Estimating Sustainability Impact, Total Cost of Ownership and Dependability Metrics on Data Center Infrastructures

G. Callou*, P. Maciel*, F. Magnani†, J. Figueiredo*, E. Sousa*, E. Tavares*, B. Silva*, F. Neves† and C. Araujo*
Federal University of Pernambuco, Brazil
Email*: {grac, prmm, jjcf, etgs, eagt, bs, cjma}@cin.ufpe.br
Email†: {magnani, fneves}@ufpe.br

***Abstract*—The advent of cloud computing has demanded more computational resources from data centers. Prominent issues are dependability, total cost of ownership and sustainability, which are significantly impacted by the redundant infrastructures needed to support cloud computing services. In this context, models are important to support data center designers to estimate previous issues before implementing the final infrastructure. This paper presents a set of formal models for estimating sustainability impact and dependability metrics in data center infrastructures with the support of an integrated environment, namely, ASTRO. Besides, we adopt reliability importance to identify the most critical components in system reliability in order to propose different data center power infrastructures.**

I. INTRODUCTION

Data center availability has accomplished greater concerns due to increased dependence on Internet services (e.g., Cloud computing paradigm). For companies that heavily depend on the Internet for their operations, service outages can be very expensive, easily running into millions of dollars per hour [1]. A widely used design principle in fault-tolerance is to introduce redundancy to enhance availability. However, since redundancy leads to additional use of resources and energy, it is expected to have a negative impact on sustainability and the associated cost.

Efforts have been conducted to address sustainability in data centers, in which energy consumption management techniques and the adoption of clean energy sources are fundamental for reducing the energy consumption and environmental impacts. In this context, sustainable data centers are those that are built using the least amount of the appropriate materials and consume the least amount of appropriate sources of energy throughout their lifetime [2].

At present stage, data center designers do not have many mechanisms to support sustainability and dependability evaluation. For instance, two different data center architectures with similar availability may have very different sustainability impact. Additionally, a growing concern of data center designers is related to the identification of components that may cause system failure as well as systems parts that must be improved before implementing the architecture.

In this work, we propose a set of formal models for quantifying sustainability impact, total cost of ownership (TCO) and dependability metrics for data center power infrastructures. The adopted approach takes into account a hybrid modeling technique that considers the advantages of both stochastic Petri nets (SPN) [3] and reliability block diagrams (RBD) [4] to evaluate system dependability. In addition, reliability importance index (RI) [4] is adopted to propose different data center architectures with higher availability. RI is a prominent index to evaluate the necessity as well as the feasibility of redundancy in a system with less effort. An integrated environment, namely, ASTRO has been developed as one of the results of this work to automate dependability, sustainability and TCO evaluation.

This paper is organized as follows. Section II presents related works. Section III introduces basic concepts on data center infrastructures, dependability, sustainability and reliability importance. Section IV describes the adopted methodology. Section V presents the sustainability, dependability and energetic models. Section VI presents a real-world case study and Section VII concludes the paper as well as presents future works.

II. RELATED WORKS

In the last few years, some works have been developed to evaluate the data center infrastructures and a subset has also considered the impact on sustainability. [5] describes opportunities for energy integration in the context of combined cooling, heating and power systems. [6] proposes a risk anatomy to detect single point of failure in data center power distribution systems. In [7], the authors argue that other issues, in addition to operational energy consumption, should be considered to quantify data center environmental impacts. The entire lifecycle of the system must be examined, beyond operational energy to also include material use and manufacturing.

In [8], the authors present an approach to estimate the sustainability impact as well as the availability of data center architectures. The adopted approach does not consider any process to indicate equipments that most impact the system reliability and availability. In [9], the authors proposes an approach that quantifies the most cost effective and sustainable data center system. However, the approach does not consider any comparison between sustainability impact and availability of data center architectures.

III. Preliminaries

A. Data Center Infrastructure

This work considers a generic data center system, which essentially consists of the following subsystems, in addition to the building facility: (i) IT infrastructure; (ii) cooling infrastructure; and (iii) power infrastructure.

