# Evaluating the Cooling Subsystem Availability on a Cloud Data Center

Demis Moacir Gomes*, Patricia Takako Endo†, Glauco Gonçalves‡, Daniel Rosendo*, Guto Leoni Santos*,
Judith Kelner*, Djamel Sadok* and Mozghan Mahloo§
*Federal University of Pernambuco, Recife, Brazil
E-mail: {demis.gomes, daniel.rosendo, guto.leoni, jk, jamel}@gprt.ufpe.br
†University of Pernambuco, Caruaru, Brazil
E-mail: patricia.endo@upe.br
‡Federal Rural University of Pernambuco, Recife, Brazil
E-mail: glauco.goncalves@ufrpe.br
§Ericsson Research, Stockholm, Sweden
E-mail: mozhgan.mahloo@ericsson.com

*Abstract*—**A data center is divided into three basic subsystems: information technology (IT), power, and cooling. Cooling plays an important role related to data center availability, and a failure in this subsystem may cause an interruption of services. Generally, a redundant cooling subsystem is implemented based on replace of failed component by a standby one. However, it also can be based on a rotation of computer room air conditioners (CRACs). This paper proposes scalable models that represent the cooling subsystem behavior to evaluate the impact of cooling failures on the data center availability. Models are based on the TIA-942 standard and represent Tiers I and II. We validate our model with other from the literature, and results show that the CRACs' rotation, although used in real data center scenarios, has similar results in availability when compared to the replace strategy.**

## I. INTRODUCTION

Unplanned data center failures are expensive and require special attention. According to the International Working Group on Cloud Computing Resiliency, a disruption of credit card authorization service incurs a loss of around $2,600,000 per hour[1], and cloud service providers, like Amazon and Microsoft Azure, have a cost of $336,000 per hour in a case of failure [1]. To avoid financial and reputational losses, a cloud provider must understand the weaknesses of their data center infrastructure, in order to guarantee highly available services.

Availability of a data center can be impacted by failures on its three main subsystems: information technology (IT), power, and cooling. Although the cooling subsystem costs are around 5-20% of a data center total expenditure [2], this subsystem receives less attention in the literature of data center availability when compared to IT and power subsystems. However, the cooling subsystem plays an important role in the data center availability; it is responsible for maintaining the IT subsystem at a proper temperature to avoid hardware and software failures.

Cooling subsystems are basically composed of cooling towers, chillers and computer room air conditioners (CRACs). They can be implemented in different ways [3], and have different redundancy models, operating times, and repair times. In general, data centers implement cooling subsystems using a replace strategy; a component in standby mode replaces another, in case of failure. Another strategy establishes a rotation among active components; for instance, a CRAC works during a regular period of time and it is then turned off, being replaced by another. Analyzing the cooling subsystem and how its components impact on the overall data center availability can offer good insight to improve the data center operation.

According to [4], data center infrastructure can be implemented considering one of four tiers (from I to IV) and can obtain different availability levels, according to the number of redundant components and redundant paths in the tier. Some works, such as [5] and [6], model a data center cooling architecture up to tier II, but such models are not scalable with respect to the number of components. Besides, these works do not model the rotation of CRACs. This work presents scalable models for tiers I and II using Petri Nets, measuring the impact with and without the CRAC rotation strategy. We validate the results of tier I against other models from literature and simulate the availability in Tier II.

This work is organized as follows. Section II describes important concepts to understand the models; related work is presented in Section III. Tier I and II models are depicted in Sections IV and V, respectively. The results of evaluation and validation are presented in Section VI. Section VII concludes the paper and discusses future works.

## II. BASIC CONCEPTS

### A. Petri Nets

The computational modelling of a system may be established by analytical models, non-state space models, or state space models [7]. Analytical models provide instantaneous results, but can ignore relevant factors due to assumptions. Non-state space models, such as Reliability Block Diagram (RBD), allow a high abstraction level, however, it is not able

---

[1] IWGCR: Downtime costs per hour. Available at http://iwgcr.org/?p=404. Last Accessed: 2017, February.

to represent the whole behavior of a system, such as its internal processes. On the other hand, state space models, such as Petri Nets and Markov Chains, allow the behavioral representation of a system with greater granularity. As Markov chains work only with markovian distributions, Petri Nets would be the best choose to represent a distributed and concurrent system with specific and dynamic behavior [7].

