City, University of London Institutional Repository

Stochastic Modelling of the Effects of Interdependencies between Critical Infrastructure

Robin Bloomfield[1,2], Lubos Buzna[3], Peter Popov[1], Kizito Salako[1], David Wright[1]

[1]Centre for Software Reliability, City University London. 10, Northampton Square, College Building, EC1V 0HB, London, UK.
{reb, ptp, kizito, dw}@csr.city.ac.uk

[2]Adelard LLP. 10, Northampton Square, College Building, EC1V 0HB, London, UK.
reb@adelard.com

[3]University of Zilina, Univerzitna 8215/1, 010 26 Zilina, Slovakia.
lbuzna@ethz.ch

Abstract. An approach to Quantitative Interdependency Analysis, in the context of Large Complex Critical Infrastructures, is presented in this paper. A Discrete state–space, Continuous–time, Stochastic Process models the operation of critical infrastructure, taking interdependencies into account. Of primary interest are the implications of both model detail (that is, level of model abstraction) and model parameterisation for the study of dependencies. Both of these factors are observed to affect the distribution of cascade–sizes within and across infrastructure.

1 Introduction

Dependencies within and between *Critical Infrastructures* (CI) have been recognised as important for achieving (or undermining) acceptable system safety, security and dependability [1, 2]. There is a growing body of research into the quantitative modelling of Complex systems, their dependencies and the implications, thereof, for the occurrence and sizes of cascades [1, 3-6]. The need to understand (inter)dependencies is evidenced by the occurrence of spectacular, catastrophic *cascades[1]* as a direct result of dependencies. One such example is the North American Blackout that occurred on the 16th of August, 2003 which affected an estimated 10 million people [7]. Yet another example is the explosion that occurred on 11 December, 2005 at Buncefield Oil

---

[1] A cascade may be defined as a causally related sequence of undesirable events. However, later in this paper, we will use a definition of cascade that does not require the events to be causally related.

Storage Depot, Hertfordshire, in the United Kingdom. The explosion affected part of the local *Information Infrastructure* and, ultimately, led to patient records for hospitals in the wider area being affected [8]. There are 2 points to note from these events. Firstly, the extent of the damage caused in each of the incidents was difficult to predict at the time. Certainly, had the cascades' occurrence and evolution been better predicted (or detected earlier), preventative and mitigation measures might have limited the consequences of the cascade. Such uncertainty is characteristic of many cascade events in CIs and suggests that CI dependencies, and there implications, are not yet well understood. Secondly, in both examples there were dependencies present that exacerbated the cascades. Indeed, investigations undertaken after the cascades occurred exposed the role of a number of dependencies in facilitating the cascades. Via these dependencies (e.g. the geographic proximity of IT database systems to the fuel depot in the Buncefield incidence) the state of some CI component (e.g. explosion at Oil depot) was related to the state of some other component (e.g. database storing healthcare records), possibly in another CI. Therefore, a change in the number, or nature, of the dependencies in a CI may affect the occurrence, and size, of cascades.

In this paper we present an approach to modelling CIs, taking dependencies into account. Using the models we follow 2 lines of inquiry. Firstly, we study how the strength of dependencies affects the occurrence and size of cascades in the CIs. Upon varying the strength of dependencies we estimate, via Monte–Carlo simulation, the distributions of cascade sizes for the various CIs. A comparison of these distributions indicates which CI are affected by a change in the strength of the dependence. Secondly, we explore the question of what the consequences of a less detailed model are for modelling the occurrence and size of cascades. Certainly, due to the size and complexity of CIs, it is unreasonable to model "everything" in the real systems. Therefore, given some level of abstraction for the model how much benefit, if any, is gained by using a more detailed level of abstraction?