The IT infrastructure consists of three main components - servers, networking and storage devices. Cooling infrastructures are basically comprised of computer room air conditioning (CRAC) units, chillers and cooling towers. The cooling infrastructure may account around 40% of the total power consumption of the data center [10].

The power infrastructure is responsible for providing uninterrupted, conditioned power at the correct voltage and frequency to the IT equipment hosted in data center racks. From the electric utility, the power typically, goes through step down transformers, transfer switches, Uninterruptible Power Supplies (UPS), Power Distribution Units (PDU), and finally to rack power strips.

B. Dependability

The dependability [4] of a system can be understood as the ability to deliver a set of services that can be justifiably trusted. Indeed, dependability is related to disciplines such as fault tolerance and reliability. Reliability is the probability that the system will deliver a set of services for a given period of time, whereas a system is fault tolerant when it does not fail even when there are faulty components. Availability is also another important concept, which quantifies the mixed effect of both failure and repair process in a system. If one is interested in calculating the availability (A) of a given device or system, he/she might need either the uptime and downtime or the time to failure (TTF) and time to repair (TTR). Considering that the uptime and downtime are not available, the later option is the mean. If the evaluator needs only the mean value, the metrics commonly adopted are Mean Time to Failure ($MTTF$) and Mean Time To Repair ($MTTR$). For more details, the reader should refer to [4], which also provides the equations for estimating dependability metrics.

Dependability metrics might be calculated either by using RBD or SPN (to mention only the models adopted in this work). RBDs allow to one represent component networks and provide closed form equations. Nevertheless, when faced with representing dependent activities, some drawbacks may limit the model. On the other hand, state-based methods can easily consider those dependencies. However, they suffer from the state-space explosion. Some of those formalism allow both numerical analysis and stochastic simulation, and SPN is one of the most prominent models of such class.

C. Measure of Component Importance

Component importance is a metric that indicates the impact of a particular component in the system's overall reliability. Various measures are available for estimating component importance, which often relates the contribution of a component to the system's failure. This paper adopts Reliability Importance to identify the most critical components for system reliability.

1) Reliability Importance: (or Birnbaum Importance) of a component i corresponds to the amount of improvement in system reliability, when the reliability of such a component is increased by one unit [4]. In other words, RI is a partial derivative of system reliability with respect to each individual component failure rate. The RI of component i can also be estimated as:

$$I_i^B = R_s(1_i, \mathbf{p}^i) - R_s(0_i, \mathbf{p}^i) \quad (1)$$

in which I_i^B is the reliability importance of component i; $\mathbf{p}^i$ represents the component reliability vector with the ith component removed; 0_i represents the condition when component i is failed; and 1_i describe the condition when i is working.

I_i^B depends on the structure of the system and the reliability of other components. In other words, the RI of a component i is completely determined by the reliabilities of the other components, excluding i [4].

D. Sustainability

Life-cycle assessment (LCA) [11] is a common approach to quantify environmental sustainability. In this work, a LCA-based approach is adopted to estimate the sustainability impact of an equipment in terms of its lifetime exergy (available energy) consumption. It should be noted that while the proposed approach is applied to analysis of data center power infrastructures, the model is sufficiently general to be applied to any component of a data center infrastructure.

The approach essentially divides the life-cycle into two phases: (1) embedded phase, which involves all impacts related to product design decisions (including material extraction, manufacturing and supply chain impacts, as well as end-of-life); and, (2) operational phase, which involves all impacts related to decisions during product use (such as operational and maintenance cycles).

