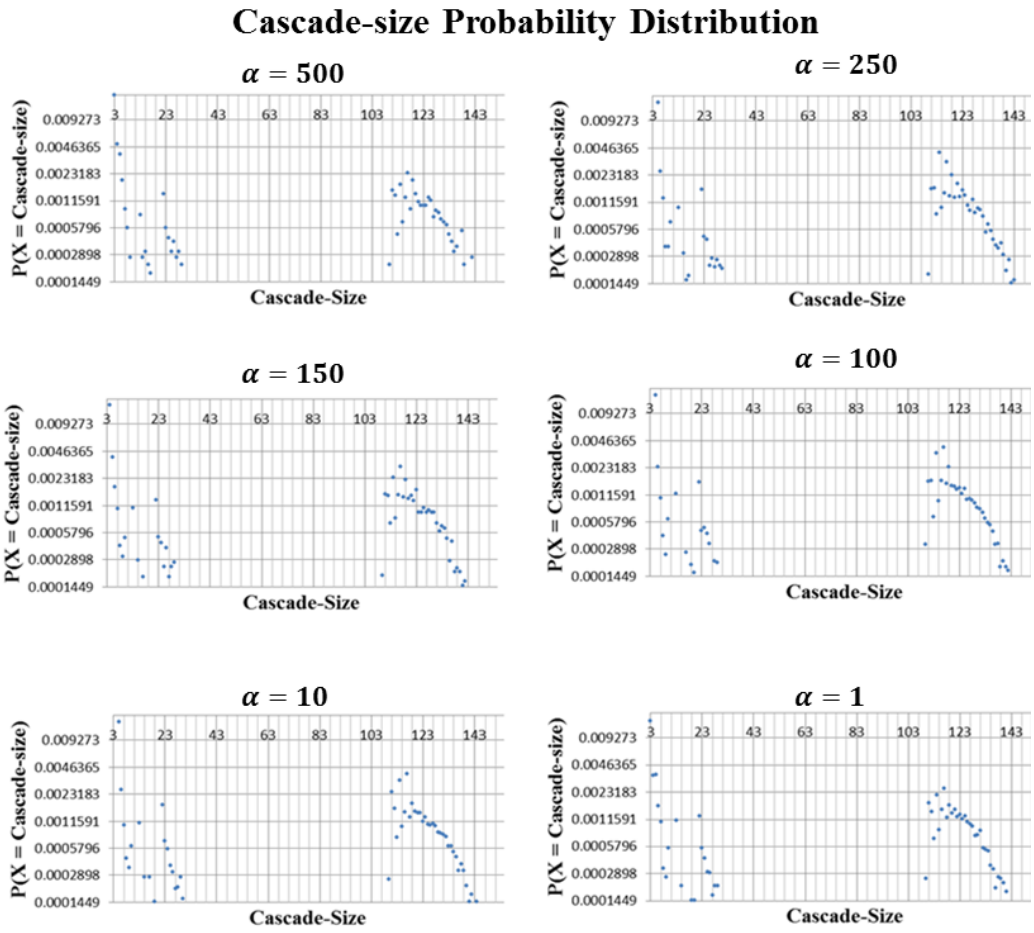

Figure 11: Cascade-size *probability mass function*s for α values 1, 10, 100, 150, 250 and 500.

1. **Functional coupling via network topology**: due to the functional couplings between the modelled networks, *it only takes a few particular MEs to fail for a cascade of a given size to occur*. Telecommunication centres in the model provide examples of functional coupling. Each such centre contains multiple telecommunication MEs (such as Add-drop multiplexers), and all of these MEs rely on the same primary and secondary power sources. That is, via a single power network ME (such as a medium voltage power cable) the power network provides power to a given telecommunication centre, and when the power network fails a backup generator supplies power to the centre for some 50 hours on average[22]. Consequently, whenever both of these power sources are unavailable *all of the MEs at the affected telecommunications centre become inoperable*.

[20] In this range, one distribution lies above another.

[21] That is, over the entire probability distribution.

[22] Consequently, in our model, the failure of a Power network ME that is a coupling point is more likely to be followed by the failure of a backup generator than the failure of another power network ME.

2. **Stochastic association**: The relationship between a stochastic child and its parents is such that the value of α – the strength of the stochastic dependence – only begins to take effect after at least one of the parents has experienced failure. Prior to such an initial failure, if there are no MEs anywhere in the model in a failed state, the experiments have identical stochastic behaviour regardless of α's value.

Consequently, for most of the experiments reported here (apart from the $\alpha = 500$), the relative frequency with which a coupling point that is a power ME fails, followed by the failure of a backup generator and all of the reliant telecommunication MEs at a given centre, is almost unchanged by varying α.

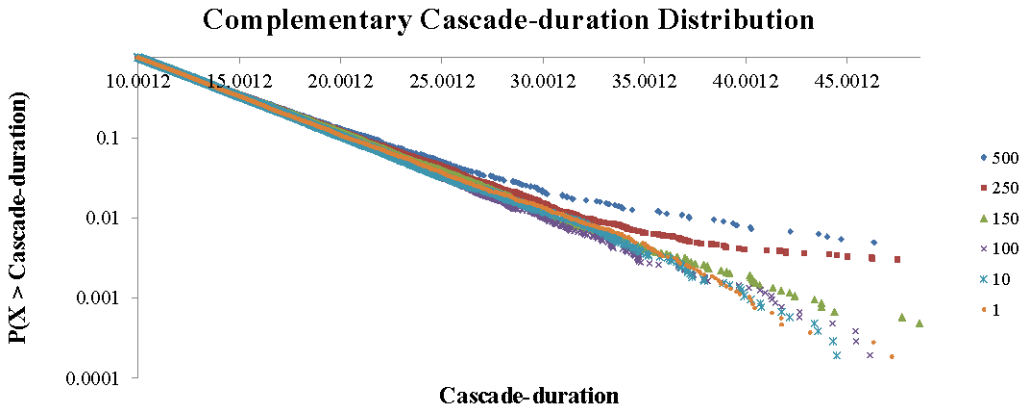

Figure 12: Cascade-duration distributions resulting from α values 1,10, 100, 150, 250 and 500.

Compare this with Figure 12 which depicts *cascade duration* distributions. There is a linear portion of the graph indicating exponential distributions of cascade durations. We see, however, that the tails of some of the distributions deviate from this (the probability of long durations are greater than what would be expected under an exponential probability law), beginning at some value of α greater than 150 and less than 250.