To illustrate the analysis approach we model interconnected CIs in the Rome area. The data and parameter values for the model are based on:

- a model of CIs in the Rome area developed within the IRRIIS[2] project [9] and inspired by a Telecommunications blackout that occurred in Rome [10-12];
- a *Preliminary Interdependency Analysis* (PIA) carried out to define and limit the scope of the model and to identify dependencies [10-12]. The model scope includes a specification of what the model's level of abstraction should be, which entities should be modelled explicitly, and what the state–spaces for the modelled entities are. The identified dependencies are used to define, in part, how modelled components are correlated. Such data is necessary for a mechanism we use to take dependencies into account in the stochastic models of CI;
- failure and repair rate field–data for Power and Telecommunication network components and equipment. The data was provided by SIEMENS and Telecom Italia [12];

---

[2] Integrated Risk Reduction of Information–based Infrastructure Systems (IRRIIS) is an EU project concerned with developing both a platform for simulating CI and technologies for mitigating against the negative consequences of CI interdependencies.

- realistic parameter values for power network components including voltage levels, thermal limit capacities, and line impedances. This data was provided by SIEMENS;
- a compilation of real–life data on thousands of cascades from all over the world. This cascade data was compiled over several years by TNO (*Netherlands Organization for Applied Scientific Research*) [12].

The outline of the paper is as follows. In Section 2 we discuss our approach to modelling CI. Section 3 outlines the implementation of simulations based on our models. Section 4 discusses the results of simulation-based studies we conducted while Section 5 summarizes conclusions and details directions of further work.

2 Stochastic Modelling of Critical Infrastructures

The stochastic models of CI we build are *Stochastic/Random processes* that are, themselves, made up of dependent *Stochastic/Random processes*. The examples given in Section 1 hint at uncertainty associated with predicting the occurrence and size of cascades. In our model such uncertainty in cascades results from uncertainty in what the next state of each component is, given the current state of the model. Each component in each CI is modelled as a *Random process*. These Random processes, as a consequence of identified dependencies (correlations) within and between the CIs, are probabilistically dependent. So, the interconnected CIs are, themselves, modelled as dependent Random processes. Also, there are factors external to the CIs, such as weather or terrorist attacks, which are modelled as Random processes that interact with the other Random processes. These external, environmental entities can put a strain on the CI, significantly affecting CI operation. So, the dynamical behaviour of the CIs is modelled as a Random process consisting of all of these aforementioned interacting, dependent, Random processes.

To define each Random process we define both a state–space and probabilities that govern the transitions of the component from state to state. In particular, for every distinct pair of states, i and j, we define conditional, *instantaneous transition rates*. These rates parameterize probability distributions used in a *competing risk*s model to determine what the potential next state for a component is. They also determine the potential *sojourn time* for the component; that is, how long before the component enters into its potential next state. For example, in the present work the components have the states {*OK,Failed*}. As a consequence of the components' state–spaces there are 2 instantaneous transition rates associated with each component; *a failure and repair rate*.

The existence of dependencies between components justifies correlation between the components. CI components can be said to be dependent if correlation exists between their states and/or state transitions. While there are other definitions of dependence in use this definition allows us to model a wide class of phenomena. All the dependencies that we have come across in the literature [13, 14] imply correlations. We implement this notion in our models as follows. Given that a component changes state the dependencies specify which other components will have their state–transition behaviour affected. So, there is a notion of *parent* (a component whose state changes af-

fect the stochastic behaviour of some other components) and *child* (a component whose stochastic behaviour is affected by state changes in some other node) components. Each component has 2 related sets; the set of all of its parents and the set of all of its children. Parents can be children and children can be parents. In addition, the parent child relationship can be cyclical so parents can be children of their children (or children can be parents of their parents). By specifying all such parent–child relationships in the model we define a *Directed–Graph* representing all of the pairs of correlated components. We refer to this graph as "*the graph of stochastic associations for the model*". Whenever a parent component undergoes a state change the *failure and repair rates* of its child components take on values that are conditional on the current states of all the components parents. This mechanism is *the primary way in which we model dependence*. The occurrence and effects of various dependence inducing phenomena may be modelled this way, including the consequences of human operator actions, natural disasters, geographic proximity terrorism and weather, on CI operation.