IV. Methodology

The methodology's first step (Figure 1) concerns understanding the system, its components, their interfaces and interactions. This phase should also provide the set of metrics (e.g. availability, reliability importance, sustainability indicators, economic losses). The next broad phase aims at grouping components in order to generate subsystems, which are adopted to mitigate the complexity of the final system model evaluation. For each subsystem, the dependability metrics are computed. These metrics might be calculated either by using RBD or SPN (to mention only the models adopted in this work). After obtaining the subsystem models, the next step will be

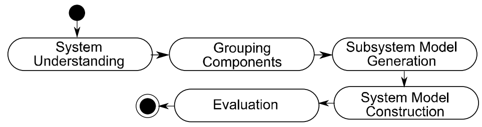

Fig. 1. Methodology

composing the subsystem models. Lastly, the system model is evaluated to obtain the metrics of interests.

V. Models

This section presents the models adopted for quantifying system dependability and sustainability.

A. Energetic Model

The system under evaluation can be correctly arranged, but they may not be able to meet system demand for electrical energy or thermal load. For instance, consider the cooling infrastructure depicted in Figure 2 (a). Assume that the amount of heat that should be extract from the data center room corresponds to 10kJ (Figure 2 (b)), and the maximum cooling capacity of the adopted CRACs and Chiller are 8kJ and 18kJ, respectively. Figure 2 (c) depicts a possible energetic flow, which extracts heat using both CRACs (with a load of 5 kJ in each component) that transfer the heat to the Chiller (10 kJ). Considering another example, in which instead of adopting two CRACs, only one is considered in the infrastructure. As shown in Figure 2 (d), the system will be able to extract 8kJ of heat. Thus, the system will not be able to satisfy the required thermal load.

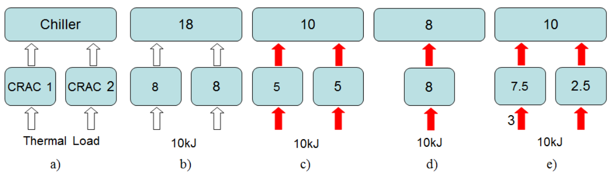

Fig. 2. a) System example; b) Maximum Cooling Capacity; c) Successful Energetic Flow; d) Failed Energetic Flow; e) Representation with Weight.

To perform such an evaluation, an infrastructure model is converted into a graph, in which the components are represented by nodes and the respective connections are modeled by edges. The edges have weights that are adopted to compute the amount of energetic flow directed to a component. For instance, consider Figure 2 (e) in which a weight of 3 unit is defined for one edge and 1 unit for the other edge. In this case, CRAC1 component extracts 3 times more heat than CRAC2. Next, an algorithm traverses the graph, checking the component capacities to meet the load for electrical energy or heat extraction.

Let $G = (N, V, w, f_p, s, t)$ be a graph which represents the energetic model in which:

- N represents the set of nodes (i.e., the components);
- $V \subseteq (N \times N)$ denotes the set of edges (i.e., the component connections);
- $w : V \rightarrow \mathbf{R}$ the weights of the edges (i.e., the value adopted for computing the amount of energetic flow directed to a component);
- $f_p : N \rightarrow \mathbf{R}^+$ represents the function that relates each node with its corresponding maximum power;
- $s \in N$ is the source or root node (i.e., an abstract component that represents the initial node of the infrastructure);

```
1  verify(G, D_a, a) {
2    O = {n | (a,n) ∈ V};
3    if(O = ∅) return TRUE;
4    ws = Σ_{n∈O} w(a,n);
5    ∀n ∈ O, if (f_p(n) < (D_a × w(a,n)/ws + n.c) {
6      return FALSE;
7    } else {
8      n.c = n.c + D_a × w(a,n)/ws;
9    }
10    ∀n ∈ O, return verify(G, out(n, D_a × w(a,n)/ws));
11 }
```

Fig. 3. Energetic Algorithm

- $t \in N$ is the target node (i.e., an abstract component that represents the final node of the infrastructure);

Besides, G has also the following properties: (i) acyclic, assuming no cycles and (ii) directed, assuming directed edges which represents the energetic flow. Each node n has also an attribute c ($n.c$), which represents the current capacity utilized by the component. Initially, the current capacities of all nodes are set to 0. The algorithm is depicted in Figure 3, which adopts a depth-first approach to traverse the graph.