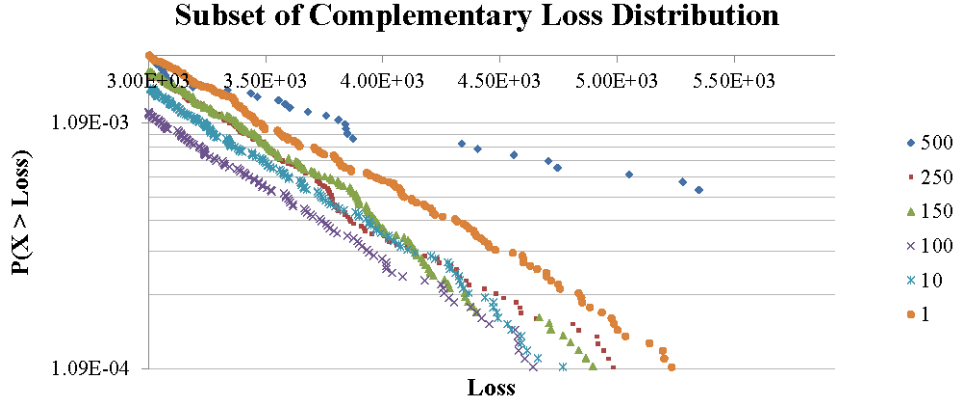

Figure 13: Distributions of "Loss" due to cascades under different values of α, where the "Loss" due to a cascade is defined as *cascade size* $\times$ *cascade duration*.

Similar to the *duration* distributions, nonlinearity also occurs in the tails of "loss" distributions (see Figure 13), however these are now noticeable at some α value greater than 250 and less than 500.

An alternative way of seeing this nonlinearity in the "loss" distribution is by comparing the proportion of loss due to each cascade size in each of the experiments. For a given value of α the proportion of loss due to a cascade of size k is computed as follows. Let $L(i)$ be the loss associated with cascade i. Then, the proportion is

$$\frac{\sum_{i \in \{\substack{cascades\ of \\ size\ k}\}} L(i)}{\sum_{j \in \{all\ cascades\}} L(j)} .$$

These proportions take values in the unit interval $[0,1]$ and are plotted in Figure 14. The horizontal axis depicts the proportions for cascade sizes in the $\alpha = 1$ experiment while the vertical axis depicts the proportions for cascade sizes in the $\alpha = 1,\ 250$ and 500 experiments. That is, each data point is a pair of proportions related to the same cascade size, at least one of which is the proportion from the $\alpha = 1$ experiment. Note that the cascade size related to each point is not explicitly shown in the figure. In the plot, the proportions from the $\alpha = 1$ experiment are represented as triangles on the diagonal line. Data points that do not fall on this diagonal (the unit

slope) line indicate deviations between the $\alpha = 1$ experiment and the other experiments.

For values of α between 1 and 250, most of the points tend to lie on, or close to, the unit-slope line (i.e. the circles and triangles are close to lying on the unit slope), with some occasional outliers that indicate an order-of-magnitude, or more, difference. We, however, see a more pronounced difference between the $\alpha = 1$ and 500 experiments (i.e. the diamonds do not tend to lie close to the unit slope). This observation is similar to the observation made with the "cascade size" measure: yet again, increasing the strength of the stochastic association does not lead to visible changes of model behaviour, *until* some critical threshold is reached. And, the results of this section highlight how changes in α over certain ranges lead to noticeable change for some risk-measures, but not others (compare the cascade-size, cascade-duration and cascade-loss distributions). The range of α values we have investigated suggest that precisely *which* α values trigger noticeable changes *can* vary from measure to measure. A note of caution is in order, therefore: it is prudent that a diverse collection of risk-measures be used when applying PIA (as in any risk-modelling of sufficient complexity), studying model behaviour over a range of plausible parameter values, to fully understand the properties and limitations of the resulting models.

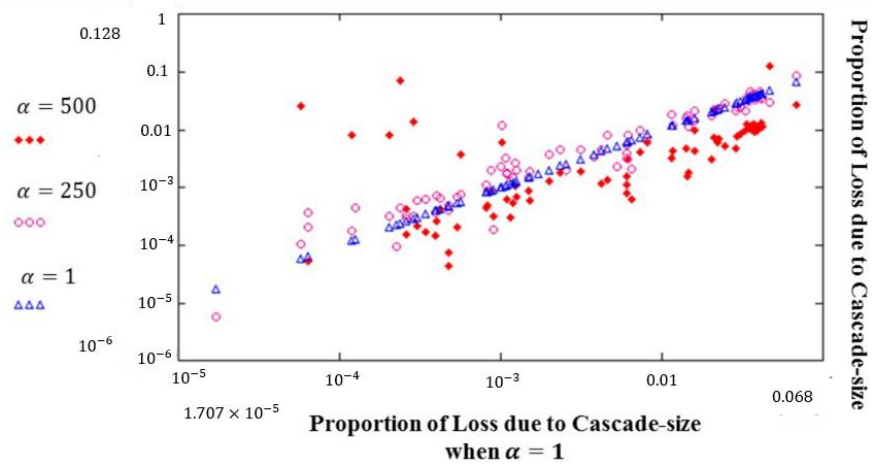

Figure 14: The proportion of loss due to a cascade of a given size can differ with different values of α.

7. Discussion

This paper demonstrates, on a realistic example, an approach to the evaluation of systemic risk and the significance of interdependencies between critical infrastructures.

This paper presents a systematic method of building, via an iterative process of model refinement, models to support the analysis of CI resilience. The method allows one to operate at different levels of abstraction, ranging from purely probabilistic study, at one extreme, to high-fidelity analysis using a number of deterministic models which capture, explicitly, the "physics" of the underlying CIs, at the other extreme. In-between these extremes lies a multitude of hybrid models with both stochastic and deterministic phenomena. Once a hybrid model of interdependent CIs *has* been developed, suitable risk-measures can be defined to study the system properties of interest (such as system resilience). While these measures are dependent on the level of detail used in the model (e.g. the component related risk-measures, such as those in Figure 8, cannot be used unless the model is defined at the component level), the measures, themselves, do not have to be part of the model definition. In fact, many of the risk-measures summarized in this paper were defined *after* the simulations were completed, and computed using the traces of the system state recorded during the simulations. This allowed us to run the simulation campaigns *only once*, and subsequently use the traces to compute different risk-measures of interest whenever the need for a new one occurred. This was important as the length of simulation can be significant, sometimes requiring a number of days to obtain sufficient statistical confidence in the results.

We demonstrate the PIA method in the appendix and note that the method was also successfully applied to build a model of another system of similar complexity, in which the focus was on modelling the resilience of power systems against cyber-attacks [31, 32]. This is an example of modelling, in detail, the effect of an adverse environment on critical infrastructure.

While the PIA method is flexible, allowing the modeller to choose different levels of abstraction, it is outside the scope of this paper to provide advice on how one can establish the appropriate level of abstraction for one's needs. Clearly, this decision is of paramount importance: it is not known what trade-offs are being made by abstraction and whether abstraction approaches that are *guaranteed* to capture essential behaviour *can* be developed. On the

one hand, being tempted to simplify the system model too much, one may "throw away the baby with the bathwater" and, thus, one might not observe serious risks (For example, there is some evidence from modelling of the UK power network that DC approximations underestimate the development of large cascades [48]). On the other hand, trying to increase the level of detail by too much may make the model intractable or, even if it *is* tractable, the insight may not be worth the extra effort. On the positive side, however, the flexibility of the PIA models and the availability of the PIA tool makes it possible for one to experiment with alternatives relatively easily, varying the level of detail when building the system model.

We gained some intriguing insight from our forays into CI-interdependency analysis using PIA.