In addition to the primary mechanism for modelling dependence there are certain parent components whose state changes *deterministically* affect the states of some of its children. For example, the failure of a power component could result in the overloading of power lines in the power network. In some of our models which power lines are overloaded is determined by using a "*Linearized, DC, load flow approximation*" [6] to calculate real power flow across the power network. A static load profile (i.e. the consumption and supply of power remains unchanged) is used, as a first step, in our modelling. Another example of modelled deterministic consequences can be found in the Telecommunications network. Each Telecommunications node has a primary power source (power from a local Power distribution company) and a secondary power source (generators or batteries). Given that both the primary and secondary power sources are in failed states the Telecommunications nodes will, with certainty be in an inoperable state.

Component failures may result in failure cascades. Failure cascades can have different causes, may occur over different time–scales, might involve different components and could have different consequences, depending on which network the cascade occurs in. For instance, a sudden surge of power may result in a cascade of power line trips due to power line overloading. The effects of this sort of cascade can be seen almost immediately; some parts of the power network may lose power. Compare this with the aforementioned Buncefield explosion (see Section 1) which caused hospital records to be affected, long after the explosion occurred. Arguably, this cascade is quite different from the previous cascade. However, note two characteristics common to both cascade examples. Firstly, there is an interval of time within which at least one component is in a failed state. Secondly, within this interval of time there is some time point, t, at which there is a maximum number of simultaneously failed components. This suggests the following definition for cascade. Any maximal[3], uninterrupted time interval continuously throughout which at least one component of the relevant sub-network (e.g. Telco Network, Power distribution Network, etc.) is in a

---

[3] In the sense that both the start point and the end point of that interval is either a start or endpoint of the simulation, or is an endpoint outside which the number of failed components is zero.

failed state defines a cascade of that network. Over any such time interval the maximal number of simultaneously failed components is the cascade–size. So, by definition, the size of any cascade is an integer ≥ 1.[4] Certainly, this definition of cascade–size is not unique and may not be the preferred definition of cascade for all situations. For instance, it may be more interesting to "weight" nodes in terms of functional importance, location in the network or economic impact of disruption. Our models allow for the cascade–size to be defined in terms of such alternative approaches, using the "*rewards*" functionality provided by the tool, *Möbius* [15, 16].

3 Simulating Critical Infrastructure

The CI models were used to estimate cascade–size distributions. This is achieved using the *Möbius* tool [15-17]. In particular, the model is created in *Möbius* using the *Stochastic Activity Networks* formalism[5]. This allows us to simulate Continuous–time, Discrete state–space, Random processes using event–driven, Monte–Carlo simulation. Three interacting CIs in Rome were modelled. However, we discuss the results for only the *Telecommunications network* and the *Power distribution network*. The *Power Transmission* network is the 3rd modelled network. We also discuss the results for the aggregated system comprising of all 3 CIs. We refer to this as the "Rome Power–Telco System" or "the entire model". Two studies were conducted using Möbius-based simulations of the Rome Power-Telco System. One study looks into the effect of the "*strength of dependence*" on the distribution of cascade-sizes, while the other study compares modes at different "*levels of abstraction*". Each study consists of a comparison between 2 experiments; a "*base-level*" experiment and a "*comparison*" experiment. The "*base-level*" experiment is the experiment that has been calibrated using all the data sources at our disposal. Consequently, it is the starting point for any comparison experiments as these are only "slight" modifications of the base–level experiment. For the "*strength of dependence*" study the comparison experiment will have parameter values almost identical to the base experiment except that the strength of some dependence is set at a noticeably different level. For the "*levels of abstraction*" study the comparison experiment uses an alternative, less sophisticated algorithm for determining line trips in the power network. So, 3 experiments in total were conducted. Each experiment simulates 10^5 hours of operation (just over 11 years and 4 months) in each of at least 15,000 simulation replications from which sample-mean, cascade–size occurrence rates were obtained. Experiment 1 is the base-level experiment, against which the other experiments will be compared. Experiment 2 changes the strength of certain dependencies while keeping the mean number of cascades in the power distribution network approximately the same as Experiment 1. Experiment 3 substitutes the "*Linearized, DC, load – flow approximation*", used in Experiment 1, for a simple algorithm that governs line trips in the power distribution network. The

---

[4] While this definition implies that a cascade may be "trivial" (the failure of a single node would be defined as a cascade) this is simply a classification choice we have made largely for ease of presentation.

