CITY UNIVERSITY LONDON

Stochastic Modelling of Cyber Attacks in Industrial Control Systems

Dr Peter Popov

In collaboration with Robin Bloomfield, Aleksandr Netkachov and Kizito Salako

Centre for Software Reliability

City University London

p.t.popov@city.ac.uk

College Building, City University London

Northampton Square, EC1V 0HB

Tel: +44 207 040 8963 (direct)

+44 207 040 8420 (sec. CSR)

CSR Building confidence in a computerised world

www.csr.city.ac.uk

Talk outline

- Risk analysis of complex industrial systems
  - Complexity makes the analysis very difficult
    - Identifying hazards and all “interesting events” is very difficult
  - Stochastic models are a way of addressing this difficulty
- Preliminary Interdependency Analysis
  - Method, Modelling dependencies, Parameterisation
- Tool support
- Modelling complex industrial control systems
  - NORDIC32 + a model of protection and control based on IEC 61850
  - Model of an Adversary
  - Simulation engine
  - Results
- Conclusions and Future work

Projects relevant to work

Sponsored by:

- EU: SESAMO (2012-2015) (Security and Safety Modelling)
- EU: AFTER (2011-2014) (A Framework for electrical power systems vulnerability identification, defence and restoration)

A new grant has just been announced:

- UK EPSRC Research Institute in Trustworthy Industrial Control Systems, "Communicating and evaluating cyber risk and dependencies" (2014 - 2017)

Based on:

- EU: IRRIIS (2006-2009) (Integrated Risk Reduction of Information-Based Infrastructure Systems)
- PIA:FARA (2009 - 2010) (Probabilistic Interdependency Analysis: framework, data analysis and on-line risk assessment), funded by the UK Technology Strategy Board (TSB).

Critical Infrastructure Interdependencies

- A key issue for achieving CI resilience and CI protection
  - risk of CI disturbances propagating across dependencies' links
- A complex phenomena, yet not well understood

Buncefield explosion

Geographical dependencies
Infrastructures affected due to proximity of explosion site

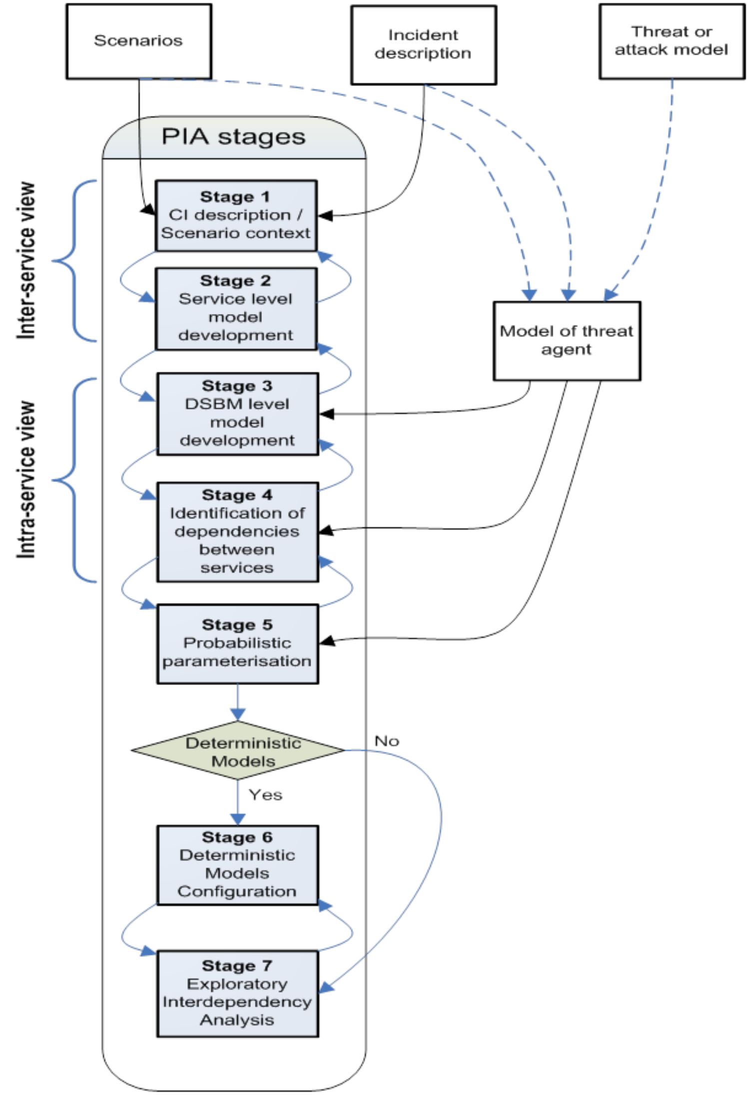
Transport:	smoke affected visibility at Heathrow, M1 closed for two days
Energy:	explosion destroyed adjacent business park incl. 92 companies (damages over £70m)
Information infrastructure:	headquarters of IT company destroyed by blast, with multiple cascading effects

Information infrastructure dependencies
Cascading effects of the damage sustained by Northgate Information Solutions

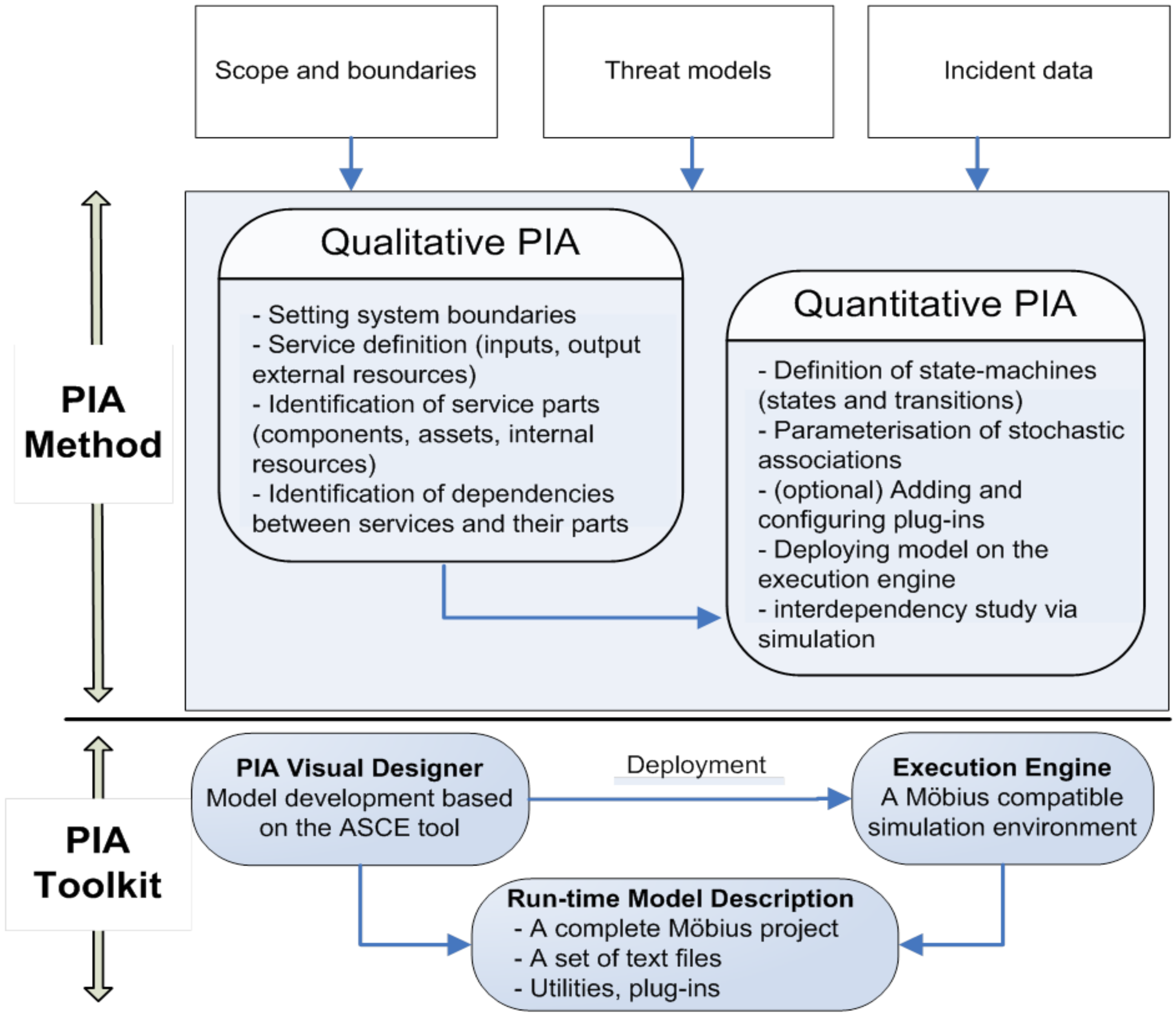
Health:	five hospitals lost access to patient records and admission/discharge systems and reverted to manual systems for a week
Finance:	£1.4 billion payroll scheme lost due to explosion — recovered in time

PIA - Interdependency Analysis

- PIA is an approach (method) to interdependency analysis which consists of two steps
  - **Preliminary** Interdependency Analysis (Pre-IA) – HAZOP like analysis of interdependency ***discovery***
  - **Probabilistic** Interdependency Analysis (Pro-IA) – ***quantitative model*** of interacting CIs, each represented as a collection of services, which in turn may have their own network and components:
    - Typically very large number of components (***hardly amenable to analytic solutions)***,
      - parameterization becomes problematic)
    - Probabilistic behaviour (rates/distributions of Time-To-Failure and Time-To-Repair)
    - Engineering (typically deterministic) models (e.g. various flows models) are needed for high fidelity studies.