According to [7], a Petri Net $C$ can be defined as a four-tuple $C = (P, T, I, O)$, where $P = \{p_1, p_2, ..., p_n\}$ is a finite set of places, $n \geq 0$; $T = \{t_1, t_2, ..., t_m\}$ is a finite set of transitions, $m \geq 0$; $I : T \rightarrow P^{\infty}$ is the input function (that represents a mapping from transitions to a set of places); and $O : T \rightarrow P^{\infty}$ is the output function (that represents a mapping from transitions to a set of places).

Graphically, a Petri Net is composed of circles (white circles represent places, and black circles represent tokens), rectangles (transitions), and arcs (see Figure 1). Places describe passive components (Figure 1.a), while transitions are active ones. There are two types of transitions: timed (Figure 1.b) and immediate (Figure 1.c). The timed transition is activated through a time parameter that follows a given probability distribution, commonly exponential. Immediate transitions are activated instantly. A set of places and transitions representing a system component can be called a building block.

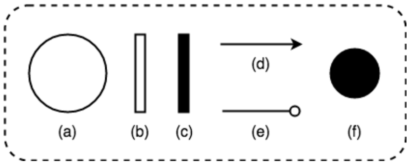


Fig. 1. Petri Net components

The connection between places and transitions is made through directed arcs (Figure 1.d). Transitions are only enabled to fire if all preconditions are fulfilled, i.e., if there are enough tokens (Figure 1.f) in the input places. When a transition fires, it consumes tokens from its input places and produces tokens at all of its output places. On the other hand, if the arc is an inhibitor one (Figure 1.e), the precondition on firing is that there is no token in the input place. To reduce the number of arcs, transitions, and places in a Petri Net, a transition can be activated with a guard function, a boolean expression composed of places, transitions, and tokens. For instance, consider the guard function $GF_{systemDown} = (\#machine_{on} = 0) AND (\#routers_{on} = 0)$, it means that the transition $GF_{systemDown}$ will be enabled only if there are no tokens in places $\#machine_{on}$ and $\#routers_{on}$.

Through a Petri Net, among other features, we can estimate system availability. The availability can be defined as system uptime over total system time, where total time is described as the sum of system uptime and system downtime. These concepts can be associated with the average behavior of the system for the purpose of availability calculation. In the following formula, the availability is calculated by division of the Mean Time To Failure (MTTF) by the Mean Time Between Failures (MTBF). The MTBF also is defined as the sum of MTTF and MTTR (Mean Time to Repair), indicating the time between the detection of a failure and the detection of next failure, as showed in Equation 1.

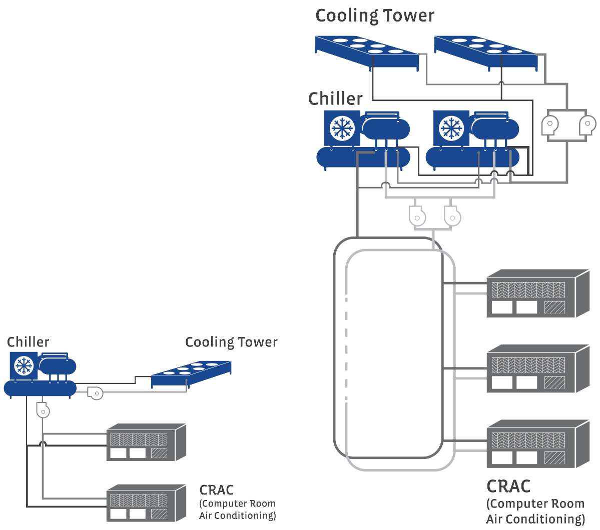
$$availability = \frac{MTTF}{MTBF} = \frac{MTTF}{MTTF + MTTR} \qquad (1)$$

### B. Cooling Subsystem Components

The TIA-942 from Telecommunications Industry Association (TIA) [4] and the ANSI/BICSI-002 (American) from Building Industry Consulting Service International (BICSI) [8] are the main data center availability standards. They define fundamental aspects, best practices, and recommendations regarding data center design and infrastructure. TIA-942 refers to four tiers from I to IV, while BICSI-002 defines five classes, from 0 to 4. Although there are differences, the two first classes of the BICSI-002 standard (0 and 1) are compatible with Tier I of the TIA-942 standard, achieving the same level of availability. Thus, we follow the architecture divided into four tiers, in which the availability ranges from 99.6% on Tier I to 99.95% on Tier IV. According to these standards, a generic data center system is basically composed of three subsystems: power infrastructure, cooling infrastructure, and IT infrastructure [9].