[5] Stochastic Activity Networks are a generalisation of Stochastic Petri–nets.

mean number of cascades in the whole model is kept approximately the same as that for experiment 1.

There are 3 model parameters relevant for the studies. The *Conditional Power-substation Failure rate coefficient* is the amount by which the failure rates of Medium and High voltage substations are scaled when the substations have lost communication with the *Supervisory Control and Data-Acquisition* (SCADA) system. So, this parameter is used to alter the strength of Power components dependence on Telco components for communication. Similarly, the *Conditional Telco-component failure rate coefficient* is the amount by which the failure rate of Telco nodes that have lost their primary source of power supply, but still have a secondary power source, is scaled. So, this parameter is used to model the strength of Telco components' dependence on power sources for their operation. Finally, the *Conditional probability of power-line overloading* is the conditional probability of a given power-line overloading, given the failure of some other power component in the Power Network. This parameter is used in experiments that do not use the *Linearized, DC, load flow approximation* to determine power–line overloading.

4 Discussion of Simulation Results

For the base–level experiment the *Conditional Power-substation Failure rate* coefficient has a value of 10^3 and the *Conditional Telco-component failure rate* coefficient has a value of 10^5. Notice, from Fig. 1, that about 10 cascades of size greater than 4 occur in the combined Telco–Power system. However, the Power distribution network has about 8.87×10^{-3} of its cascades having a size greater than 4. So, the Power distribution network appears to contribute relatively little to the cascade size for the larger cascades. In contrast, there are about 2.53 cascades of size greater than 4 that occur in the Telco network; two orders of magnitude more than the related number for the Power distribution network. While this does not fully account for the number of cascades greater than 4 in the combined model it suggests that a significant number of relatively large cascades occur in other parts of the Rome model not depicted, i.e. the Power transmission network.

The "Strength of dependence" experiment explores the dependence of Power components on Telco components and vice–versa. The value for the *Conditional Power-substation Failure rate coefficient* is set to 4800, which is higher than the corresponding value of 1000 in the base–level experiment. So, each power sub–station's dependence on Telco–services in this experiment is almost 5 times stronger than in the base–level experiment. Contrastingly, the *Conditional Telco-component failure rate coefficient* parameter is set to 1.0 for this experiment, where it had a value of 10^5 in the previous experiment. So, each Telco component experiences a dependence on Power services that is *5 orders of magnitude weaker* than it was in the previous experiment. The DC approximation to AC power flow is still used. The resulting Cascade–size distribution is given in Fig. 2. The parameter values were chosen so that the mean number of cascades in the entire model, 286, is comparable with the respective value in the base–level experiment, 285.
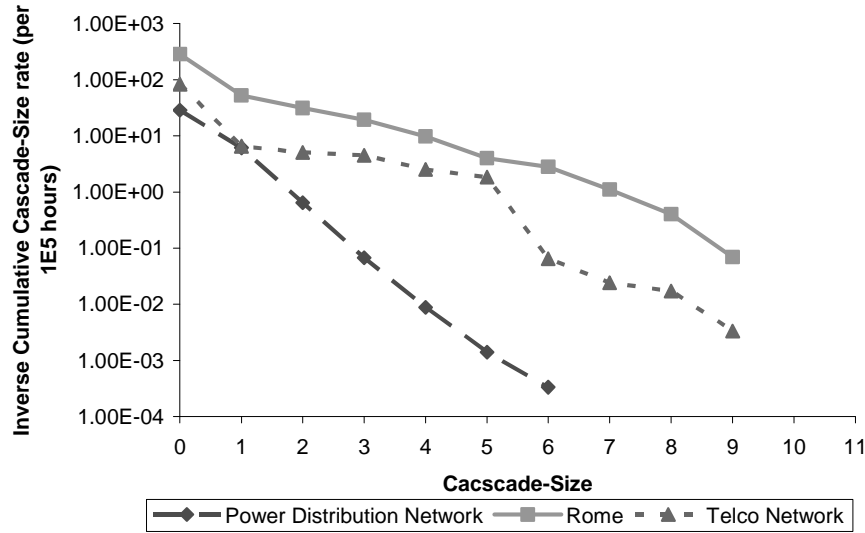
Fig. 1. For the base–level experiment this graph depicts the sample–mean number of cascades of size strictly greater than the indicated size (on the horizontal axis) that occur in each sub-network. For instance, the mean number of cascades that occurred in each of the sub–networks is given by the respective value of each graph corresponding to the cascade–size value of 0. So, the mean number of cascade events in the Power distribution network is 28.72, the mean number of cascade events in the entire model is 285.12 and the mean number of cascade events in the Telco network is 83.18.