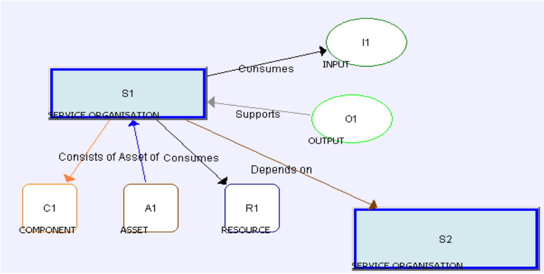


An overview of the PIA method



Preliminary Interdependency Analysis (Pre-IA)

- ‘Preliminary’ because one should start by establishing ***basic understanding***
- Service oriented, systematic elaboration of model components
  - “Quick and easy wins” rather than expensive and time-consuming detailed modelling and analysis
  - HAZOP style Identification of dependencies of assets/components/resources within and across organizations/departments
  - Basis for more detailed models



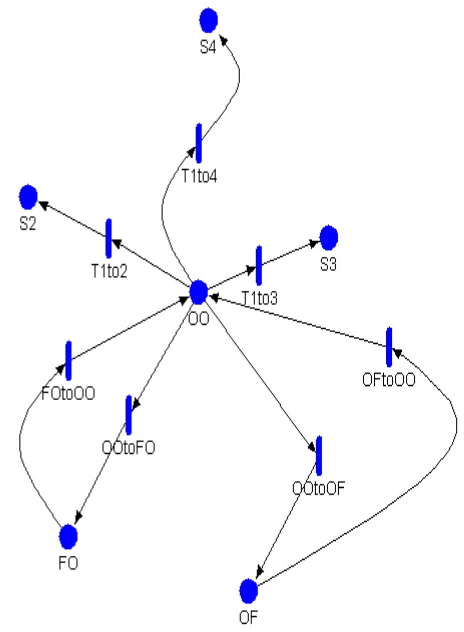
- Examples
  - Rome telecommunications incident (developed in IRRIIS)

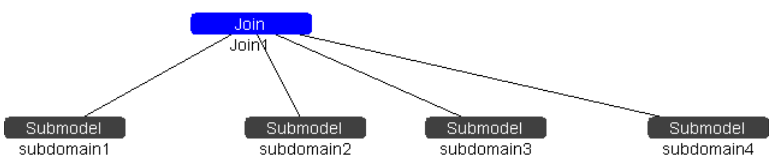
Probabilistic PIA (Pro-IA)

- We deal with both uncertainty in the real world (aleatory) and in our knowledge of it (epistemic)
  - behaviours, structures (especially for Information infrastructures)
- The measures of interest are probabilistic
  - overall aggregated ***risks*** (e.g. size of cascades vs. frequency)
  - probability of specific events (e.g. service loss, failure scenarios, "weakest link")
- Pro-IA allows for modelling approximations and efficiencies
  - consequence and environment models, infrastructure models
  - explore cascade mechanisms
  - can explore many thousands situations (very large state space)
  - can search for interesting cases, link to trials/demos
- important role to ***complement*** deterministic, qualitative, trails and analytic approaches

Pro-IA models

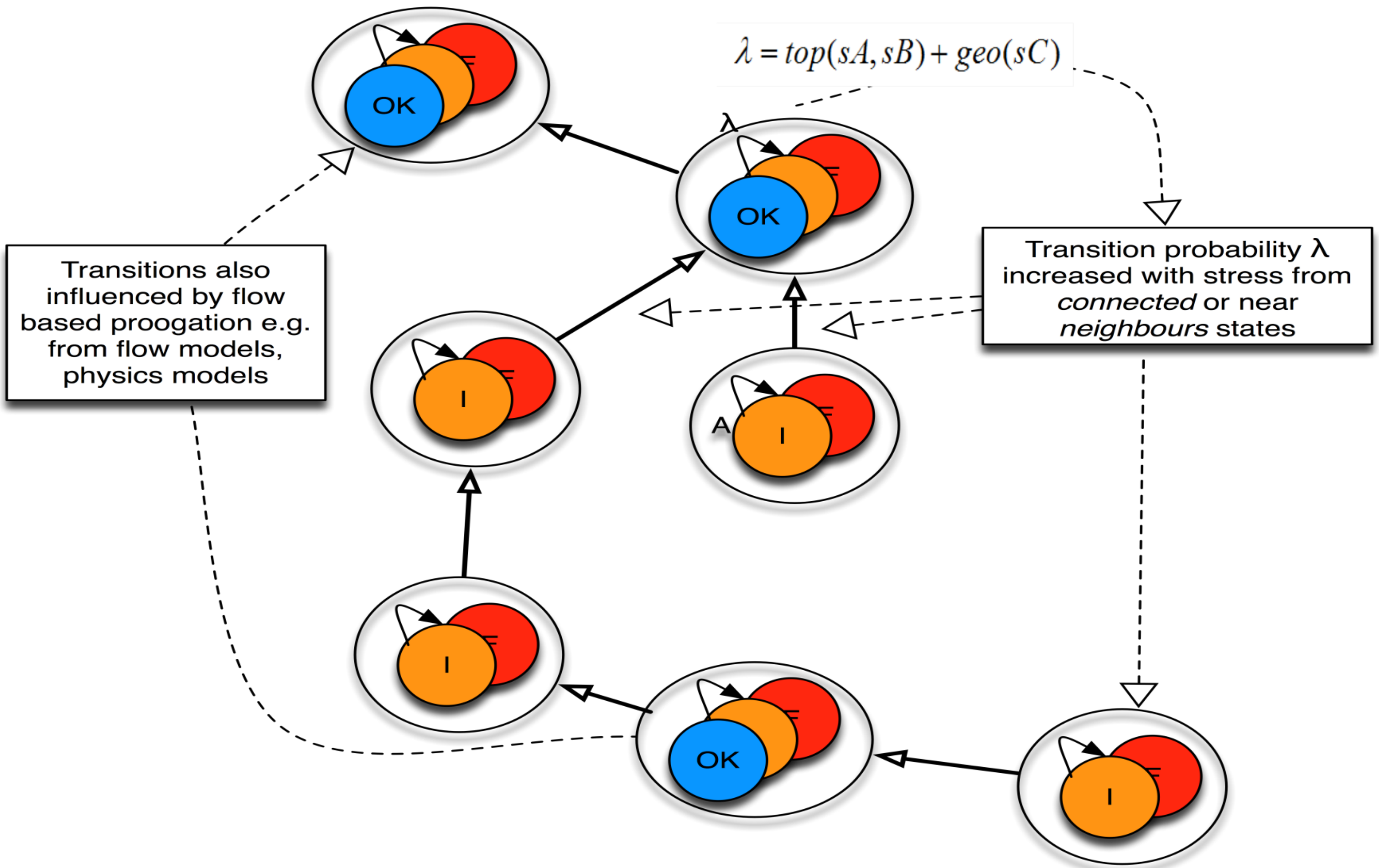
- We used SANs (stochastic activity networks) and Möbius Modelling Tool (by the performability group at the University of Illinois at Urbana Champaign, USA) to define parameterised continuous time semi-Markov models
- Finite state atomic components that "interact" with each other to make ***impairment*** and failure "contagious":
  - Each component is modelled as a state-machine (a semi-Markov process)
  - rates (distributions) of transition between states are ***functions of the states of the 'neighbour' components*** ("model of stress").
- Embedded deterministic sub-models that can relate the "dynamics" of some subsets of the components on the state of other subset of components, e.g.:
  - DC/AC approximate power flow model for power flow components
  - telecommunication service model.
- Components coupled via geographic location.
  - Spatial dependencies are important
  - ***BUT not the only ones worth studying! (design faults, viruses are not spatial)***