1) *Further confirmation that the "N-1" criterion used in power system reliability assessment* [44, 45] *is, by itself, too simplistic*, and our model captures (as shown in Figure 7) the complex distribution of the size of a cascade, including some very large cascades which occur with non-negligible probability.
2) *"Full vs Partial" model view*: A strength of the case study is that it illustrates *how* considering single infrastructures might underestimate the risk of larger failures (Figure 9) and demonstrates the capability of the tools and methods we have for developing multi-infrastructure models for systemic risk analysis. It also illustrates how reliable the infrastructures are and how relatively rare these widespread failures are. We have also demonstrated a significant difference between estimates of systemic risk, where the estimates are based on two models with differing levels of detail – one "full view" model with explicitly modelled inter-CI dependence and another "partial view" model without it. The results as depicted in Figure 9 show a clear difference between the respective model's distributions of cascade-sizes. In particular, the "full-view" model's cascade-size distribution exhibits a power-law, indicating that larger cascades are significantly more likely than if the distributional law was, say, Poisson.
3) *Ranking of component "importance" to determine sensible investment allocation for improving the weakest components in the CIs*: Ranking is dependent on the ranking criteria. As part of our investigations, we considered a number of alternative definitions for a network component's "importance". Understandably, the usefulness of such notions for decision making heavily depends on what goal a practitioner might have in mind by their use. For instance, given the occurrence of a critical event on the network – such as the shorting of a high-voltage power line – a network operator might seek to mitigate the immediate ongoing impact on the network or, perhaps, choose to expedite the recovery of the network from such a disruption. Clearly, these goals are related and both of them seem important for network resilience. However, our initial findings suggest that ordering of components according to notions of "criticality" aligned with these goals can be very different. In section 6.1, using three related notions of "criticality" – cascade membership, component connectivity and component (un)availability – produces significantly different component rankings: some components that are very critical because, historically, they have had a high propensity of being part of large cascades, are less critical in that they have significantly higher availability than other network components. A component relatively unlikely to fail may be one to take notice of, if a network disruption is of the kind that typically results in cascades.
4) *Sensitivity Analysis*: Although the primary "accelerant" for cascades in our model is the physics of the network – that is, redistribution of power flow – in section 6.3 we detail the results of our investigation into how our particular choice of stochastic coupling (between network components) affects different properties of the network. In some respects the effects of increased coupling were negligible, even for large values of the coupling: for instance, increasing the strength by two orders of magnitude had little effect on the distribution of cascade-sizes (see Figure 11). This is primarily due to whether a given coupling strength makes the failure of children sufficiently likely before the recovery of a failed parent: very large values of the coupling are required for this to be the case since, typically, failure rates are significantly smaller than repair rates. In other respects, however, the coupling strength induced nonlinear relationships across the distributions of both the loss incurred and duration of cascades (see Figure 12 and Figure 13). So, while there is some inertia in terms of an increase in the occurrence of large cascades, when they do occur many children remain in a failed state for longer because they are strongly coupled with parents who are in a failed state, thereby increasing the impact of the cascade. While an increased dependence between the components may not make large cascades more likely, it may have the effect that both the duration and impact of such large cascades, when they do occur, may tend to be significantly increased.

The modelling effort, itself, also provides interesting insights. A contribution of this work is the following: given a collection of desirable (i.e. justified by real world data) properties for both the marginal and joint probability distributions that model the behaviour of the MEs, our methodology can be used to construct consistent joint probability distributions of the components. That is, we specify a mechanism by which the MEs are made probabilistically dependent on each other: a model feature achieved by defining so-called stochastic associations between the components. The resulting joint distributions are consistent with the, aforementioned, desirable properties a modeller has in mind, but these distributions may not be unique or the most suitable. Indeed, our

methodology allows one to successively refine the distributions resulting in models that better approximate phenomena observed in the actual system; the only challenges being what the definition of the dependence relationships should be and what the parameterisation of the resulting model should be. In this way our approach provides a rigorous, consistent way in which domain experts can build models that take their assumptions and expectations into account. The point here being that system complexity makes the calculus of uncertainty tricky even for a domain expert; so a rigorous model that is clear enough to be critiqued will be useful in propagating expert judgement to observable, understandable consequences. This is in a similar spirit to [49] where details of a process of expert elicitation and model refinement are given.

We plan to extend this work. Of immediate concern are the following areas:

- We would like to extend the approach and tool support to be able to account for the evolution of model parameters or even the *structure of the modelled systems*. This seems important in practice, e.g. to be able to adequately model the evolution of systems over long periods of time during which the modelled infrastructures will be subjected to changes, and to take into account the impact of "covert channels" in ICT systems [50];
- Cyber-security of *Industrial Control Systems* (ICS) is also a priority. In this paper we did not specifically address cyber security. Although the approach is general enough and allows for cyber security threats to be modelled at high level of abstraction using the proposed approach addressing cyber security threats in greater detail would require extending the proposed modelling approach. Various ideas have already been tried, including some of ours [32]. Other noteworthy examples are applications of the ADVISE formalism [51] and the approach to assessing SCADA system cyber-security in [52].

Alternative sources of data for model validation continue to be sought. For instance, work with available data on interdependency related incidents [30] offers ball park figures for the values of stochastic associations.

Validation of the methodology itself, in practice, requires the PIA approach to be embedded in an iterative engineering process that challenges and verifies the approach. In other work [33], we have used the *Claims, Argument, Evidence* (CAE) framework to explore the justification for the models and data used. We would expect such validation to provide assurance in practice.

It would be interesting to apply PIA to other types of infrastructure; in particular, we are working with domain experts to extend the Rome model. We also intend to further our investigation of the Rome scenario as follows:

- In the studies presented in this paper, we assumed unlimited maintenance resource. In practice, however, this is unrealistic. Evidence-based planning of the level of maintenance is clearly important. Looking at this aspect, especially in combination with economic loss models, will offer useful practical insight. For the current model, this problem was studied (only informally) by analysing the simulation trace.
- Finally, investigating models with more sophisticated dependence mechanisms may offer useful insight. For instance, the rate of failure of a power line is, arguably, dependent on the power flow through it; a dependence that we have not seen modelled in the spirit of the work presented here.

8. Conclusions

We have presented, in detail, a risk-based approach to interdependency analysis of critical infrastructures. The method proposed, PIA, starts with a qualitative phase and, via a set of focussed refinements, may be evolved into a quantitative method for assessing the risk due to interdependencies between CI. The method can be used at different stages of a CI's lifecycle: the planning of investment across the CI, operational and resilience planning, and the evaluation of various risk mitigation techniques.

Due to its nature of iterative refinement, PIA, by design, can be applied at different levels of abstraction: from very abstract modelling – in the hope that the modeller will quickly capture 'important' interdependencies – to high fidelity simulations integrating many different aspects (both stochastic and deterministic) of the behaviour of the Cis.

An essential aspect of quantitative PIA is how it allows one to assess risks due to rare events (via very extensive simulation campaigns). These can be either:

- millions of repetitions of short periods, e.g. hours, which are important for operational and emergency planning, or
- thousands/millions of repetitions of very long periods, e.g. tens and hundreds of years of operation, which are important for short and long term planning. An example where such simulations might be of interest are studies of the impact of climate change on the resilience of infrastructures.