According to [10], the sum of cooling and power subsystems' costs can be equal to, and sometimes even exceed, the cost of the IT components' subsystem. The cooling subsystem of a data center can be implemented in several ways. In this work, our focus is on the implementation of this subsystem considering Tiers I and II from the TIA-942 standard. There are different technologies for cooling subsystems [3], however, in this work, we are considering a chilled water system, which is composed of five main components: CRACs, chiller, cooling tower, pipes, and pumps. Figure 2(a) depicts a Tier I cooling subsystem. As one can note, in Tier I, there is only one path to each component and there is no component replication, i.e., the failure of any component turns the subsystem unavailable.

Pipes are responsible for linking the main components, and pumps are responsible for pushing the chilled water from chiller to CRAC, and from the chiller to the cooling tower. A CRAC draws heat from the environment, heating the chilled water that is conducted to the chiller. The chiller has an evaporator, that cools the water received from the CRAC via the Chilled Water Return (CWR) pipe and sends the chilled water to CRAC through a Chilled Water Supply (CWS) pipe. The cooling process generates heat in the condenser, that sends warm water to the cooling tower through the Condenser Water Supply (CDWS) pipe and returns it via the Condenser Water Return (CDWR) pipe. The cooling tower rejects heat to the outside environment. On the other hand, in Tier II (Figure 2(b)) each component is replicated with $N + 1$ redundancy, except pipes. Despite redundancy, there is still a single path from the cooling tower to the chiller and from the chiller to the CRAC.



(a) (b)

Fig. 2. Cooling architecture of Tier I (a) and Tier II (b) (adapted from [8]).

## III. RELATED WORK

There are few works that model the impact of cooling failures on data center availability. Liu et al [11] present models for sustainable data centers with renewable energy. An analytical model is developed to study the impact on costs and availability when using outside air and/or solar energy.

Alissa et al. [12] describe a study regarding cooling failures on data centers. The impact of a failure is measured by a metric called available uptime (AU), defined as the time until the IT components turn themselves off due to high temperature after a failure of the cooling subsystem.

Callou et al. [5] [6] propose models of a cooling subsystem using SPN. The cooling components are replicable by submodels, which may lead to scalability problems. The authors model architectures compatible with Tier I and Tier II. However, they do not take into account data center standardization.

Our work proposes, as [5], models for cooling subsystem in Petri Nets. However, our models are more scalable through addition of tokens, and follow the TIA-942 standard; we also consider CRAC rotation.

## IV. COOLING SUBSYSTEM TIER I MODELS

In this work, we are proposing SPN models that represent Tier I and Tier II cooling subsystems. Our goals are: $i)$ propose a scalable and validated Tier I model, according to [5], and $ii)$ propose a scalable Tier II model with different CRAC operation types.

### *A. Basic Cooling Subsystem Tier I*

Figure 3 shows our Tier I model based on [5]. Note, for now, we are not considering pipes and pumps, and the cooling subsystem works without redundancy, composed of five CRACs (we are assuming that the data center needs at minimum five CRACs to work properly), one chiller, and one cooling tower. Each component is modelled as a building block composed of two places ($ON$ and $FAIL$) and two transitions (one from active to failed place ($F$) and another from failed to active place ($R$)). One important feature of our proposal is that each transition $CRAC\_F$ and $CRAC\_R$ is configured as infinite server; this allow to model the independence of redundant components and their simultaneous failures, and also to provide a scalable model.