PIA approach to modelling (inter)dependencies

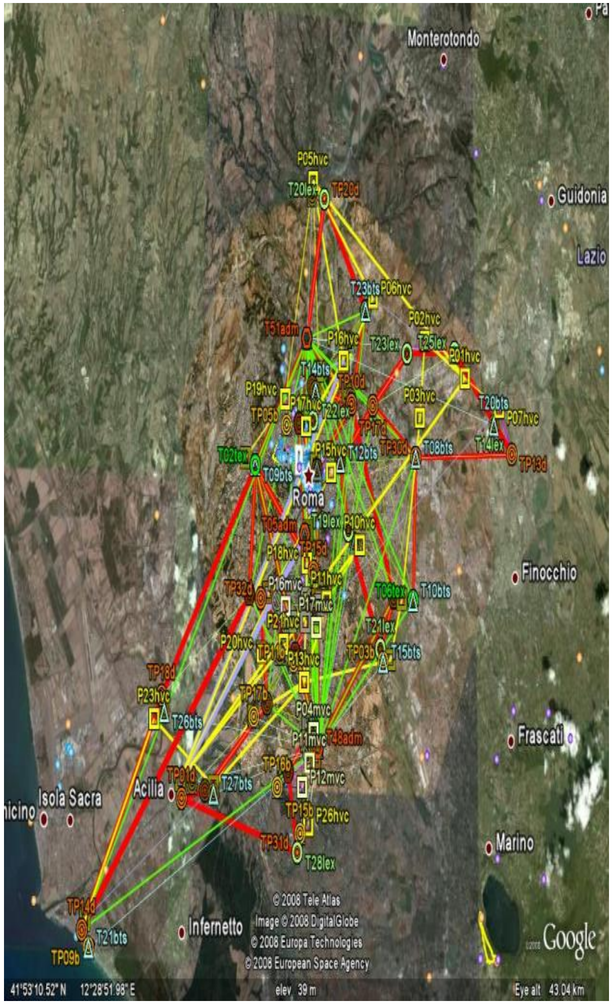
Stochastic associations - sources of dependency and cascades



CITY UNIVERSITY LONDON

The Rome Scenario

- Service layer – 5 services:
  - Power Grid: Power Transmission and Power Distribution
  - Telecommunications: Fibre-optics network, fixed lines telephony, GSM
- Physical layer;
  - **830 modelled physical elements** - nodes and links (high-voltage cabins, trunks, fibre cables, transmitters, gateways)
- Dependencies
  - deterministic based on functional dependencies (telecommunications need power, power components controlled remotely via telecommunication channels)
  - stochastic associations – spatial proximity and cross-CI functional dependencies;
  - Non-probabilistic models (causality, flow models which may lead to overloading and tripping)
- Parameter values;
  - Probabilistic models: Failure rates, Repair rates
  - Deterministic: flows, capacity (of lines, batteries), power load, voltage levels, line resistance (ETHZ);

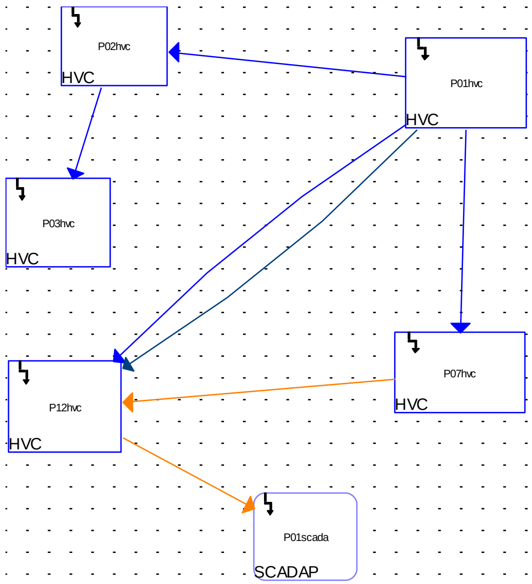


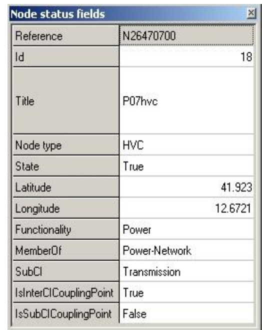
PIA:FARA Toolkit Prototype

- The toolkit consists of:
  - PIA Designer – an interactive tool to allow a modeller to 'design' an interdependency study.
    - Supported by Adelard's ASCE visual editing tool (designed to support documenting safety-cases and customised for the needs of PIA)
  - PIA Run-time support – execution environment based on the Möbius tool (and in particular its SAN formalism) with very ***extensive customisation***
- PIA Designer - a 2-layer approach:
  - ***Intra***-services model - networks behind the individual services are explicitly modelled (as SANs with dependencies between the modelled elements)
  - ***Inter***-services model – explicitly models (inter)dependencies between the services that belong to different Intra-service models;
    - Coupling points – path for interdependencies to propagate between services;
  - Deterministic models added via ***plug-ins*** to the system at run-time (DLLs and initialisation files, e.g. XML)
  - Exporting the model for 'execution' on a run-time environment such as Möbius's SAN execution engine.
  - Visualisation of the probabilistic model simulation traces (using the Möbius built-in provisions or custom built utilities)

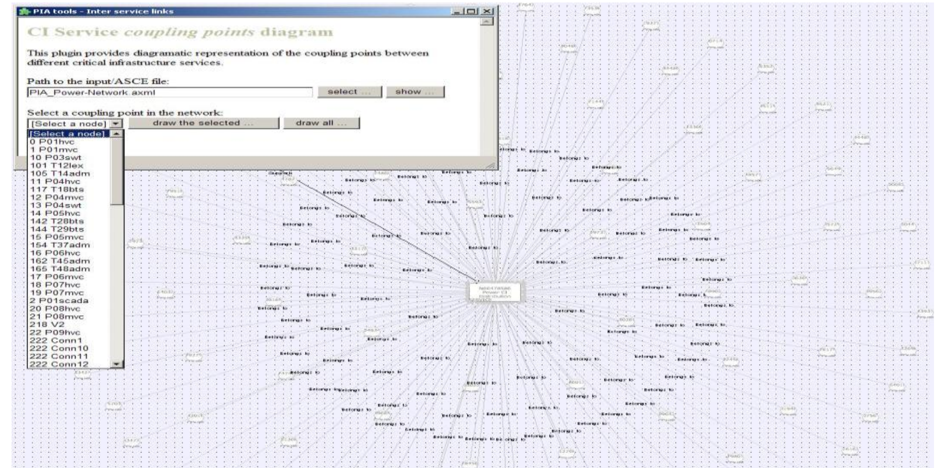
CITY UNIVERSITY LONDON

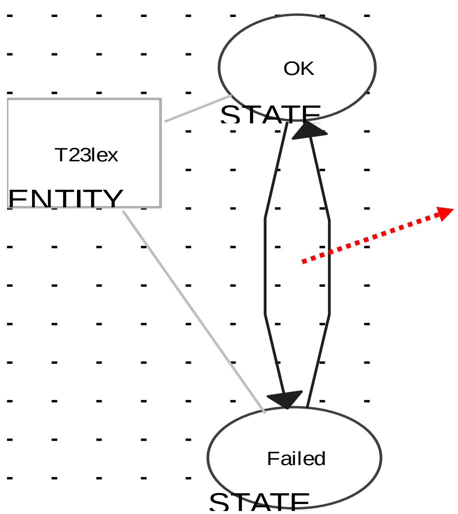
PIA:FARA Toolkit

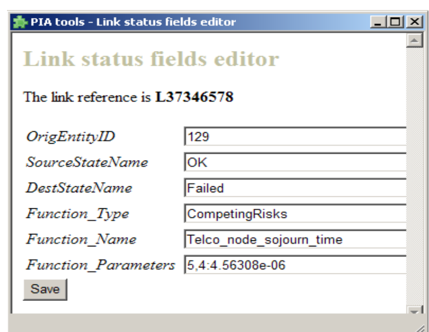


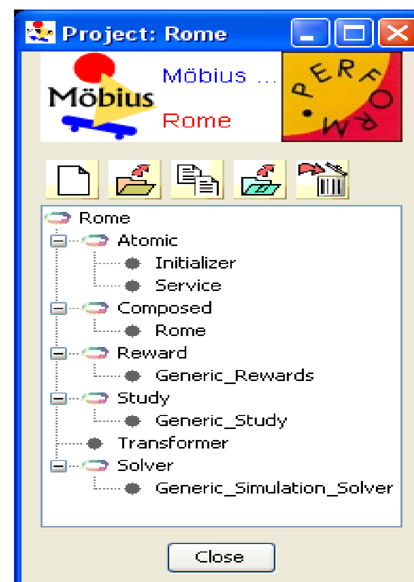


Reference	N26470700
Id	18
Title	P07hvc
Node type	HVC
State	True
Latitude	41.923
Longitude	12.6721
Functionality	Power
MemberOf	Power-Network
SubCI	Transmission
IsInterCICouplingPoint	True
IsSubCICouplingPoint	False