Acknowledgements

This work has been carried out with the financial support of a number of research and innovation projects: the EU FP6 IRRIIS (IP) IST Project N° 027568, FP 7 AFTER Grant agreement N° 261788 and the Innovate UK (formerly the Technology Strategy Board of the UK) PIA:FARA, Grant number TSB BK030F.

References

1. Hollnagel, E., D.D. Woods, and N. Leveson, *Resilience Engineering: Concepts and Precepts*. 2007: Ashgate Publishing, Limited.
2. Standardisation, I.O.f., *Risk Management - Principles and guidelines.* 2009a.: Geneva.
3. Standardisation, I.O.f., *Risk management – Risk assessment techniques*. 2009b.: Geneva.
4. Ericson, C.A.I., *Hazard Analysis Techniques for System Safety*. 2005: John Wiley and Sons.
5. Neil, M. and N. Fenton, *Risk Assessment and Decision Analysis with Bayesian Networks*. 2012: CRC Press. 524.
6. Anonymous, *Interdependencies and Reliability in the Combined ICT and Power System: An overview of Current Research (status: under review).* Applied Computing and Informatics, 2016.
7. Adelard. *Assurance and Safety Case Environment (ASCE) tool*. 2012; Available from: http://www.adelard.com/asce/.
8. The Performability Engineering Research Group at The University of Illinois at Urbana-Champion. *The Mobius Modelling Tool*. 2012; Available from: https://www.mobius.illinois.edu/.
9. Rinaldi, S.M., J.P. Peerenboom, and T.K. Kelly, *Identifying, Understanding, and Analyzing Critical Infrastructure Interdependencies.* IEEE Control Systems Magazine, 2001. **21**(6): p. 11-25.
10. Bloomfield, R., N. Chozos, and P. Nobles, *Infrastructure interdependency analysis: Introductory research review*. 2009, Adelard LLP.
11. Bloomfield, R.E., N. Chozos, and K. Salako, *Current Capabilities, Requirements and a proposed Strategy for Interdependency Analysis in the UK*, in *CRITIS 09* 2009, Springer.
12. Ouyang, M., *Review on Modeling and Simulation of Interdependentcritical Infrastructure Systems.* Reliability Engineering and System Safety, 2014(121): p. 43-60.
13. Marchi, S.d. and S.E. Page, *Agent-Based Models.* Annual Review of Political Science, 2014(17): p. 1-20.
14. Vicsek, T., *Complexity: The Bigger Picture.* Nature, 2000. **418**: p. 131.
15. Miller, J. and S.E. Page, *The Standing Ovation Problem.* Complexity, 2004. **9**(5): p. 8-16.
16. Forrester, J.W., *Industrial Dynamics*. 1961, Cambridge, MA: MIT Press.
17. Kollikkathara, N., H. Feng, and D. Yu, *A System Dynamic Modeling Approach for Evaluating Municipal Solid Waste Generation, Landfill Capacity and Related Costr*

Management Issues. Waste Management, 2010. **30**(11): p. 2194–2203.

18. Sterman, J.D., *Business Dynamics: Systems thinking amd Modelling for A Complex World* 2000, New York, NY: Mc-Graw Hill.
19. Kwiatkowska, M., G. Norman, and D. Parker. *Advances and Challenges of Probabilistic Model Checking. .* in *48th Annual Allerton Conference on Communication, Control and Computing*. 2010. IEEE Press.
20. Kwiatkowska, M., G. Norman, and D. Parker, *PRISM: Probabilistic Model Checking for Performance and Reliability Analysis*, in *ACM SIGMETRICS Performance Evaluation Review*. 2009, ACM. p. 40-45.
21. Kwiatkowska, M., G. Norman, and D. Parker. *Stochastic Model Checking*. in *Formal Methods for the Design of Computer, Communication and Software Systems: Performance Evaluation (SFM'07)*. 2007. Springer.
22. Pintilie, M., *Competing Risks: A Practical Perspective*. 2006, Hoboken, NJ: John Wiley and Sons.
23. David, H.A. and M.L. Moeschberger, *The Theory of Competing Risks*. 1978, London: Charles Griffin and Co.
24. Newman, M.E.J., *Complex Systems: A Survey.* American Journal of Physics, 2011. **79**: p. 800-810.
25. Newman, M.E.J., *The Structure and Function of Complex Networks.* Society for Industrial and Applied Mathematics, 2003. **45**: p. 167-256.
26. Johansson, J. and H. Hasselb, *An approach for modelling interdependent infrastructures in the context of vulnerability analysis.* Reliability Engineering and System Safety, 2010. **95**(12): p. 1335-1344.
27. Svendsen, N.K. and S.D. Wolthusen, *Connectivity models of interdependency in mixed-type critical infrastructure networks.* Information Security Technical Report, 2007. **12**(1): p. 44-55.
28. Bobbio, A., et al., *Unavailability of critical SCADA communication links interconnecting a power grid and a Telco network* Reliability Engineering and System Safety, 2010. **95**(12): p. 1345-1357.
29. Utne, I.B., P. Hokstad, and J. Vatn, *A method for risk modeling of interdependencies in critical infrastructures.* Reliability Engineering and System Safety, 2011. **96**(6): p. 671-678.
30. Wright, D., et al., *Industrial Sector-Based Modelling of 1337 Critical Infrastructure Incidents in the European Union*. 2010, Centre for Software Reliability. p. 58.
31. Netkachov, O., P. Popov, and K. Salako, *Quantification of the Impact of Cyber Attack in Critical Infrastructures.* Lecture Notes in Computer Science, 2014. **8696 LNCS**: p. 316-327.
32. Netkachov, O., P.T. Popov, and K. Salako. *Model-based Evaluation of the Resilience of Critical Infrastructures under Cyber Attacks*. in *9th International Conference on Critical Information Infrastructure Security (CRITIS 2014)*. 2014. Limassol, Cyprus.
33. Netkachova, K., et al., *Using Structured Assurance Case Approach to Analyse Security and Reliability of Critical Infrastructures*, in *Computer Safety, Reliability, and Security: SAFECOMP 2015 Workshops, ASSURE, DECSoS. ISSE, ReSA4CI, and*

SASSUR, Delft, The Netherlands, September 22, 2015, Proceedings, F. Koornneef and C. van Gulijk, Editors. 2015, Springer International Publishing: Cham. p. 345-354.