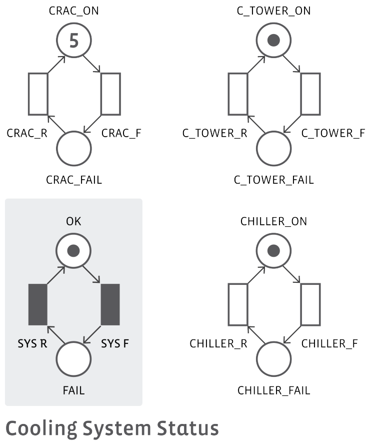


Fig. 3. Model for tier I architecture without pipes and pumps

There is a building block that represents the cooling subsystem status (with two places named $OK$ and $FAIL$ connected with two immediate transitions). Then, as the availability of the subsystem is calculated as the probability of all components being working, our model calculates the probability that place $OK$ has one token ($P(\#OK = 1)$).

The immediate transitions are configured with guard functions to guarantee the correct behavior: $SYS\_F$ is enabled when any component fails, and $SYS\_R$ is enabled when all components are running. Table I shows the guard functions, all with priority equal to one.

TABLE I
GUARD FUNCTIONS OF BASIC COOLING SUBSYTEM - TIER I

| **Transition** | **Guard Function** |
|---|---|
| $SYS\_F$ | $(\#CRAC\_ON < 5)$ OR $(\#CHILLER\_ON = 0)$ OR $(\#C\_TOWER\_ON = 0)$ |
| $SYS\_R$ | $(\#CRAC\_ON >= 5)$ AND $(\#CHILLER\_ON >= 1)$ AND $(\#C\_TOWER\_ON >= 1)$ |

### *B. Cooling Subsystem Tier I with Pumps and Pipes*

In order to improve the cooling subsystem model, we propose a model considering pumps and pipes (Figure 4), based on the architecture shown in Figure 2(a), but with five CRACs. Pipes and pumps follow the same characteristics as other components described in basic cooling subsystem (subsection IV-A). The main difference is that there are two pipes from the chiller to the cooling tower ($PIPE\_CDWS$ and $PIPE\_CDWR$, indicating condenser water pipes), and two pipes from the chiller to the CRACs ($PIPE\_CWS$ and $PIPE\_CWR$, the chilled water pipes).

There is one pump between the chiller and the CRAC ($PUMP\_TO\_CRAC$) and another between the chiller and the cooling tower ($PUMP\_TO\_C\_TOWER$). If a pump fails, the pipes referring that flow are disabled and the subsystem fails. For instance, if $PUMP\_TO\_CRAC$ fails, the immediate transition $C\_D$ (that means chiller pipes disabled) is triggered, disabling the chilled water supply and chilled water return pipes. Immediate transition $C\_E$ (that means chiller pipes enabled) is fired when the place $PUMP\_TO\_CRAC\_ON$ has one or more tokens, indicating a repair of this pump. Both $C\_D$ and $C\_E$ have guard functions as shown in Table III. The same behavior occurs in condenser water pipes when the $PUMP\_TO\_C\_TOWER$ fails ($CD\_D$) and is repaired ($CD\_E$).

TABLE II
GUARD FUNCTIONS OF COOLING SUBSYSTEM WITH PUMPS AND PIPES - TIER I

| **Transition** | **Guard Function** | **Priority** |
|---|---|---|
| $SYS\_F$ | $(\#CRAC\_ON < 5)$<br>OR $(\#CHILLER\_ON < 1)$ OR $(\#C\_TOWER\_ON < 1)$<br>OR $(\#PIPE\_CDWR\_ON < 1)$ OR $(\#PIPE\_CDWS\_ON < 1)$<br>OR $(\#PIPE\_CWR\_ON < 1)$ OR $(\#PIPE\_CWS\_ON < 1)$ | 1 |
| $SYS\_R$ | $(\#CRAC\_ON >= 5)$<br>AND $(\#CHILLER\_ON >= 1)$ AND $(\#C\_TOWER\_ON >= 1)$<br>AND $(\#PIPE\_CDWR\_ON >= 1)$ AND $(\#PIPE\_CDWS\_ON >= 1)$<br>AND $(\#PIPE\_CWR\_ON >= 1)$ AND $(\#PIPE\_CWS\_ON >= 1)$ | 1 |
| $CD\_D$ | $\#PUMP\_TO\_C\_TOWER\_ON = 0$ | 2 |
| $CD\_E$ | $\#PUMP\_TO\_C\_TOWER\_ON > 0$ | 2 |
| $C\_D$ | $\#PUMP\_TO\_CRAC\_ON = 0$ | 2 |
| $C\_E$ | $\#PUMP\_TO\_CRAC\_ON > 0$ | 2 |

Subsystem availability is calculated as the probability of having one token on active state ($ON$) in each component. Note that, even though the pump is not present in the formula, when it fails automatically this event disables the pipes, leading to system unavailability.