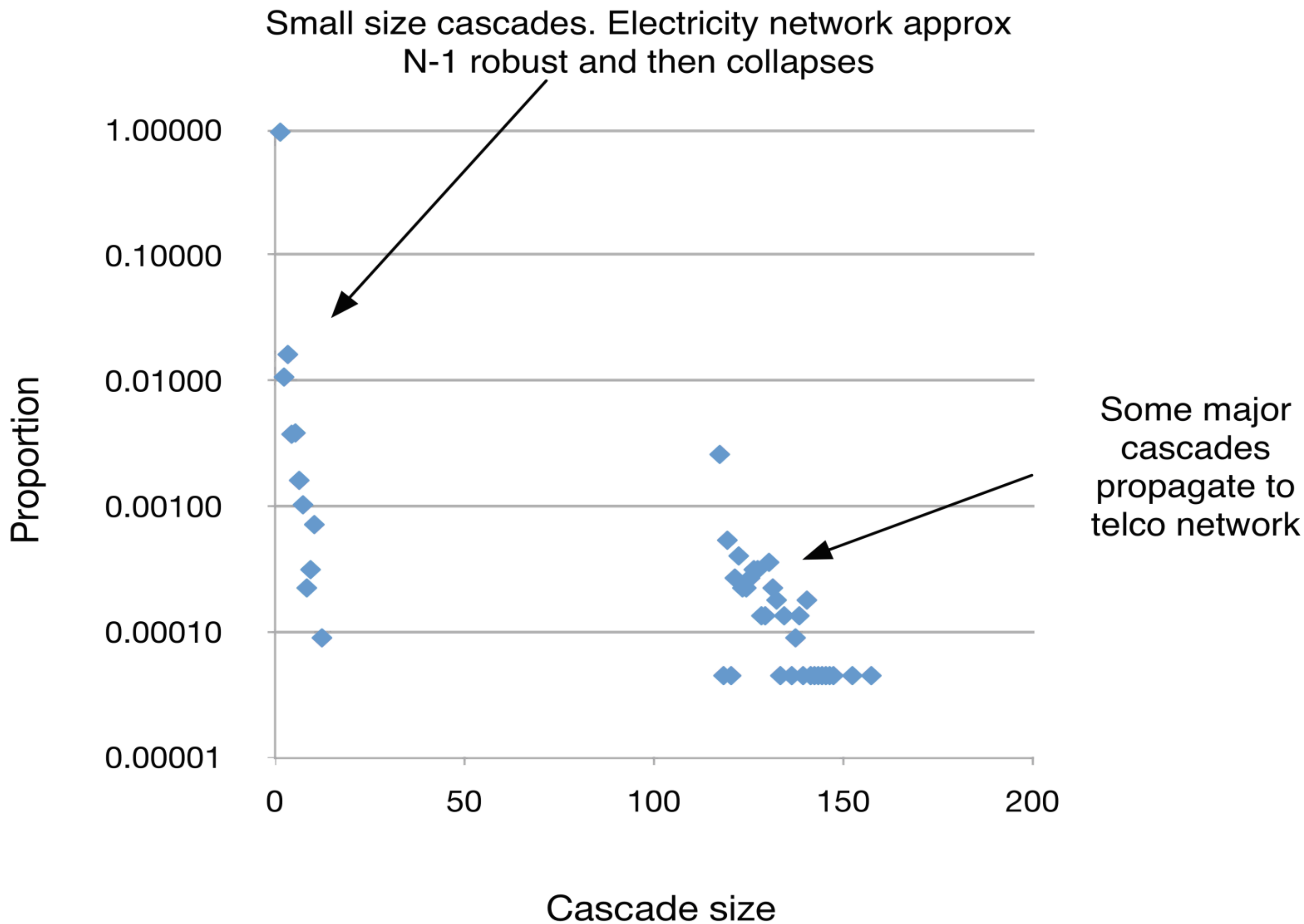






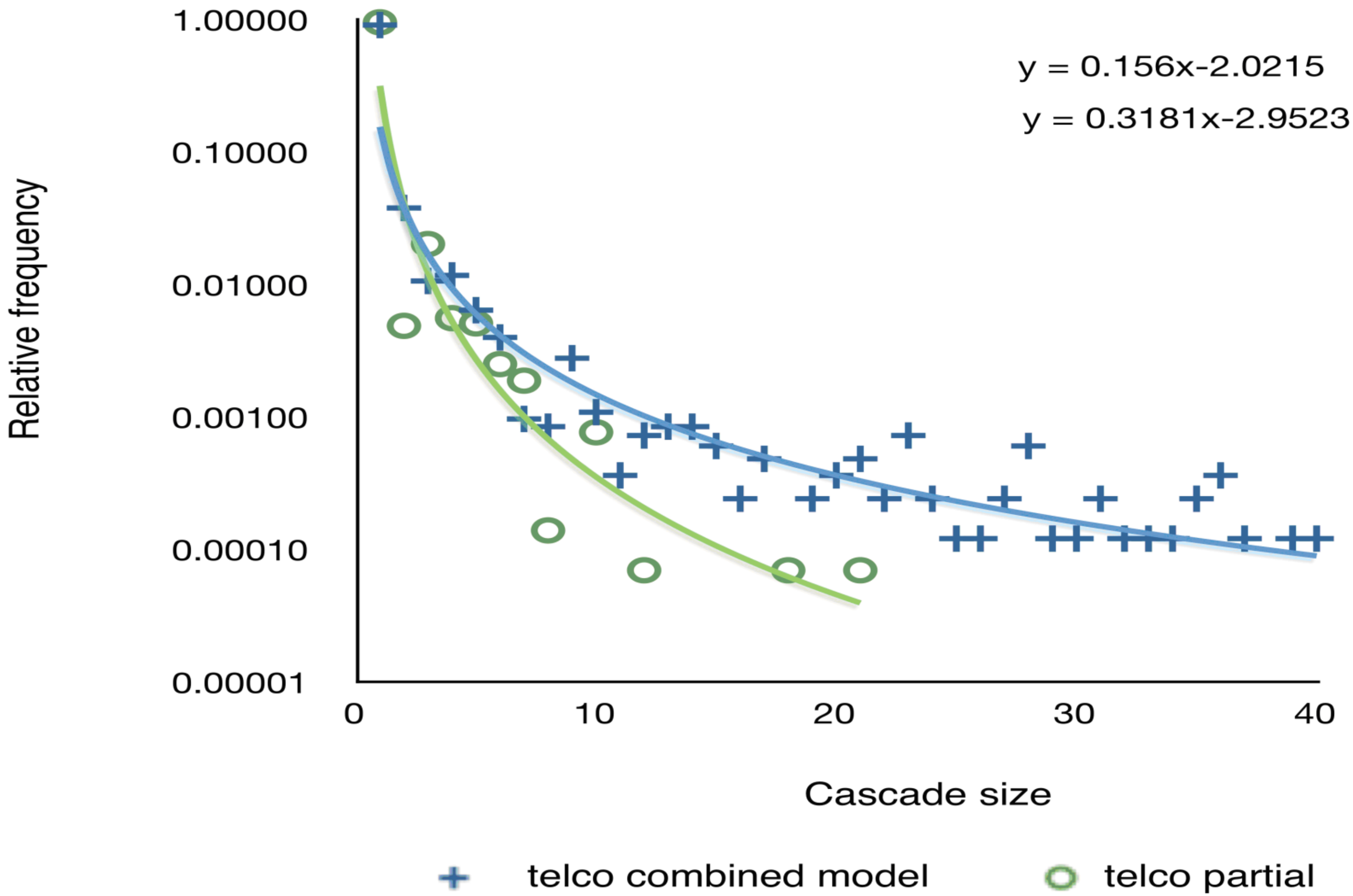
CITY UNIVERSITY LONDON

Results



CITY UNIVERSITY LONDON

Results (2)