34. Doob, J.L., *Stochastic processes,* . 1990: Wiley.
35. Ross, S.M., *Stochastic processes,* . 1996: Wiley.
36. Sanders, W.H. and J.F. Meyer. *Stochastic Activity Networks: Formal Definitions and Concepts*. in *Lectures on Formal Methods and Performance Analysis*. 2001. Berlin: Springer.
37. David, H.A. and M.L. Moeschberger, *The Theory of Competing Risks*. Vol. 39. 1978. 103.
38. Bloomfield, R., et al. *Stochastic Modelling of the Effects of Interdependencies between Critical Infrastructure*. in *(CRITIS 09) International Workshop on Critical Information Infrastructures Security*. 2010.
39. Hammersley, J.M. and D.C. Handscomb, *Monte Carlo Methods* 1975, London: Methuen.
40. Klein, R., *The EU FP6 Integrated Project on Dependent Critical Infrastructures: Summary and Conclusions.* Critical Information Infrastructure Security: 5th International Workshop, CRITIS 2010, Athens, Greece, September 2010, Revised Papers, 2011.
41. City University London. *Probabilistic Interdependency Analysis: framework, data analysis and on-line risk assessment*. 2009; Available from: http://www.city.ac.uk/centre-for-software-reliability/research/research-projects/piafara-probabilistic-interdependency-analysis-framework,-data-analysis-and-on-line-risk-assessment.
42. Lindley, V.D. and D.N. Singpurwalla, *On exchangeable, causal and cascading failures.* Statistical Science, 2002. **17**(2): p. 209-219.
43. Carreras, B.A., et al., *Evidence for Self-Organized Criticality in a Time-series of Electric Power System Blackouts.* Transactions on Circuits and Systems, 2004. **9**(51): p. 1733 – 1740.
44. North American Electric Reliability Corporation (NERC), *Reliability Concepts*. 2007.
45. North American Electric Reliability Corporation (NERC), *Reliability Assessment Guidebook*. 2012.
46. Nedic, P.N., et al., *Criticality in a cascading failure blackout model.* Electrical Power and Energy Systems, 2006. **28**: p. 627-633.
47. Lu, W., et al., *Blackouts: Description, Analysis and Classification.* Proceedings of the 6th WSEAS International Conference on Power Systems, 2006.
48. Simonsen, I., et al., *Transient Dynamics Increasing Network Vulnerability to Cascading Failures.* Phys. Rev. Lett, 2008. **100**(21): p. 4.
49. Jensen, F.V., *An introduction to Bayesian networks*. 1996: UCL Press (Published in North America by SpringerVerlag NewYork Inc). 208.
50. Hearing Before The Subcommittee On National Security, Homeland Defense And Foreign Operations Of The Committee On Oversight And Government Reform, and H.o.R.t.C.F. Session, *Cybersecurity: Assessing The Immediate Threat To The United States*. 2011.

51. Ford, M.D., et al., *Implementing the ADVISE security modeling formalism in Möbius*, in *The 43rd Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*. 2013, IEEE: Budapest, Hungary.

52. Ten, C.-W., C.-C. Liu, and G. Manimaran, *Vulnerability Assessment of Cybersecurity for SCADA Systems.* IEEE Transactions on Power Systems, 2008. **4**(23): p. 1836-1846.

53. Klein, R. *Information Modelling and Simulation in Large Dependent Critical Infrastructures – An Overview on the European Integrated Project IRRIIS*. in *3rd International Workshop, CRITIS 2008*. 2008. Rome, Italy: Springer.

Appendix A

Rome case-study – an example of applying the PIA method

One of the case-studies used to validate the PIA method was based on a real incident, which occurred in Rome, Italy. The incident initially affected the telecommunications in the Rome metropolitan area, and subsequently the power system. In this appendix the case-study is referred to as the *Rome scenario*. Further information about the incident can be found in [28, 40, 53].

We applied the method described in section 3 of the main text and built a *Rome scenario* model which, through a number of refinements, resulted in a probabilistic model (a complex *stochastic activity network* (SAN)) used to conduct the studies described in section 6 of the paper.

The process of building the model is described in detail below using the tool support briefly described in section 1 of the paper. The illustrations presented below include screen shots of the user interface of the tool, PIA:FARA. Recall, from section 3, the two levels of abstractions, *HLSM* and *DSBM*, used to model the entire system under study – one in which every service (CI) is represented as a "Black Box", and the other by the network of components the service relies upon, respectively. These abstraction levels are referred to in the tool as "INTER" and "INTRA" models, respectively. Clearly, each of these models can be applied to capture connectivity between, either, physical assets or their stochastic properties.

Each PIA model is represented as a PIA:FARA project, a combination of diagrams and text. Figure 15 gives an overview of the Rome Scenario project, which consists of a number of models, which fall broadly into four categories:

- Physical topology/network. The models of this category are either "INTRA" models, (defined in the tool as type INTRA_PN), and represent the detailed topology of an individual CI (or service), or "INTER" model (defined in the tool as type INTER_PN), i.e. represent explicitly the connections between CIs/services modelled in detail (with a respective "INTRA" model). In Figure 15 we have two INTRA PN models, POWER and TELCO, and an instance of INTER_PN, InterCIPN model, respectively.
- Stochastic association topology. These models capture the concept of stochastic association described in section 4 of the paper. Similarly to the models of physical topology, the models of this kind are "INTRA" or "INTER" models and are defined in the tool as types INTRA_SA and INTER_SA, respectively. The project shown in Figure 15 contains two INTRA_SA models, PowerSA and TelcoSA, and an instance of type INTER_SA, InterCISA, respectively.
- State machines. They model the behaviour of the individual modelling elements (MEs) included in the model as is detailed in Stage 3 of section 3.2 of the paper.
- Simulation plugins. This model, or rather a list of definitions, describes third parties software components needed for a particular simulation study. The components must be compliant with an interface defined by us so that they can be "plugged in" to the simulator built by the tool for an existing model. The existing model does not need any alteration. The plugins are typically used to allow the modeller to study effects beyond the pure stochastic behaviour of the MEs. For power systems, for instance, a useful plugin implements power flow calculations which establish phenomena such as line overloading, occurrence of islands in the power networks, etc. For telecommunication networks, a useful plugin will establish the connectedness of the network, etc.

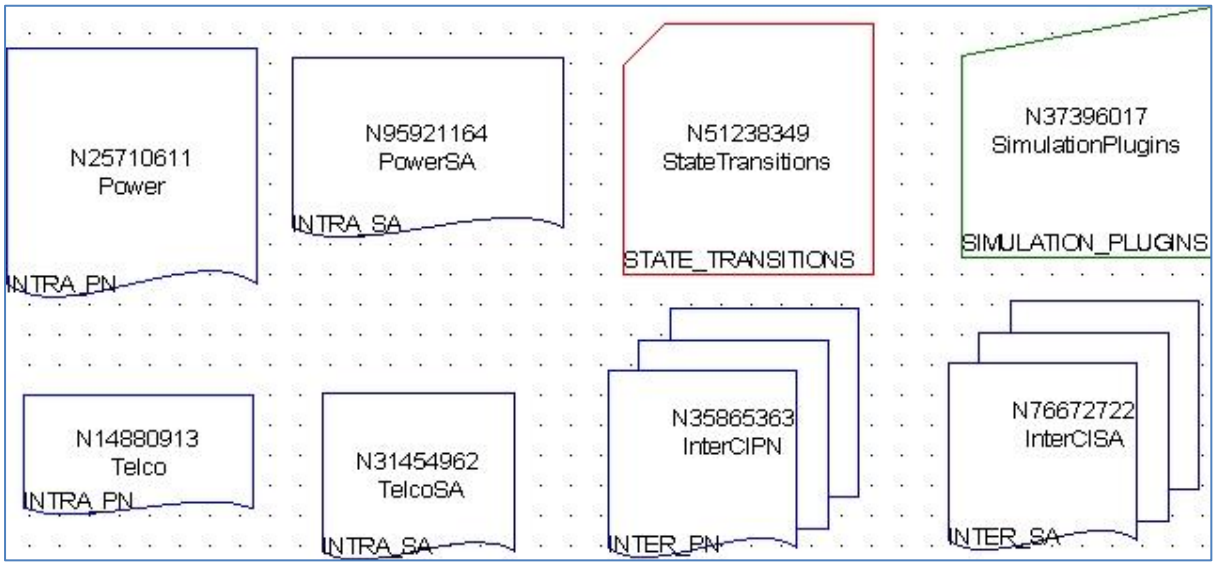

Figure 15. An overview of the project representing the Rome scenario.