The Power distribution network still contributes relatively little to the large cascades. There is no noticeable change, even though a change exists, in the contribution of Power distribution cascades to the total number of cascades in the model. The Telco network appears to exhibit a rapid drop from a frequency of 75.1 for single failures to a frequency of 9.13×10^{-3} for triple failures. So, there were hardly any cascades of size 3 among the cascades in the Telco network. However, this steep fall is followed by a steep rise so that there are an estimated 3.4 cascades of size 6 occurring; *a change in cascade–size rate of 2 orders of magnitude* between cascade–size 2 and cascade–size 6. While in both experiments a significant number of cascades of size 6 occur in the Telco network it would seem that the effect of having weaker dependence on Power–services in the current experiment is to reduce the number of cascades of size 3. The cascade–size distributions for the entire model are virtually identical, up until cascades greater than size 5 (see Fig. 3). Only in the extreme right of the graph are the curves an order of magnitude apart. So, the Telco nodes' having 5 orders of magnitude less of a dependence on Power components has little effect, globally, on the distribution of cascades if compensated by a comparatively modest increase in the dependence of Power nodes on Telco services. In Fig. 3 even though the

graphs are arguably similar there is, nevertheless, a strict ordering between the graphs for "large" cascades (cascades with sizes greater than 6). "Large" cascades are strictly less likely to occur in the base–level experiment. Similar orderings are visible and more pronounced for the cascade–size distributions in both the Telco network and the Power distribution network. In particular, consider the Power distribution network (see Fig. 4), where the ordering is maintained and, additionally, the distributions appear to diverge. One of the distributions is fairly linear on a log-linear plot and the other has approximately 2 linear regimes (one between cascade sizes 1 and 3 and the other between 4 and 6). This suggests 3 exponential laws governing the cascade-size distributions, with one of the distributions having 2 distinct laws characterizing it. Statistical tests of significance and estimating the parameters for these laws is of immediate interest. In summary, globally there appears to be little effect on cascade – size distributions from a weaker dependence of Telco components on Power services and stronger dependence of Power nodes on Telco services. Locally, however, there are differences between the distributions.
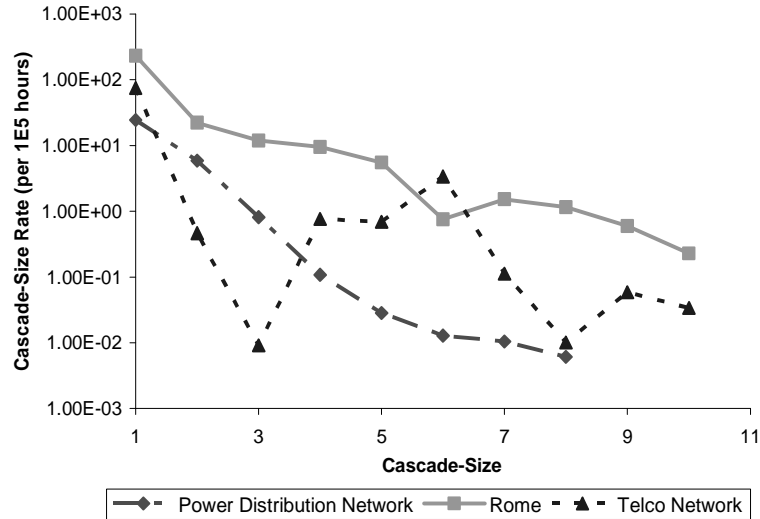