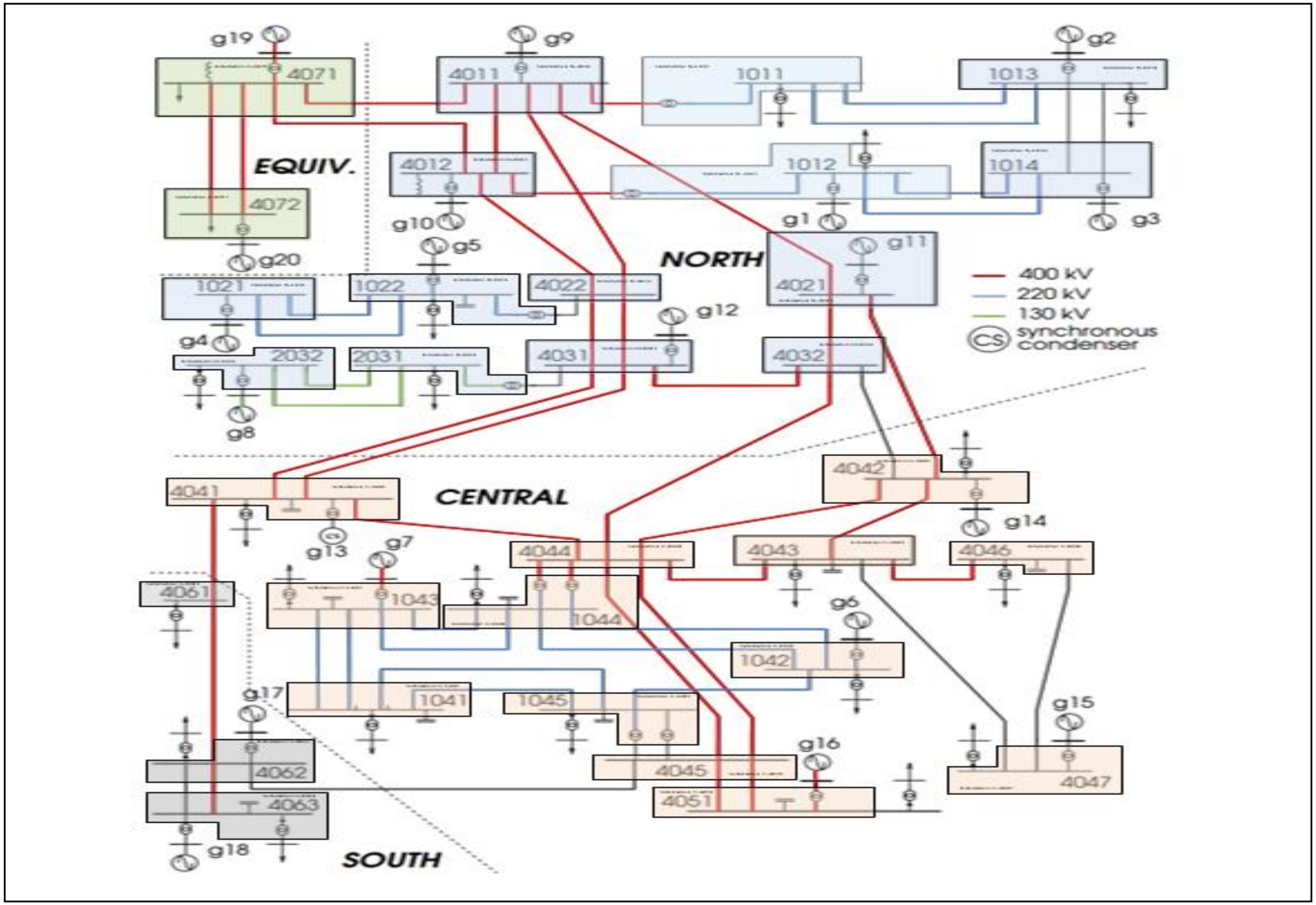


AFTER / SASAMO case study

NORDIC 32

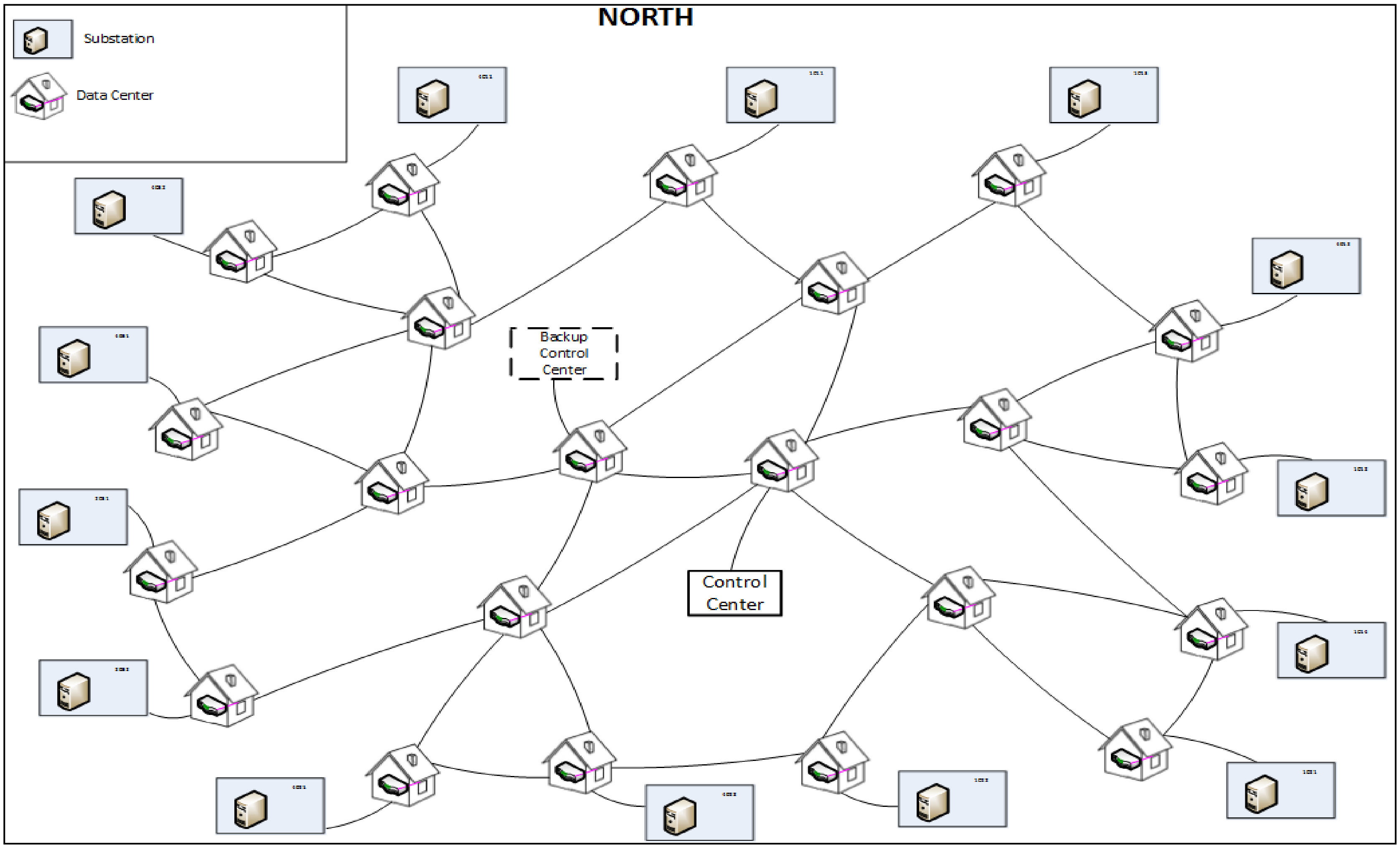
- Power transmission network – a reference network used widely in research
  - 32 sub-stations (more details are provided later)
- ICT network
  - SCADA system modelled at ***high level of abstraction***
  - Control network in substations is compliant with IEC 61850 (an international standard defining an architecture and communication stack for substation protection and control)
- Model of cyber attacks
  - Model of an Adversary adapted to the specific context
- The PIA principles applied:
  - Stochastic dependence between the modelling elements
  - Hybrid models (i.e. stochastic and deterministic, e.g. Power flows)
  - Rewards – specific to the context, e.g. the power loss due to accidental failures or malicious activities, probability of large cascades.