Stage 1. We used the descriptions of the actual incident [53] and carried out an in-depth analysis of the system to identify *other likely failure modes* that could have emerged in the case study. We drew the system boundaries and defined the system to consist of two critical infrastructures: Power (Power CI) and Telco (Telco CI). Power CI consists of 2 CI services: Transmission and Distribution, while the Telco CI contains 3 services: Public Switched Telephone Network (PSTN), Global System for Mobile Communications (GSM) network and Synchronous Digital Hierarchy (SDH) network.

Stage 2. HLSM development. As indicated above the two topologies of the CIs were modelled in detail using the respective INTRA_PN models. These models were continuously refined, increasing the level of detail with each refinement.

An alternative approach to representing the system structure would have been to use 5 INTRA_PN models to capture the services listed. We instead, chose to define, for the MEs, an attribute which will indicate which of the services within a CI the particular ME belongs to. Either way for each ME we could tell which CI and which service within the respective CI it is a member of.

Node status fields

Reference	N6478096
Id	30
Title	P13hvc
Node type	HVC
State	True
Latitude	41.8231
Longitude	12.48
Functionality	Power
MemberOf	Power-Network
SubCI	Transmission
IsInterCICouplingPoint	True
IsSubCICouplingPoint	True

Figure 16. An illustration of the attributes attached to the MEs in the system model.

Figure 16 illustrates the parameters used with an ME. The specific set of parameters was defined for the Rome scenario and *may be altered* for other studies. For instance, we used MEs attributes defining their geographical location (Latitude and Longitude), which was useful to visualise the system in Google Map. In other cases, such information may not be available – the definition of the nodes then can be altered not to include the respective attributes. For instance, Latitude/Longitude was not used with MEs referring to lines (power or telco).

Several attributes in Figure 16 illustrate how the MEs are defined to belong to a CI (the attribute 'Functionality') and to service within the CI (the attribute SubCI - Transmission).

The number of MEs which belong to each of the 5 services listed above is summarised in the table below:

Table 1 The number of Modelling Elements (MEs) belonging to different services

Service	Number of MEs
Power Transmission	74
Power Distribution	100
Telco MEs in power services (SCADA, etc.)	27
Telco PTSN	83
Telco GSM	62
Telco SDH	105
Others in Telco ("Offices", i.e. buildings where the telco components reside and lines – power and communication - within the "Offices").	378

A complex topology of physical links between the MEs was added (i.e. the INTRA_PN and INTER_PN models were defined). The MEs which belong to different services (and CI) but were connected by physical links were marked as *coupling points*. For instance, the power elements which provided power to Telco elements were marked as coupling points. Typically, the power will be supplied to an ME of type 'Offices', i.e. the premises where the Telco MEs such as routers, masts, etc. would be held. These offices are also considered as coupling points. The last two attributes in Figure 16 are of type Boolean and allow for coupling points to be defined.

The INTER_PN model included in the Rome scenario model is shown in Figure 17.

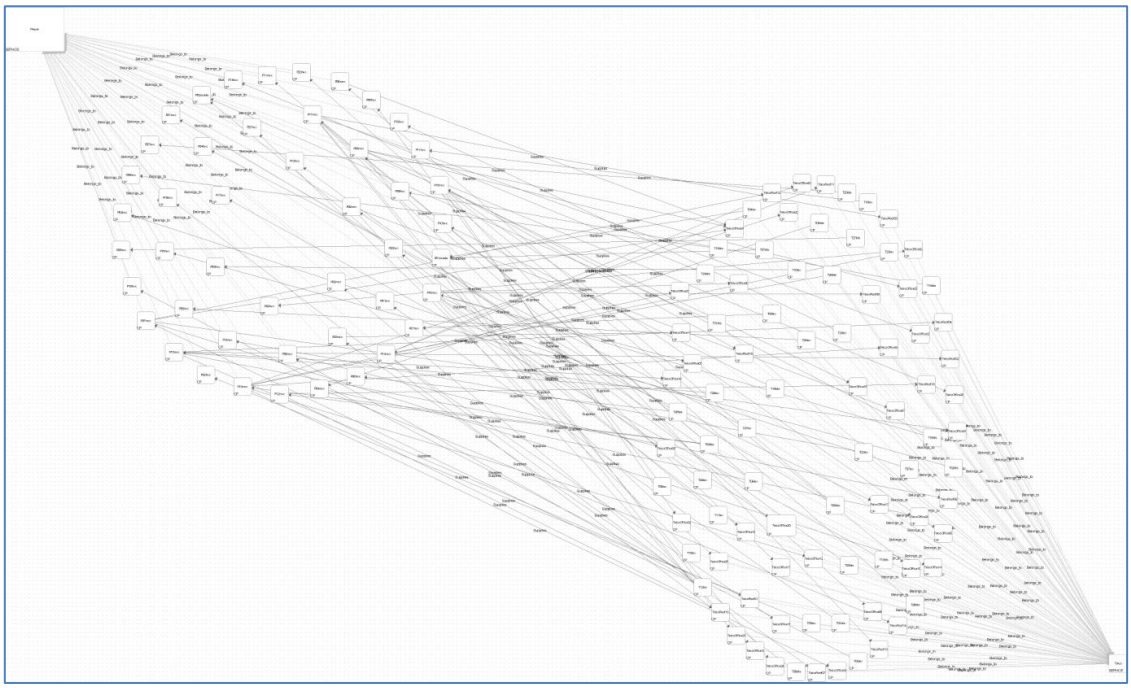

Figure 17. Illustration of the INTER_PN model. The model shows the coupling points between the two CIs: Power and Telco. Coupling points are shown as squares at the two opposite corners of the diagram: Power nodes are located in the top left corner, the Telco nodes in the bottom right corner.