The four immediate transitions ($CD\_D$, $CD\_E$, $C\_D$ and $C\_E$) have a higher priority than system status immediate transitions ($SYS\_F$ and $SYS\_R$) to avoid the double firing. We modeled four pipes and two pumps as building blocks to better represent the behavior of the subsystem, identifying which pipes are affected by a failure of a specific pump.

## V. COOLING SUBSYSTEM TIER II MODELS

Regarding Tier II, we propose two models: with and without CRACs rotation; both with pipes and pumps. All components are replicated with $N + 1$ redundancy except pipes, because according to Tier II definitions, there is only a single path (see Figure 2(b)).

### A. Basic Cooling Subsystem Tier II

Our cooling subsystem of Tier II has all building blocks of Tier I with pumps and pipes (Figure 4), with exception of CRAC, that has a cold-standby redundancy now (Figure 5). There are five active CRACs and one standby. The standby CRAC is represented by one token in place $CRAC\_OFF$.

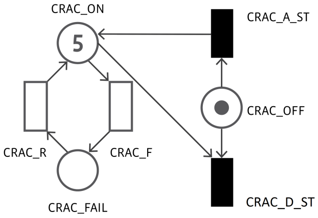


Fig. 5. Basic model for Tier II with pipes and pumps

This model uses the same guard function as our Tier I model with pumps and pipes (Table II), and additionally two others: $CRAC\_A\_ST$ and $CRAC\_D\_ST$ (Table III).

When a failure occurs, the number of tokens in place $CRAC\_ON$ becomes fewer than five, which fires the transition $CRAC\_A\_ST$, activating a standby CRAC to always maintain the five CRACs running. When the CRAC is repaired, the token returns to $CRAC\_ON$. If there are six tokens in

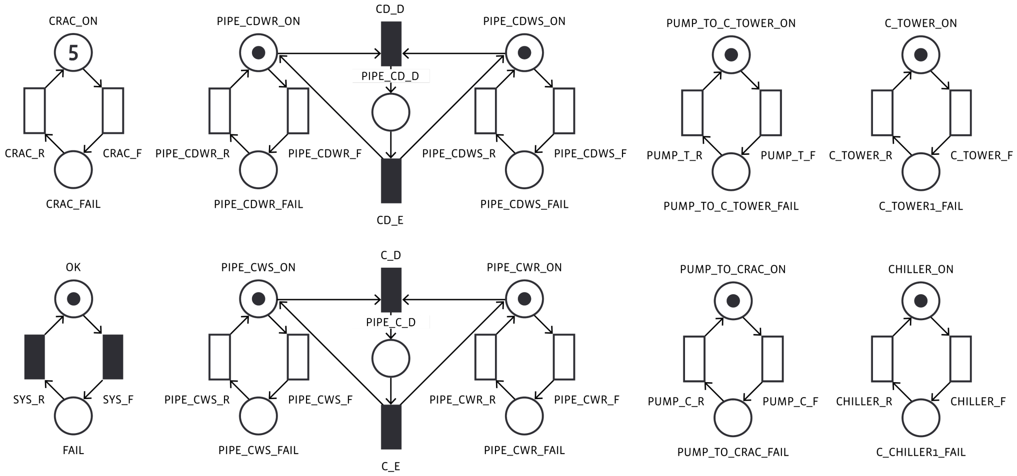


Fig. 4. Model for tier I architecture with pipes and pumps.

$CRAC\_ON$, $CRAC\_D\_ST$ is enabled, deactivating one of the six CRACs and putting one of them in standby mode.

The timed transitions referring to CRAC are infinite server type to simulate a failure of more than one CRAC, while other transitions are single server, behaving as a standby component.