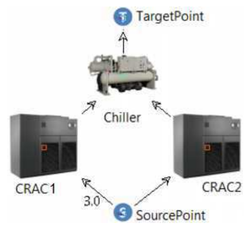

Fig. 4. An energetic model example.

As an example, Figure 4 depicts the ASTRO's graph representation for the energetic model of the Figure 2 (e). Initially, the source node (a=s) is the first node to be visited in G. The respective set of output nodes (O) are obtained (line 2) and verified if it is empty. This verification is not TRUE (line 3), since the node a has two output edges. Next, the sum of its output edge weights (ws) is calculated to estimate the amount of energy that can be transferred to (in the case of electricity) or extracted from a (when concerning heat).

For each $n \in O$ (line 5), the algorithm verifies if the output node still has any capacity to meet the required demand. Function $f_p : N \rightarrow \mathbf{R}^+$ represents the maximum cooling or power capacity of a component (which are informed by the designer); D_a represents the required (electrical or thermal) energy that must be transferred to (e.g, electricity) or extract (e.g., heat) from the node a; $w(a,n)/ws$ is the ratio of energy that needs to be transferred/extracted. Thus, $(f_p(n) < Da \times (w(a,n)/ws) + n.c)$ evaluates the capacity of n to support an additional demand from a node a. If the evaluated value is greater than the maximum capacity, the algorithm terminates. Otherwise, node n is updated (line 8).

Afterwards, the verification is recursively performed for each output node n in order to search for any inconsistence in the energetic model. The demanded energy for a node n is computed by function out, which is based on Table I. In that table, η represents the equipment's efficiency according to the 2nd law of thermodynamic; Q_{IN} is the input thermal energy; Q_F is the thermal energy CRAC's fluid; COP is the coefficient of performance; Q_q corresponds to the thermal load that flows to the cooling tower; $\mu = \frac{MaximumCoolingPower}{MaximumPowerConsumption}$. Since the physical behavior of each device is not the focus of this work, the reader should refer to [12][13].

TABLE I
EQUATIONS FOR ENERGY OUTPUT OF DIFFERENT DEVICES.

Device	Equation
Electrical	$\frac{E_c}{\eta}$
CRAC	$\frac{1}{\eta} \times Q_{IN}$
Chiller	$Q_F \times \left(1 + \frac{1}{COP}\right)$
Cooling Tower	$Q_q \times \left(1 + \frac{1}{\mu}\right)$

B. Sustainability Model

In this work, environmental impact is computed in terms of the thermodynamic metric of exergy (also called usable available energy).

1) Embedded Exergy: The embedded exergy, which involves impacts related to product design decisions, is obtained as follows:

$$Ex_{emb} = E_{man} \times [\eta_{man} + (1 - \eta_{man}) \times (1 - C_{reuse})] \quad (2)$$

where E_{man} is the energy required for manufacturing all equipments adopted in the infrastructure (see Equation 3); η_{man} is the manufacturing efficiency (2nd law of thermodynamics); and C_{reuse} is the exergy reused in other processes.

$$E_{man} = \sum_{i=1}^{n} E_{eq_i} \quad (3)$$

In the particular case of electrical energy, all energy can be theoretically converted into work, thus, the variable E_{man} in Equation 3 means the total exergy made accessible during the manufacturing phase. A fraction of E_{man} is consumed. The complementary fraction (1-E_{man}) can be reused in other processes (C_{reuse}) or destroyed (1-C_{reuse}).

In this work, the energy required for manufacturing each equipment (E_{eq_i}) has been obtained from [14].

2) Operational Exergy: The operational exergy consumption can be understood as the fraction of the heat dissipated by each equipment that cannot be theoretically converted into useful work. The following equation represents the operational exergy.

$$Ex_{op} = \sum_{i=1}^{n} Ex_{op_i} \times T_{life} \times A \quad (4)$$

where Ex_{op_i} is the operational exergy of each device (Table II); T_{life} is the lifetime (i.e., the period of analysis); and A is the system availability.