Fig. 2. Cascade–size distribution for each sub–network in the "Strength of dependence" experiment

In the "levels of abstraction" experiment we use a relatively naïve algorithm for power–line overloading on the cascade–size distributions, replacing the *Linearized DC approximation to AC power flow*. A Binomial probability distribution randomly trips power–lines, given the failure of a power component. For the occurrence of cascades in each network will it matter that we are using an extremely simplistic algorithm to model power–line trips? If it does, how significant is the change in model behaviour? The value of the parameter *Conditional probability of power-line overloading*, which we set as 3.2×10^{-3} for this experiment, defines the Binomial distribution. All other parameters are the same as the base–level experiment. The mean number of cascade events in the Power distribution network is $28 \cdot 48$, which is

comparable with $28 \cdot 72$ for the base–level experiment. The distributions for the Telco network are visibly almost identical (see Fig. 5). On the other hand, the power distribution network exhibits significant differences. A clear ordering exists, with the base–level experiment having an order of magnitude more occurrences of "large" cascades. The shape of the distributions both appear to be approximately linear and parallel on a log-linear plot. Thus, again the data suggests an exponential relationship characterising the cascade size-distributions.

In conclusion, we note that while a significant change in strength of dependence noticeably affected the Cascade-size distributions for the sub-networks, the impact on the whole model was pronounced only for very large cascades. Also, the use of an extremely simplistic algorithm for power line trips had a significant impact in the power distribution network but was negligible in the Telco network.
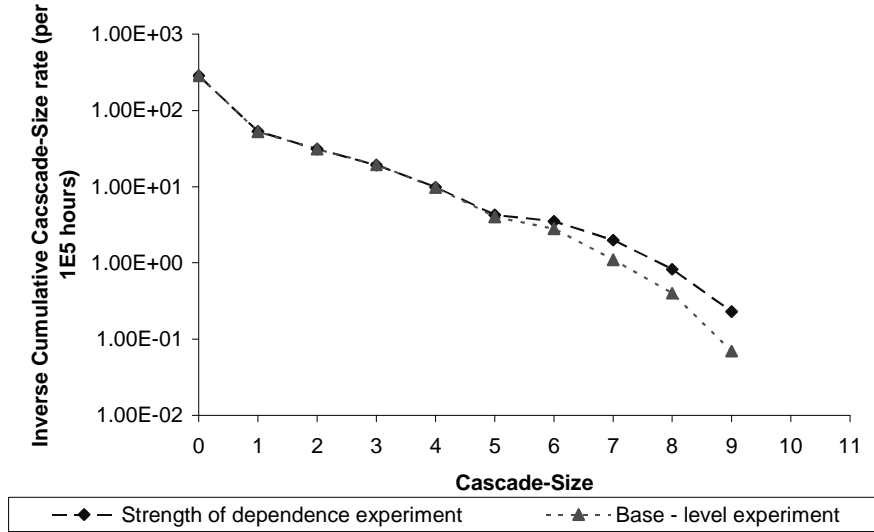

Fig. 3. Comparison of the Inverse, Cumulative Cascade–size rates for the Rome model.

5 Conclusion and future work

This paper presents a quantitative approach to modelling interdependencies in CIs and results from applying the approach to a non-trivial case-study, the Rome Power-Telco system developed within the EU IP IRRIIS.