NORDIC 32



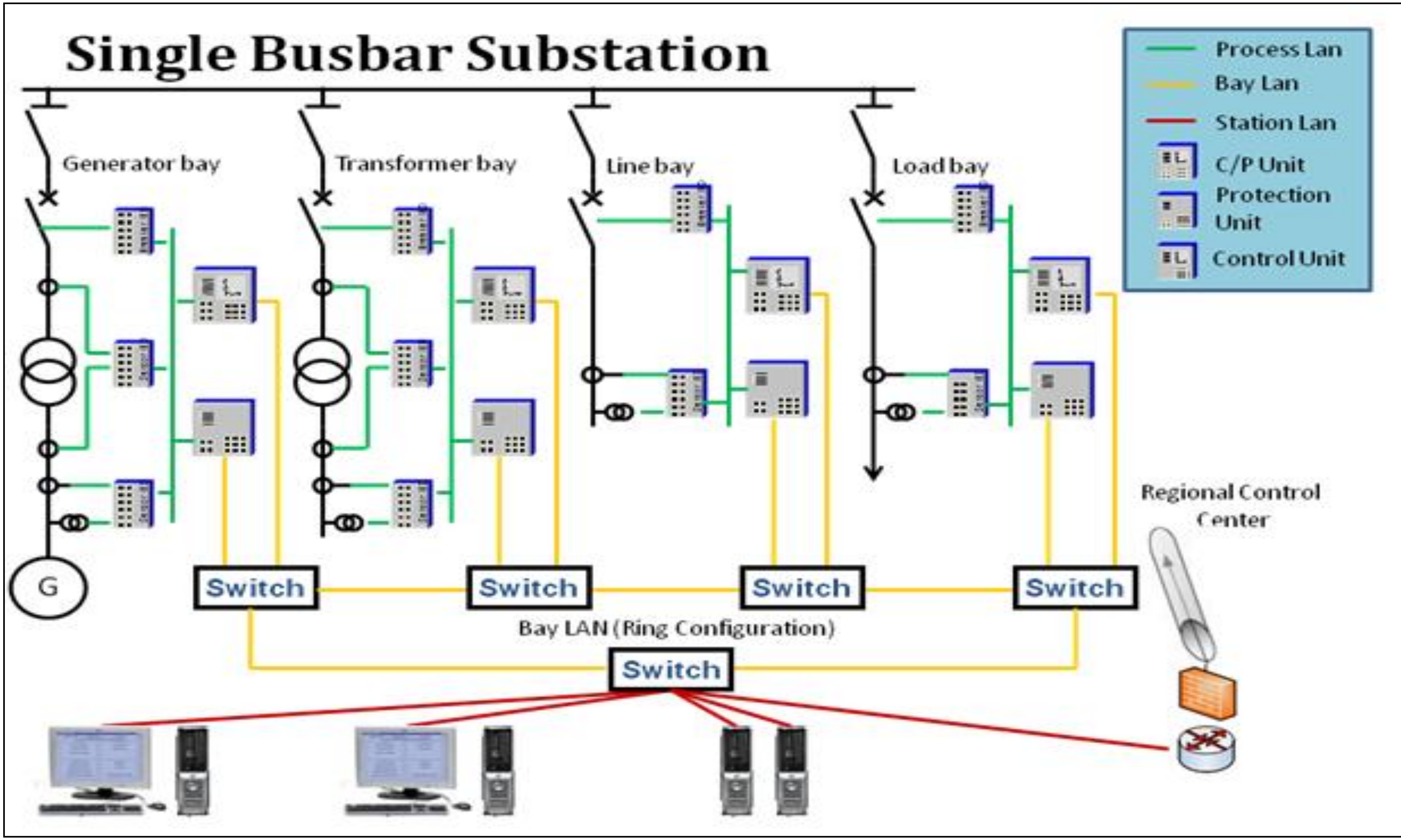
g19
g9
g2
4071
4011
1011
1013
EQUIV.
4012
1012
1014
4072
g10
g1
g3
g5
g11
g20
NORTH
1021
1022
4022
4021
400 kV
220 kV
130 kV
CS
synchronous condenser
g12
g4
2032
2031
4031
4032
g8
4042
4041
CENTRAL
g14
g13
g7
4044
4043
4046
4061
1043
1044
g6
1042
g17
1041
1045
g15
g16
4062
4045
4047
4063
4051
g18
SOUTH

ICT system


NORTH
Substation
Data Center
Backup Control Center
Control Center

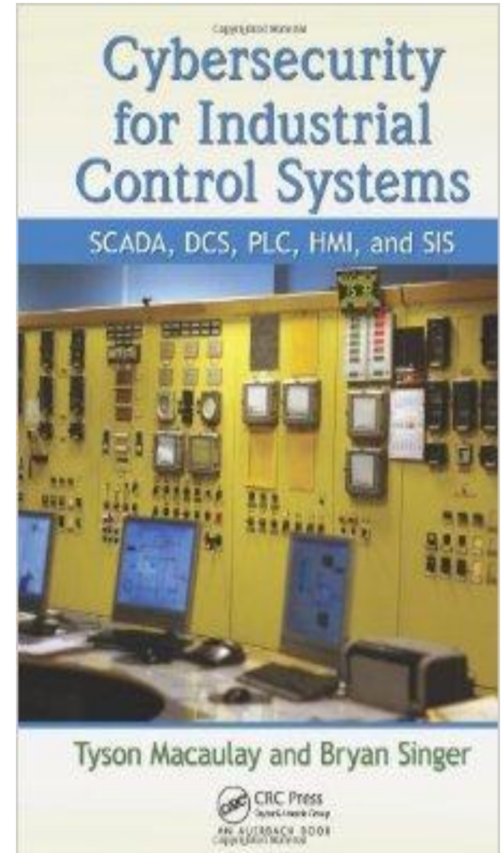
CITY UNIVERSITY LONDON

Sub-station model

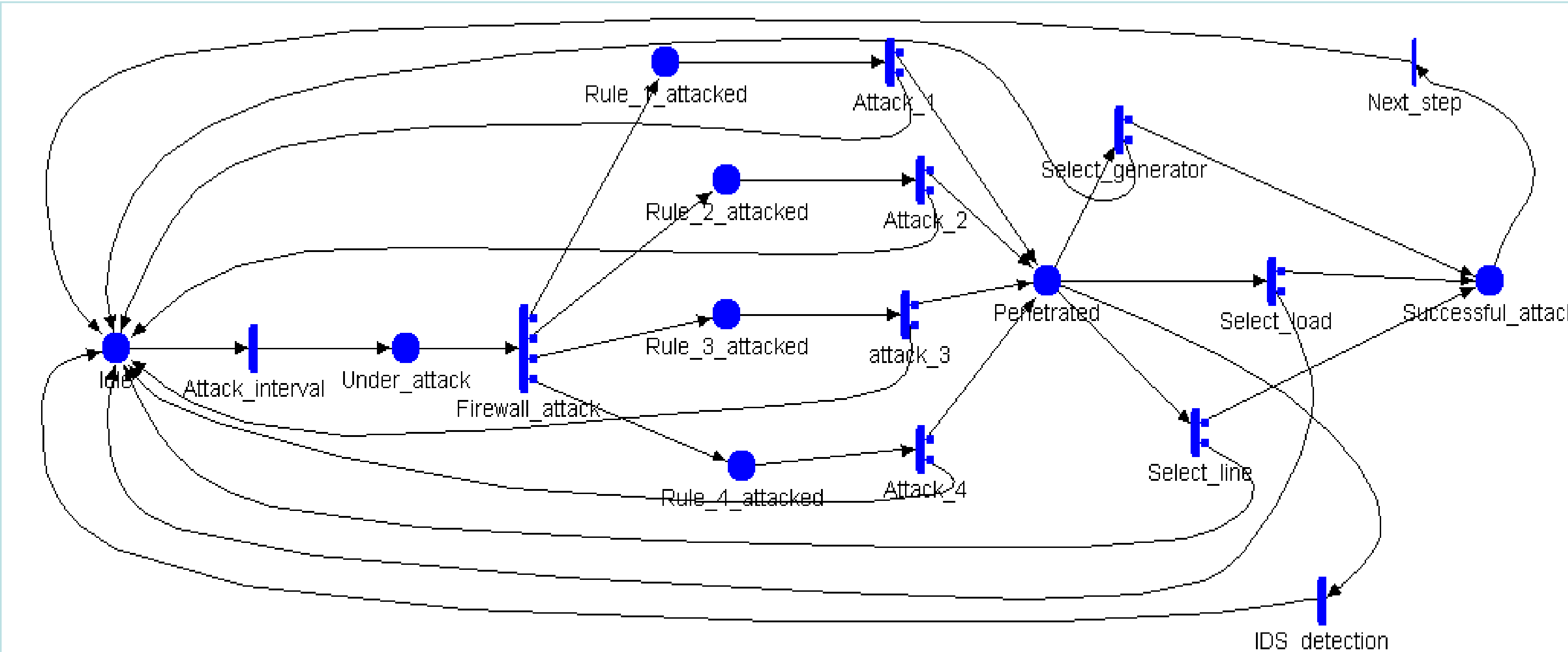


Risks in Industrial Control Systems

- Industrial Control Systems (ICS) demand ***different*** prioritisation of concerns (in comparison with enterprise systems):
  - Real-time - essential
  - High availability – paramount
  - Integrity - important
  - Privacy – ***typically*** not a concern
    - but seems important in power distribution systems
- Failures of Industrial systems have directly ***observable*** and measurable impact
  - In the enterprise systems the consequences of failures are less observable and the losses can easily be exaggerated
- Our work is on risk assessment when an ***objective*** utility/loss function can be defined



Model of Adversary



Models an attack on a firewall of a substation and the actions taken by an Adversary in case of a successful attack, which is ***switching off*** a ***single power element*** via its respective bay:

- a generator, or
- a load, or
- a line

Studies

A set of simulation experiments (studies) were completed to assess the risk of cyber attacks on the modelled power system

- We compared a ***base-line*** case with ***system under attack*** cases
  - Under the base-line case ***no attacks*** take place (the Adversary is ***inactive***)
  - Under the system under attack case the Adversary is active
- The model was ***parameterised*** as follows:
  - Transitions of the state machines representing the power and ICT elements were parameterised using ***data provided by experts***
  - For attacks we varied the rate of attacks (***sensitivity*** analysis):
    - once a year, once a month, once a week and once a day.
  - The ***chances of success*** by the adversary were also varied do that we can distinguish between poor and good ***security policies***
    - Repairs after successful attacks is achieved by either:
      - the ***standard control*** (for lines repair is almost instantaneously) or
      - dedicated measures additional: for generators and loads we modelled the repair time as an ***exponential distribution*** with an average of 3 hours (a typical figure for power systems).