As each device may consume or destroy exergy in different ways, specific equations (Table II) are adopted for each component. In this table, η is the delivery efficiency; P_{in} is the total input power of the electrical device; φ is the exergy correction; Q_{in} is the total thermal input of the device; T_{room} is the data center room temperature; T_{cold} corresponds to the CRAC's cold water temperature; μ is the ratio of the maximum cooling power by the maximum power consumption; COP is the coefficient of performance; T_{tower} is the water temperature that goes to the cooling tower; $T_{chilled}$ corresponds to the chilled water temperature; T_{amb} is the ambient temperature; T_{warm} corresponds to the warm water from chillers. Since the physical behavior of each device is not the focus of this work, the reader should refer to [12][13].

TABLE II
OPERATIONAL EXERGY EQUATIONS OF DIFFERENT DEVICES.

Device	Operational Exergy Equation
Electrical	$P_{in} \times (1 - \eta)$
Diesel Generator	$P_{in} \times \left(\frac{\varphi}{\eta} - 1\right)$
CRAC	$Q_{in} \times \left(1 - \frac{T_{cold}}{T_{room}} + \frac{1}{\mu}\right)$
Chiller	$Q_{in} \times \left(\frac{1}{COP} - \frac{T_{tower} - T_{chilled}}{T_{chilled}}\right)$
Cooling Tower	$Q_{in} \times \left(1 - \frac{T_{amb}}{T_{warm}} + \frac{1}{\mu}\right)$

3) Lifetime Exergy: In this work, the sum of the embedded and operational exergies is the lifetime exergy (LTE).

C. Dependability Models

This section presents the models adopted for quantifying system dependability and sustainability.

1) RBD Models: Reliability Block Diagram (RBD) is a combinatorial model that was initially proposed as a technique for calculating reliability on large and complex systems using intuitive block diagrams. Such a technique has also been extended to calculate other dependability metrics, such as availability and maintainability. The blocks (e.g., components) are usually arranged using the following composition mechanisms: series, parallel, bridge, k-out-of-n blocks, or, even, a combination of previous approaches. The reader should refer to [4] for examples as well as the closed-form equations.

2) SPN Models: This work adopts a particular Petri net extension, namely, Stochastic Petri Nets (SPN) [3], which allows the association of probabilistic delays to transitions using the exponential distribution, and the respective state space is isomorphic to continuous time Markov chains (CTMC) [3]. Besides, SPN allows the adoption of simulation techniques for obtaining dependability metrics, as an alternative to the Markov chain generation. Next section briefly present the proposed SPN building block for obtaining dependability metrics.

Simple Component. The simple component has two states: functioning or failed. To compute its availability, TTF and TTR should be represented. Figure 5 shows the SPN model of the "simple component", which has two parameters (not depicted in the figure), namely X_MTTF and X_MTTR, representing the delays associated to the transitions $X\_Failure$ and X_Repair, respectively.

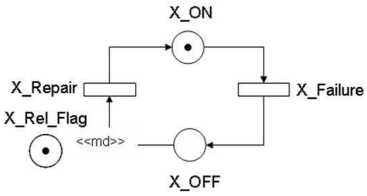

Fig. 5. Simple component model

Places X_ON and X_OFF are the model component's activity and inactivity states, respectively. The simple component also includes an arc from X_OFF to X_Repair with multiplicity depending on place marking. The multiplicity is defined through the expression IF($\#X\_Rel\_Flag = 1$) 2 ELSE 1, where place X_Rel_Flag models the evaluation of reliability/availability. Hence, if condition $\#X\_Rel\_Flag = 1$ is false (no token in place X_Rel_Flag), then the evaluation refers to availability. Otherwise, the evaluation concerns reliability.

A component is operational if the number of tokens ($\#$) in place X_ON is greater than 0 and in a failure state, otherwise. Hence, if $\#X\_Rel\_Flag = 1$, $P\{\#X\_ON > 0\}$ means the component's availability (steady-state evaluation). If $\#X\_Rel\_Flag = 0$, then $P\{\#X\_OFF > 0\}$ allows computing the component's reliability, if transient evaluation is carried out and the initial marking is $\#X\_ON = 1$ and $\#X\_OFF = 0$.