We argue in favour of using probabilistic models of interdependencies and provide details of the model of 'stochastic associations' we developed. Such models not only allow the user to incorporate any knowledge that might exist about the likelihood and

dependence between various adverse events (e.g. failures of the modelled component) and risks of environmental disruptions (e.g. natural disasters, extreme weather conditions, etc.) but also to study the impact and challenge various assumptions about these.
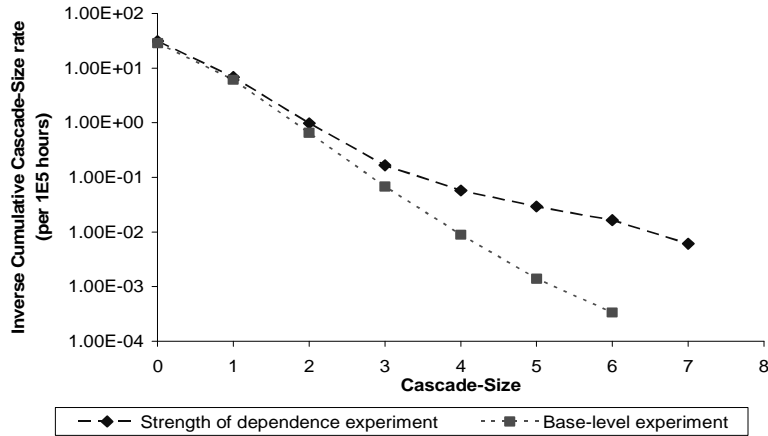

Fig. 4. Comparison of the Inverse, Cumulative, Cascade–size rates for the Power distribution network.
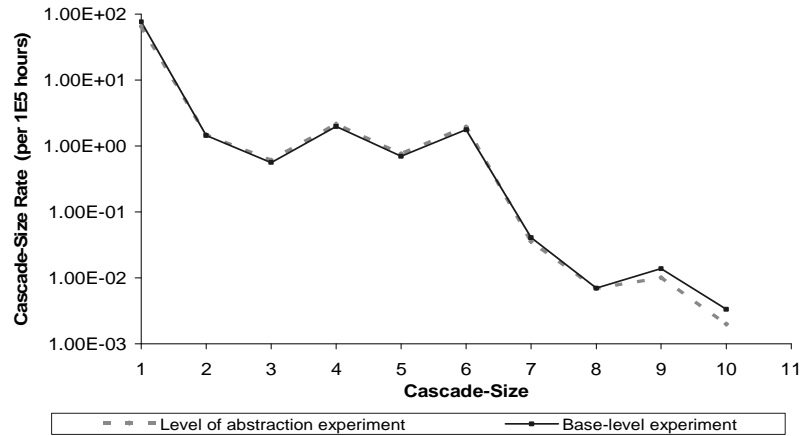

Fig. 5. Comparison of Cascade – size distributions for the Telecommunications network.

The results presented target primarily the impact of the level of abstraction on the modelling results, an important scalability issue with very large CIs. We report, that

in some cases the behaviour of 'low fidelity' models is close to the behaviour of more sophisticated models (i.e. low fidelity models can offer an acceptable accuracy). This result seems important, because it opens up a practical way of modelling very large CIs with a reasonable accuracy.

In the light of these results we propose a future modelling approach in which a network of CIs is decomposed into manageable parts. Comparative studies (between high and low fidelity models, as described above) of these parts may then be applied so that the low-fidelity models can be tuned to model accurately the behaviour of the modelled sub-systems. Then, from the perspective of each CI, the network of CI may be modelled by combining acceptable low-fidelity models of some parts with high fidelity models of other parts. Although such an approach seems plausible further work is needed to validate that the thus composed model will still be accurate. We intend to attack this issue in our future work.

In addition, the work demonstrates that a range of parameter values can give similar model behaviour, depending on what aspects of the model are of interest. Very different parameter values gave the same total number of cascades in the model. However, the *tails* of the cascade-size distributions for the whole model and the power distribution network exhibited divergence. This suggests that in parameterizing such models for use in practice care must be taken since the effect of different parameter values is to produce "regimes of agreement" between the models. In the case we observed the models may diverge for large cascades but agree reasonably well for small cascades. The magnitude of this effect differed between networks.