Stage 3. DSBM development. Here, the model of each service defined in Stage 2 has been refined to include a network of components used by the respective service. Each component is modelled as an ME. The behaviour of each ME is modelled by a state machine associated with it. An example of a state machine is given in Figure 18. The nodes (graph vertices) represent the states an ME can be in; the links/edges represent the state change (state transitions) from one state to another. In this example, shown in Figure 18, the state space of the particular ME

consists of 2 states (the possible states are “Failed” and “OK”). Eventually, state machines are refined to include a number of attributes, which are also shown in Figure 18:

- The type function that calculates the probability of that state transition, the type/family of the particular function, and the parameter values needed for the calculation of the function. The “OK-to-Failed” state transition (pointed from by the red arrow) has attributes attached to it. These attributes are specified in *Stage 5* in the process when the probabilistic model is parameterised.

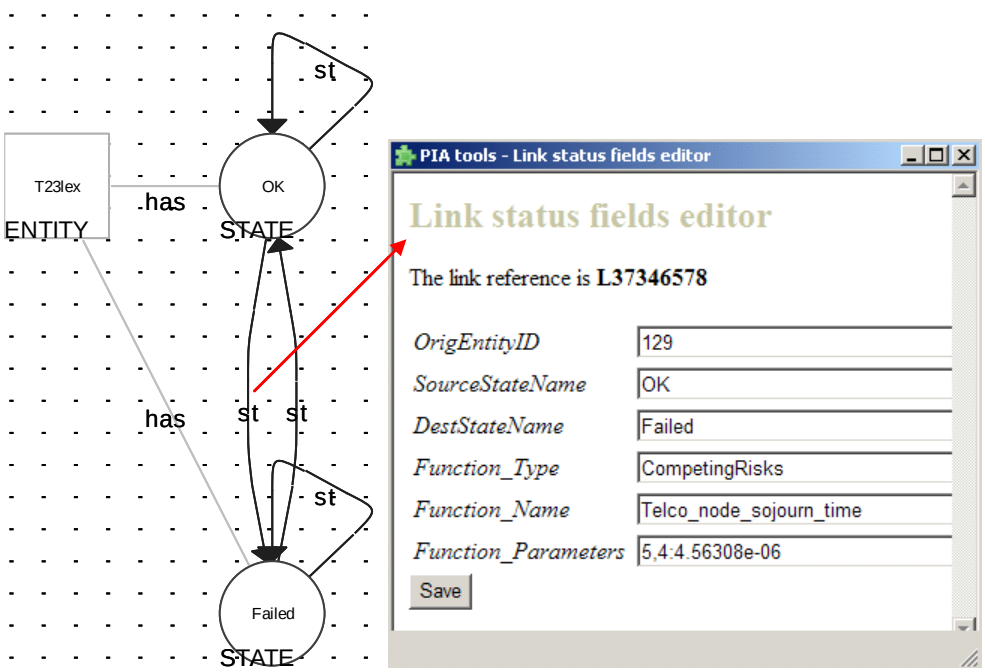

Figure 18: A state machine diagram for an ME (a local telephone exchange) with an illustration of the parameters attached to a state transition.

Stage 4. Identification of dependencies between services.

This stage, too, went through many refinements. It started with a fairly abstract model of dependencies initially between the services. For instance, various telco elements require power to function. This fairly abstract statement of functional dependence is translated to a detailed view of how the individual telco MEs depend on individual MEs powering them. In the Rome scenario, in practice, it is rarely the case that the individual telco MEs receive power directly from a power ME. Instead, typically a power line provides power to a building, which, in turn, provides power to the individual telco MEs. This line of reasoning was followed in the Rome scenario model. An ME from the Power CI is connected to an “Office” ME (the building where telco equipment resides). The entire Office ME is a part of the Telco CI. The MEs modelling a power line providing power to the office and the Office ME are treated as coupling points between the Power and Telco CIs. Typically, each “Office” ME will have a backup power source – a battery or a generator. The Telco MEs residing in the “Office” ME then will be powered if at least one of the sources of Power (power line or a battery/generator) is connected to the respective Office ME is up. In case the Office is left without power, all Telco MEs will be affected. This non-trivial model of *functional dependence* between the elements of the Power and Telco CIs is captured by the topology of the physical network (INTRA_PN and INTER_PN): power lines and backup power supply are connected to the Office ME, individual lines provide power from the Office ME to the individual telco MEs residing in it. If the modeller would like to model in detail the functional dependence of Telco elements on power this can be done by a plugin (see *Stage 6* below).

A number of stochastic dependencies are captured too, using the INTRA_SA and the INTER_SA models (Figure 19 shows the topology of the INTER_SA model, InterCISA). The graph defined in the model captures the “parent – child” relationship for MEs which belong to different CIs. INTRA_SA captures similar relationship between MEs which belong to the same CI. In case an ME has parents, the transition rates used by the state machines representing the particular child ME are modified whenever the state of the some of the parents changes. The strength of the stochastic association is captured in the definition of the transition rates, as illustrated in Figure 18 above.