Without loss of generality, the subsystems are combined by serial, parallel, (non) series-parallel, and hierarchical compositions.

VI. Case Study

This section presents a case study to illustrate the application of those models (Section V) and ASTRO environment to evaluate data center power architectures. Seven data center power infrastructures have been considered (see Figure 6). For each architecture, we estimate: (i) availability; (ii) the sustainability impact and (iii) the total cost ownership.

A. Architectures

From the power infrastructure A1 depicted in Figure 6 (a), we propose other architectures that are created according to the component reliability importance index. Originating in the utility feed (i.e., AC Source), typically, the power goes through voltage panel, uninterruptible power supply (UPS) unit, power distribution units (PDUs) (composed of a transformer and an electrical subpanel), junction boxes, and, finally, to rack PDUs (rack power distribution units).

In this work, the system is operational if the power infrastructure is able to supply energy to the IT devices. For the sake of readability, each IT device block represents 10 racks and, so, the adopted infrastructure provides electrical power to 50 IT racks. Additionally, MTTFs and MTTRs values for the components have been obtained from [15].

We started performing the dependability evaluation of the baseline architecture (A1). As will be further explained, six new architectures are generated as shown in Figure 6. Moreover, for the sake of readability, Table III presents the importance index for the components in each architecture. In that table, it is possible to note that the AC Source on architecture A1 has the highest RI value. Thus, architecture A2 (see Figure 6 (b)) is created adopting the redundancy of that equipment. A static transfer switch (STS) is also added for switching from a failed AC Source to the operational one.

TABLE III
Reliability Importance Values of Architectures A1-A7.

	ACS	STS	UPS	SDT	JB	LVP	SP	ITD
A1	1		0.140	0.137	0.136	0.136	0.136	3.3E-08
A2	1	0.303	0.302	0.295	0.294	0.293	0.293	7.2E-08
A3	1	0.303	0.010	0.295	0.294	0.293	0.293	7.2E-08
A4	1	0.303	0.013	0.013	0.294	0.293	0.293	7.2E-08
A5	1	0.303	0.013	0.013	0.002	0.293	0.293	7.2E-08
A6	1	0.301	0.140	0.137	0.002	0.136	0.292	7.2E-08
A7	1		0.140	0.137	0.136	0.136	0.1361	7.2E-08

Key: ACS-AC Source; LVP-Low Voltage Panel; STS-Static Transfer Switch; UPS-Uninterruptible Power Supply; SDT-Step Down Transformer; SP-Sub Panel; JB-Junction Box; ITD-Information Technology Device;

Considering the reliability importance values of architecture A2, although there are redundant AC Sources, these components still have the highest RI indexes (see Table III). The cost to add another ACS is very high. Besides that, the STS (second highest value in Table III) is a component that cannot be separately replicated. Therefore, architecture A3 (see Figure 6 (c)) corresponds to the architecture A2 with redundant UPS components (i.e., the next item in the list with the highest value).

Considering a similar approach, architecture A4 (see Figure 6 (d)) is created from architecture A3, adopting redundant transformers. Architecture A5 (Figure 6 (e)) is defined applying redundancy to the Junction Box of A4. Architecture A6 (see Figure 6 (f)) is conceived with redundancy applied to the low voltage panel. In such an architecture, it is possible to eliminate a STS component that switch the AC Sources. Finally, architecture A7 (see Figure 6 (g)) corresponds to the architecture A1 with all components replicated.

B. Results

Table IV summarizes the results separately for each infrastructure, in which: *Architec* represents the architecture evaluated; *Avail(%) (9's)* is the availability level (A) with the respective number of nines (-*log*[1 - A/100]); *EMB(GJ)* is the embedded exergy in gigajoule; *OPER(GJ)* is the operational exergy in gigajoule; *LTE(GJ)* is the lifetime exergy consumption (*EMB(GJ)* + *OPER(GJ)*); and *TCO(USD)* is the total cost of ownership in U.S. dollars. For this case study, TCO takes into account the cost to build and to maintain the system in operational state over a period of 1 year.