In the current work, as a first approximation, static load profiles were used in the study. Dynamic load profiles that capture daily, or seasonal, variations in load may have the effect of significantly changing the probability of large-scale cascades. We will experiment with, and study the effects of, dynamic load profiles as an extension of the current work.

A related problem is addressing aspects of 'observability' of the state of the entire CI. The Rome system consists of networks, operated by organisations which may have detailed knowledge of their own network but very limited knowledge of the state of the networks operated by other operators. Approaches for dealing with such limited observability are of practical interest. We have scoped in [12] an approach to on-line risk estimation (RE) based on the probabilistic models described in this paper. Validating RE in terms of achievable accuracy of risk predictions (i.e. whether the periods with predicted high risk of disruption will indeed tend to be highly positively correlated with actual disruptions) is a problem, which we are currently working on.

In addition to the relative confidence intervals used for determining acceptable convergence of the estimated distribution data points further statistical tests of significance will be carried out on the data to increase confidence in the results. Also, the data suggests a number of exponential relationships governing some of the cascade-size distributions. The parameters for these laws will be estimated as part of future work.

6 References

1. Bloomfield, R., et al., *Report on Service Oriented Interdependency Analysis (IRRIIS Project Deliverable D2.2.4)*. 2007, City University London: London.
2. Bloomfield, R., et al., *Infrastructure interdependency analysis: an introductory research review*. 2009, Adelard: London.
3. Boccaletti, S., et al., *Complex networks: Structure and dynamics*. Physics Reports, 2006. **424**: p. 175-308.
4. Reka., A. and A.L. Barabasi, *Statistical mechanics of complex networks*. Reviews of Modern Physics, 2002. **74**.
5. Pederson, P., et al. *Critical Infrastructure Interdependency Modeling: A Survey of U.S. and International Research*. 2006 [cited INL/EXT-06-11464].
6. Schläpfer, M., T. Kessler, and W. Kröger. *Reliability Analysis of Electric Power Systems Using an Object-oriented Hybrid Modeling Approach*. in *Proceedings of the 16th Power Systems Computation Conference*. 2008. Glasgow.
7. U.S.-Canada Power System Outage Task Force. *Final Report on the August 14, 2003 Blackout in the United States and Canada: Causes and Recommendations*. 2004 [cited; Available from: https://reports.energy.gov/.
8. Board, B.M.I.I. *The Buncefield Incident, 11 December 2005: The final report of the Major Incident Investigation Board*. 2004 December 2008 [cited; Available from: http://www.buncefieldinvestigation.gov.uk/index.htm.
9. IRRIIS. *Integrated Risk Reduction of Information-based Infrastructure Systems*. 2006-2009 [cited.
10. *IRRIIS deliverable D2.2.4.,``Report On Service Oriented Interdependency Analysis''*. 2007 [cited; Available from: http://www.irriis.org.
11. *IRRIIS deliverable D2.2.2., ``Tools and techniques for interdependency analysis''*. 2007 [cited; Available from: http://www.irriis.org/File.aspx?lang=2&oiid=9138&pid=572.
12. *IRRIIS deliverable D2.2.6, "Preliminary Interdependency Analysis (PIA) as a service-oriented approach towards LCCI Interdependency Analysis (report and prototype)"*. 2009.
13. 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.
14. Pearl, J., *Probabilistic Reasoning in Intelligent Systems*. 1988: Morgan Kaufmann, Palo Alto, CA.
15. Sanders, W.H., et. al., *The Mobius Manual*. 2008.
16. Sanders, W.H. and J.F. Meyer. *Stochastic Activity Networks: Formal Definitions and Concepts*. in *Lectures on Formal Methods and Performance Analysis, First EEF/Euro Summer School on Trends in Computer Science*. 2000. Berg en Dal, The Netherlands: Springer, Berlin, 2001.
17. Courtney, T., et al., *The Möbius Modeling Environment*. Tools of the 2003 Illinois International Multiconference on Measurement, Modelling, and Evaluation of Computer-Communication Systems, 2003.