The Adversary model

We varied the preferences of the Adversary

- A ***non-intelligent*** attacker - indifferent between targets (i.e. which sub-station to attack and which bay in a sub-station to switch off)
  - Different sub-stations are not equally important – some connect large generators/loads while some other – small generators/loads
- An ***intelligent*** attacker – greater generators and loads make a sub-station more attractive for the Adversary.
- For illustration of the difference we chose:
  - 5 largest generators are the only targets for the intelligent Adversary
  - 5 largest loads are the only targets for the intelligent Adversary which represent ***positive correlation*** between the importance index and the probability for a random target to be attacked by the Adversary.

The Intelligent Adversary Profile

Generators

Substation ID	Attack Probability | attack on a generator	Generator Capacity [MW]
4072	0.50	4500
4051	0.25	1400
4047	0.10	1200
4063	0.10	1200
4011	0.05	1000

Loads

Substation ID	Attack Probability | attack on a load	Load [MW]
4072	0.50	2000
4043	0.25	900
4051	0.10	800
1044	0.10	800
4046	0.05	700

Measures of interest (rewards)

The measures used in the studies are related to the supplied power.

The studies span over a period of 10 years (an arbitrary choice).

- some power is lost due to accidental failures
- power may also be lost due to successful attacks

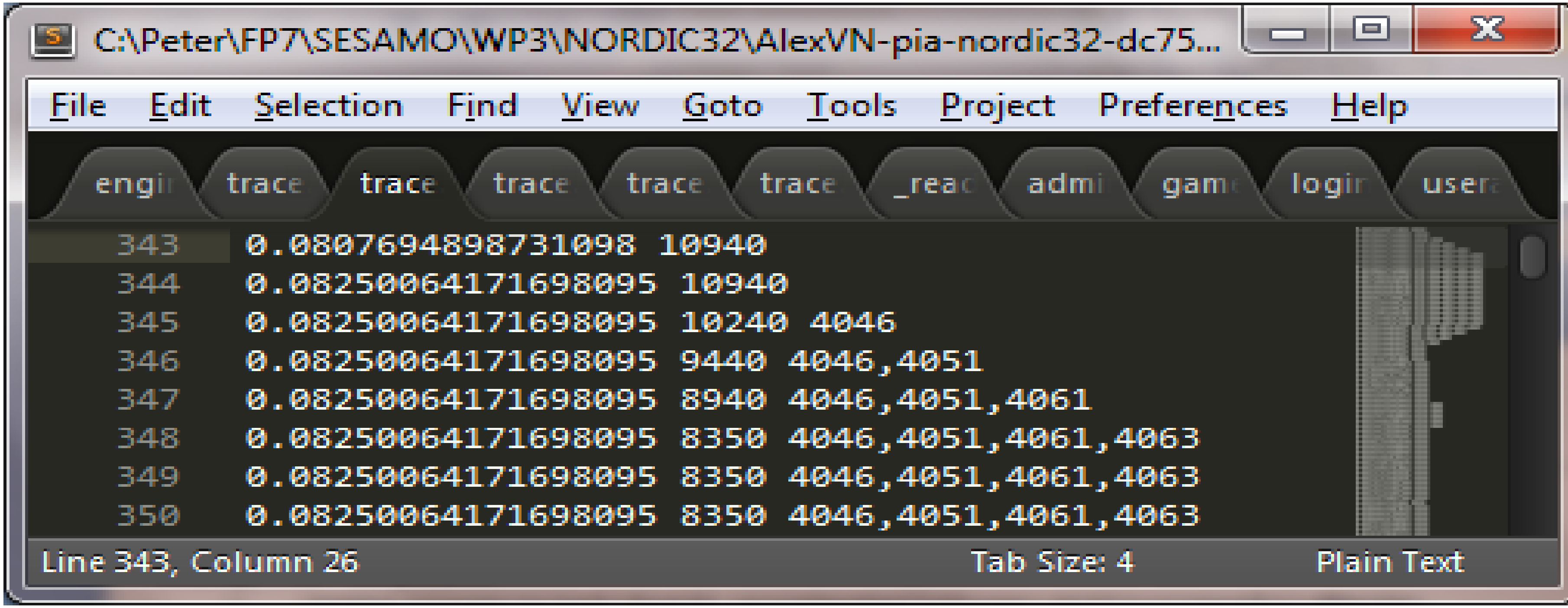
The chosen measures of interest (rewards) were computed for:

- the ***base-line*** case and
- the ***system under attack*** cases

CITY UNIVERSITY LONDON

Measures 1: Supplied Power

The supplied power, $P_i(t)$, is a ***random variable***.



We looked at two statistics:

- The average supplied power over the chosen interval of 10 years, $E[P_i(t)]$
- The standard deviation, $StD(P_i(t))$ is a measure of spread of the power delivered to consumers. Greater value indicate ***greater variability*** of power supply, i.e. more ***unstable*** power supply.

Measure 2: Probability of large outage

For each run we define a score function (an indicator) for each of the simulation runs as follows:

$$\omega_i(X) = \begin{cases} 1, if\ P_i(\ ) \leq X\ \ for\ 0 \leq t \leq 10\, years \\ 0, elsewhere \end{cases}$$

Then for a number of runs, N_r, we express the probability of large outage as:

$$P(X) = \frac{\sum_{1}^{N_r} \omega_i(X)}{N_r}$$

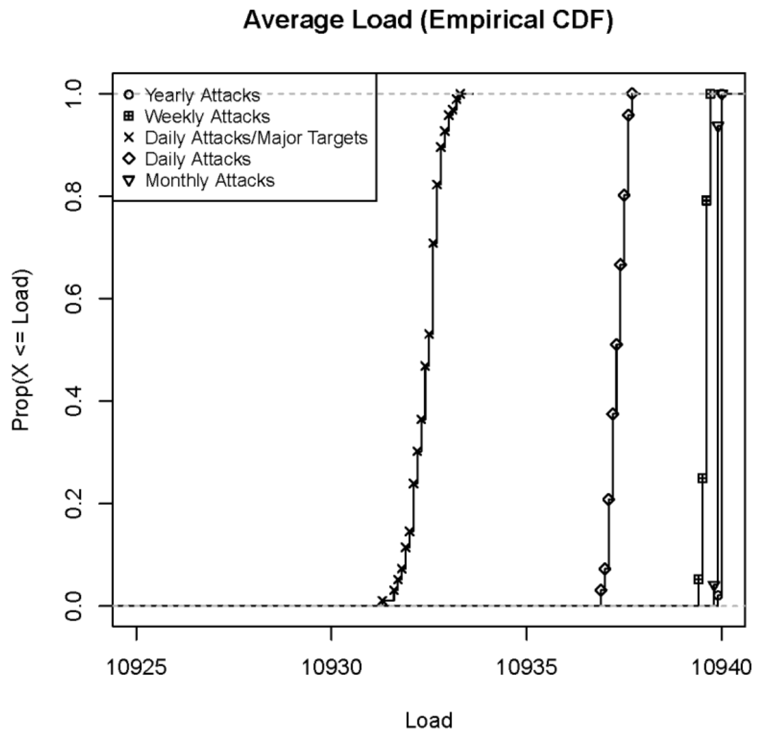
We set X as percentage of the nominal power, 10,940 MW, and compute P(X) for X = 10, 20, 30, ... 80, 90.