As expected, the availability increases when redundancy improves in the architectures. However, as previously presented in Table III, STS has a significant impact in the system availability. For instance, the availability result of A3 is slightly smaller than A2 due the fact that, besides considering an additional UPS, A3 considers an additional STS. Similarly, the expressive availability increase from A6 to A7 can be explained by the absence of STS devices in A7.

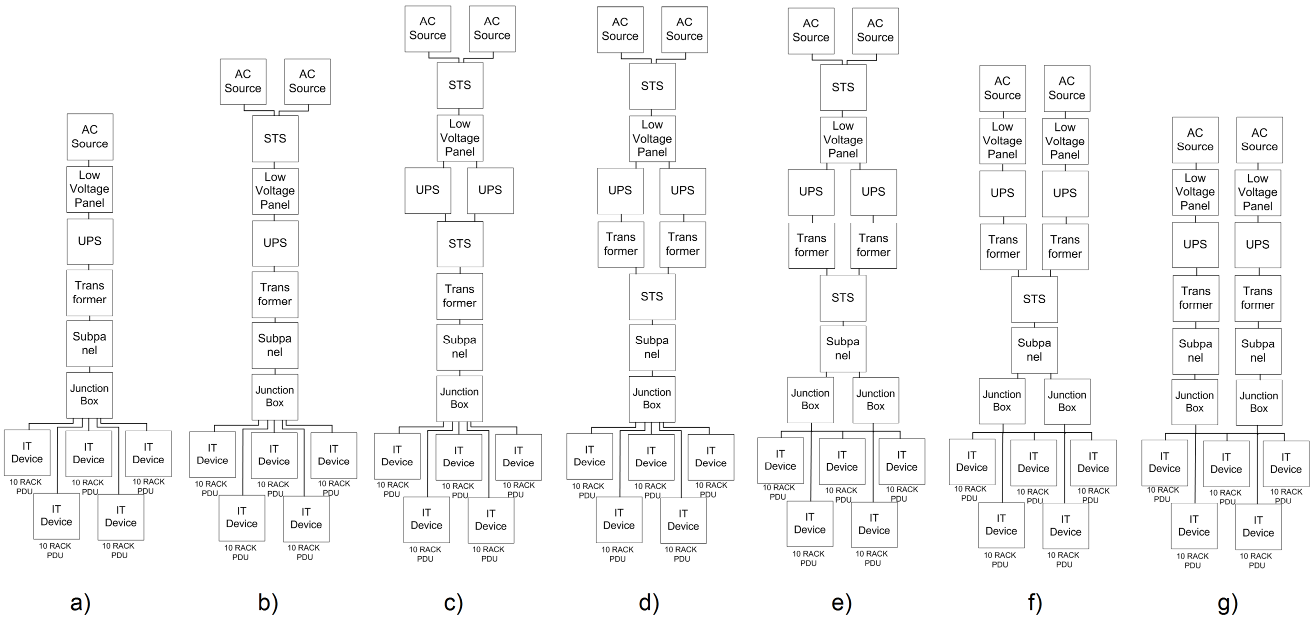

Fig. 6. a) Architecture A1; b) Architecture A2; c) Architecture A3; d) Architecture A4; e) Architecture A5; f) Architecture A6; g) Architecture A7.

TABLE IV
SUMMARY RESULTS.