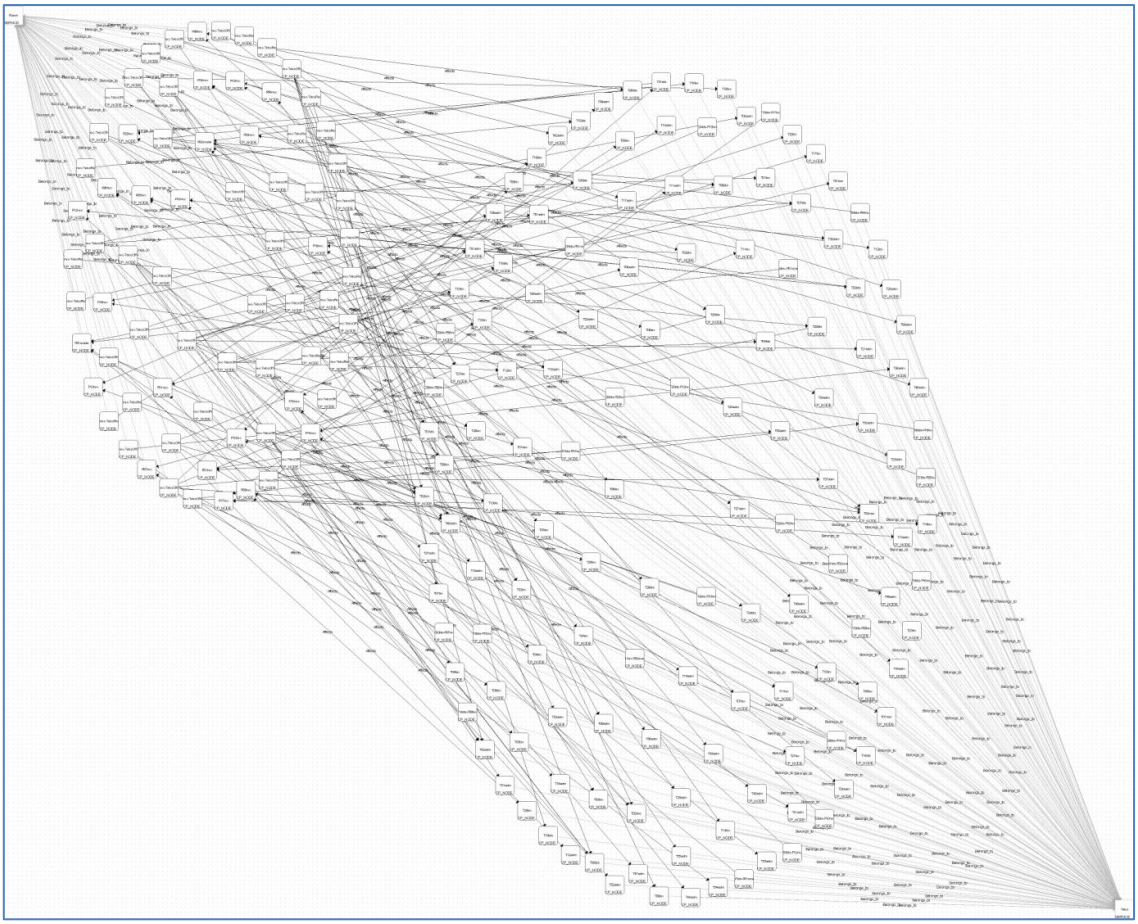

Figure 19. A model of stochastic dependencies, InterCISA, between MEs which belong to different CIs. The MEs of the two CIs are shown as squares in the top left corner (Power CI) and the bottom right corner (Telco CI).

Stage 5. Probabilistic-model development and parameterisation.

In sections 4 and 6.3 of the paper, we have already described how the values of the probabilistic parameters were elicited. The component failure-rates, under normal conditions, were provided by domain experts. For the strength of stochastic dependence, we performed a sensitivity analysis for a plausible range of values, as explained in section 6.3. Here, we illustrate how, using the tool, the elicited values can be assigned to the respective parameters.

Each of the state machines, which model the behaviour of the MEs included in the system model, is parameterised. Returning to Figure 18 we notice that for state transition "OK-to-Failed" the attributes have the following values assigned to them:

- Type of the function used to decide which transition will be next: *CompetingRisks* (see section 4). From the current state OK (the vertex at the origin of the transition edge) a transition can occur to either the other states "Failed" or the state may remain unchanged.
- The function name is: *Telco_node_sojourn_time*. This parameter adds convenience and flexibility for the Monte Carlo simulator.
- The parameter values are: *5,4:4.56308e-06*. The last number describes the transition rate associated with a transition under *normal conditions*, i.e. when all of the parents of the particular ME are in an "OK" state. The other two parameters represent in an encoded form the parameters of how the parents being in a non-OK state impact the transition rate of the child ME. The interested reader is referred to [PIA:FARA Deliverable 2] for a more detailed description of the syntax of the particular values.

Stage 6. Deterministic Models Configuration.

At this stage the modeller may decide to include in the simulation a number of plugins, which allow for more detailed modelling of various aspects of the system. Figure 20 shows a list of plugins added to the model of the Rome scenario:

- libInitialize_Nodes_Legacy.dll,

- libPlugin_PowerNetwork_LoadFlow.dll,
- libPlugin_TelcoBateryDependence.dll, and
- libTracePlugin_BinarySnapshots.dll.

These are created as dynamic link libraries (DLLs) and serve different purposes, indicated by their modelling type: INITIALISATION, DETERMINISTIC and TRACE. The first plugin is used to initialise the simulator and allows for fine grain tuning of the initialisation depending on the particular purpose of the study. The DETEMINISTIC plugins are examples of extending the simulator functionality beyond the pure stochastic behaviour. The power flow implements an algorithm of a DC power flow approximation; BatteryDependence plugin checks whether telco MEs are powered and if not, changes their state to inoperable.

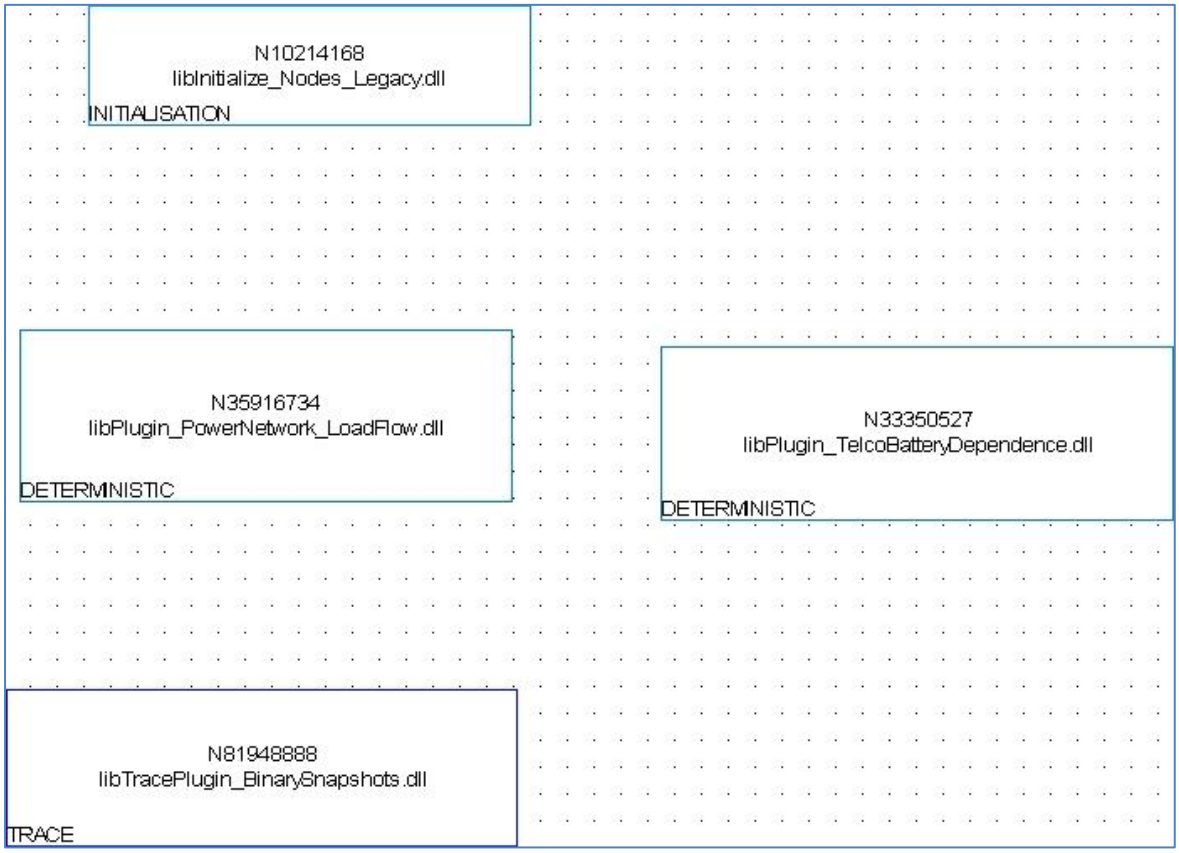

Figure 20. Plugins “model” with PIA lists the plugins of different types that are added to a particular PIA model.

Stage 7. Exploratory Interdependency Analysis.

This stage involves Monte Carlo simulation with the defined model. A trace, as defined in the TRACE plugin, is created, typically recording the events of interest, which are subsequently analysed off-line by suitable tools. Section 6 of the paper provides examples of results obtained with the Rome scenario model.