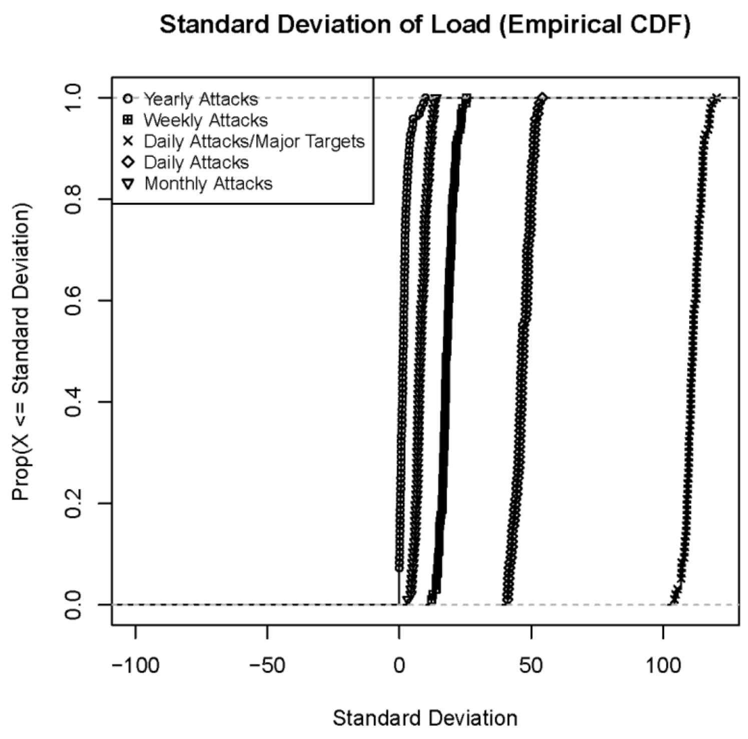
Results

- ~500 simulation runs of 10 years of operation
  - The number of events per run is in the range of 8000 – 32,000 including the attacks.
- Measure 1:
  - Over the population of 500 runs $E[P_i(t)]$ and $StD(P_i (t))$ are themselves random variable. We plot:
    - The distribution of $E[P_i(t)]$
    - The distribution of the standard deviation, $StD(P_i (t))$
- Measure 2:
  - Over the population of 500 runs we computed the probability that in a ***randomly chosen run*** the supplied power, $P_i(t)$, drops at least once to less X% of the nominal power, 10,940 MW.
  - This probability tells us the likelihood of a “large outage” to occur in the modelled system.

Measure 1: Attacks only case

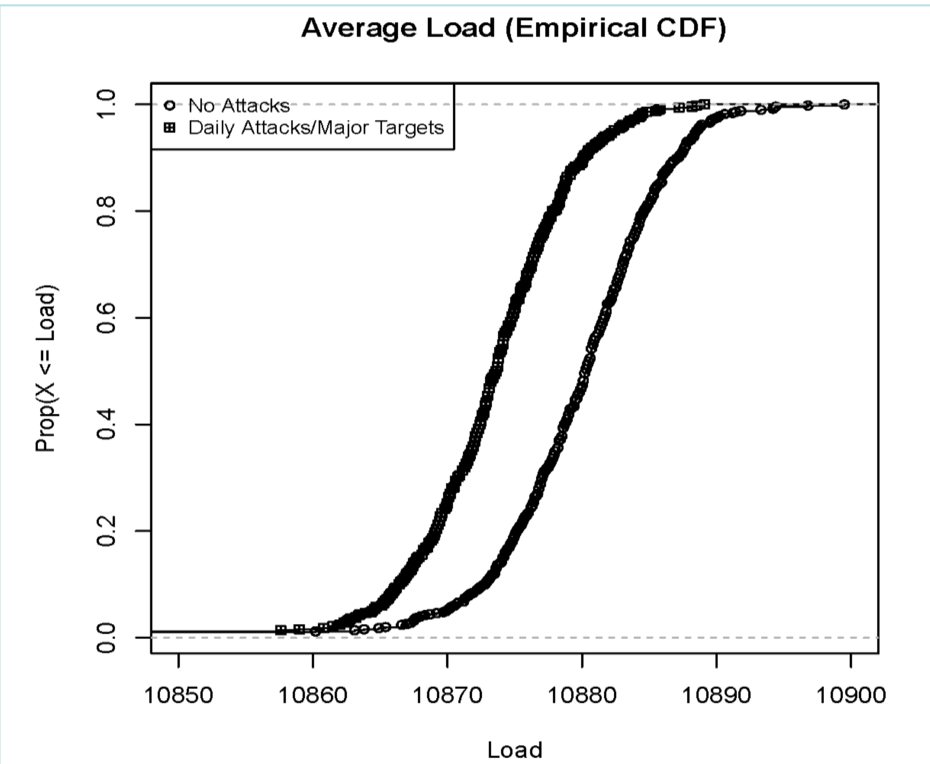
- The effect of frequency of the attacks on the power supply is shown below.
  - Power loss increases with the frequency of the attacks
  - Standard deviation increases, too.

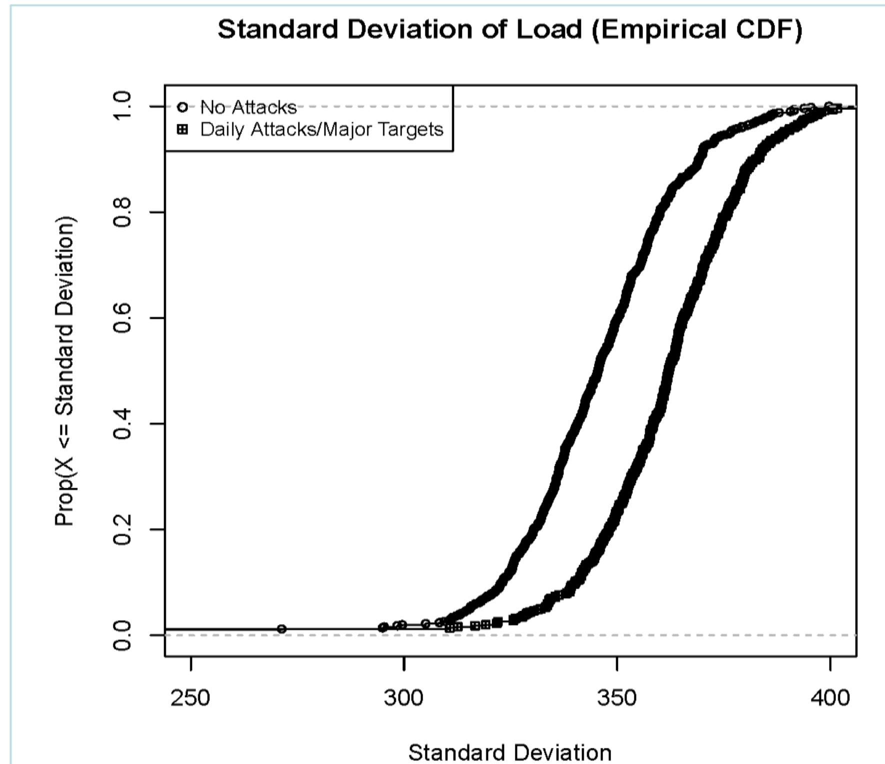




Measure 1: Failures and attacks

- The combined effect of accidental failures and the frequency of attacks on the power supply is shown below.
  - Power loss increases
  - Standard deviation increases, too





Measure 2: Probability of large outages

Probability that the ***power generation drops to X% of the nominal level*** of 10,940 MW ***at least once*** in 10 years of operation.

X[%]	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
no-attacks	0	0	0	0	0	0.466	0.99	1	1	1
daily-attacks.major (AF)	0	0	0	0.05	0.15	0.992	1	1	1	1
daily-attacks.major (NAF)	0	0	0	0	0	0	0.002	0.894	1	1
monthly-attacks (NAF)	0	0	0	0	0	0	0	0	0.808	1
weekly-attacks (NAF)	0	0	0	0	0	0	0	0.004	0.998	1
yearly-attacks (NAF)	0	0	0	0	0	0	0	0	0.114	1

major - attacks on one of the 5 larger generators or one of the larger loads.

AF - accidental failures

NAF - no accidental failure

Future work

- Extending the model of Adversary
  - More sophisticated scenarios are an obvious direction
    - attacking ***multiple*** targets by a single Adversary,
    - attacks that create ***hazards***, e.g. altering the threshold of a protection device, which will not manifest itself immediately, but may cause large outage later
  - A combination of cyber and physical attacks
  - Orchestrated (SWARM) attacks
- Looking into using simulation to help with quantification in applying fashionable theories in cyber security research
  - e.g. Nash equilibrium
- Given the great difficulty to parameterise Adversary models, ***sensitivity analysis*** for a plausible range of model parameters might be useful. This possibility was already demonstrated with the frequency of the attacks.
- The effectiveness of ***defences against cyber attacks*** in ICS can be studied, in case these can be varied and a decision is need which combination to apply. Among these defences are:
  - Frequency of repair
  - Use of sophisticated designs (e.g. using design diversity).

Conclusions

- We have built capability of quantifying the risk in complex ICS.
  - The methodology for interdependency analysis was adapted and tried on a non-trivial power system.
  - The impact of cyber security on industrial systems requires detailed hybrid models. In our view the system model must include:
    - a model of the Adversary,
    - a model of the ICS (e.g. Protection, control, etc.) and
    - a model of the controlled system itself (to evaluate more realistically the impact).
  - Tool support was developed (continuous improvements are under way)
- Initial observations:
  - Some initial indications suggest that not only naive attacks, but also attacks by an ***intelligent Adversary*** may have a ***limited impact*** on the ICS.
  - Measures of interest are important – risk perception varies with stakeholders.
    - “Black swan” events deserve particular attention
- ***Open issues*** related to methodology
  - how to do complex systems research
  - Issues of research methodology, testbeds, scaling, realism, realistic examples.
    - lack of general theories.

Questions

Thank you!