Architec	Avail(%) (9's)	EMB(GJ)	OPER(GJ)	LTE(GJ)	TCO(USD)
A1	99.8117 (2.725)	198.63	1996.09	2194.72	647984.73
A2	99.9904 (4.017)	304.84	2088.94	2393.78	667482.52
A3	99.9902 (4.012)	363.59	2178.66	2542.25	741023.44
A4	99.9913 (4.061)	363.93	2178.68	2542.62	741579.18
A5	99.9919 (4.094)	364.18	2178.70	2542.88	741732.68
A6	99.9957 (4.376)	363.83	2089.05	2452.88	738411.95
A7	99.9996 (5.450)	363.48	1999.84	2363.33	735104.79

For a better visualization, Figure 7 depicts a graphical comparison between the availability, the total cost of ownership and the lifetime exergy for each power infrastructure architecture. The baseline corresponds to the architecture A1. As the reader should note, architecture A7 is an interesting option, since is less than 15% more expensive than the baseline and provides the highest availability (i.e., twice the value of A1). Besides, the lifetime exergy consumption is smaller than architectures A2 to A6.

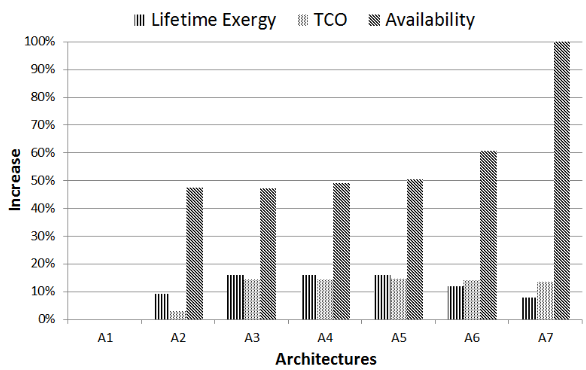

Fig. 7. Comparing Exergy, Availability and TCO.

VII. Conclusion

This work presented a set of formal models, supported by the developed environment ASTRO, to estimate the availability, TCO and sustainability impacts of data center infrastructures. The approach considers reliability importance index to propose architectures with higher availability. Moreover, the feasibility of those models and the environment are demonstrated considering the evaluation of a real world data center power infrastructure. The models as well as the environment are generic enough to allow designers to evaluate different systems. As a future work, we intend to analyze the cooling and IT infrastructures.

References

[1] D. Patterson. A simple way to estimate the cost of downtime, 2002.
[2] C Bash and et al. The sustainable information technology ecosystem. In *ITherm '08*, 2008.
[3] K Trivedi. *Probability and Statistics with Reliability, Queueing, and Computer Science Applications*. Wiley, 2002.
[4] Way Kuo and Ming J. Zuo. *Optimal Reliability Modeling - Principles and Applications*. Wiley, 2003.
[5] K.E Herold and R. Radermacher. Integrated power and cooling systems for data centers. In *ITHERM*, 2002.
[6] M. Wiboonrat. Risk anatomy of data center power distribution systems. In *ICSET'08*, 2008.
[7] J. Chang, J. Meza, P. Ranganathan, C. Bash, and A. Shah. Green server design: beyond operational energy to sustainability. In *HotPower*, 2010.
[8] M. Marwah, P. Maciel, A. Shah, R. Sharma, and T. Christian. Quantifying the sustainability impact of data center availability. *SIGMETRICS Perform. Eval. Rev.*, 37:64–68, March 2010.
[9] D. Gmach, Y. Chen, A. Shah, J. Rolia, and C. Bash. Profiling Sustainability of Data Centers. In *Sustainable Systems and Technology (ISSST), 2010 IEEE International Symposium on*, pages 1–6. IEEE.
[10] C. Belady. In the data center, power and cooling costs more than the it equipment it supports. *Electronics Cooling Magazine*, February 2007.
[11] H. Baumann and A. Tillman. The hitch hiker's guide to lca. In *Studentlitteratur AB*, 2004.
[12] Michael J. Moran and Howard N. Shapiro. *Fundamentals of Engineering Thermodynamics*. John Wiley Sons, Inc., 5 edition, 2006.
[13] Yunus A. Cengel and Michael A. Boles. *Thermodynamics an Engineering Approach*. 5 edition, 2003.
[14] Economic input-output life cycle assessment. http://www.eiolca.net/.
[15] *IEEE Gold Book 473, Design of Reliable Industrial and Commercial Power Systems*.