TABLE III
GUARD FUNCTIONS OF BASIC COOLING SUBSYSTEM - TIER II

| Transition | Guard Function | Priority |
|---|---|---|
| $CRAC\_D\_ST$ | $\#CRAC\_ON > 5$ | 1 |
| $CRAC\_A\_ST$ | $\#CRAC\_ON < 5$ | 2 |

### B. Cooling Subsystem Tier II with CRAC Rotation

The other cooling components are modeled in the same way as a basic cooling subsystem, as depicted in Figure 5. Figure 6 shows the CRAC building block with the rotation behavior.

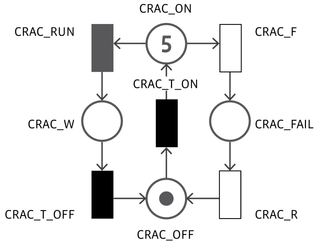


Fig. 6. CRAC building block representing the rotation of components

This model uses the guard functions of our Tier I model with pump and pipes (Table II), and additionally three others: $CRAC\_T\_ON$, $CRAC\_T\_OFF$ and $CRAC\_RUN$ (Table IV).

TABLE IV
GUARD FUNCTIONS ON COOLING SUBSYSTEM II WITH CRAC ROTATION

| Transition | Guard Function | Priority |
|---|---|---|
| $CRAC\_T\_ON$ | $\#CRAC\_ON < 5$ | 2 |
| $CRAC\_T\_OFF$ | $\#CRAC\_ON >= 5$ | 1 |
| $CRAC\_RUN$ | $\#CRAC\_OFF >= 1$ | 1 |

Now, CRACs run during a period of time, following a deterministic distribution, and one of them is replaced by the standby, firing the $CRAC\_RUN$ transition. The CRAC to be replaced waits for the standby CRAC, and it is represented as a token in place $CRAC\_W$. As the number of CRACs becomes fewer than five, the immediate transition $CRAC\_T\_ON$ fires and adds one token to the $CRAC\_ON$ place, keeping always five CRACs running. The transition $CRAC\_T\_OFF$ turns off the CRAC that was waiting for the replacement, adding a token in place $CRAC\_OFF$. This cycle continues until a CRAC fails.

The subsystem availability follows the same as Tier I with pipes and pumps. This model assumes that the time to turn on a CRAC is negligible, and this action is represented by an immediate transition $CRAC\_T\_ON$.

## VI. RESULTS

We validate the infinite server proposal of Tier I based on [5] and evaluate our proposals with and without CRAC rotation on Tier II. Each CRAC may have the following specifications: Stulz ASD 1200 CW2, 64% of airflow volume, 16.2 Ton of Nominal Sensible Cooling, with a CRAC airflow rate of 16,632 Cooling Meter per Hour (CMH), leaving temperature of 18 C and return temperature of 24 C[2]. These specifications allow five CRACs maintain a data center with until 300 kW, enough to sustain a data center of $465m^2$ with 150 racks [3].

As our model with rotation has a deterministic transition, it is not possible make a steady-state analysis, which leads us to evaluate it by simulation using the Mercury tool [13]. Parameters are shown in Table V. The $CRAC\_RUN$ value in the model with rotation is 12h, based on [14]. The simulation was carried out with a 95% confidence interval, a relative error of 10%, a batch size of 50000 and each transition have 1000 firings.

TABLE V
MTTF AND MTTR OF COOLING EQUIPMENT

| Equipment | MTTF (h) | MTTR (h) |
|---|---|---|
| CRAC [6] | 37,509 | 8 |
| Chiller [6] | 18,000 | 48 |
| Cooling Tower [6] | 24,816 | 48 |
| Pump [15] | 1,387,387 | 7.09 |
| Pipe [15] | 479,265 | 2.71 |

### A. Model Validation and Evaluation

We compare the simulation results of our Tier I model with infinite server and the model proposed by [5]. Table VI shows the experiment results, with the confidence intervals generated by the Mercury tool.

TABLE VI
VALIDATION RESULTS OF TIER I ARCHITECTURE

| Tier I | Availability | 9's |
|---|---|---|
| Basic Cooling Subsystem from [5] | 99.426 - 99.450% | 2.249 |
| Basic Cooling Subsystem | 99.440 - 99.444% | 2.253 |
| Cooling Subsystem with Pumps and Pipes | 99.434 - 99.438% | 2.249 |

Results show that our basic model with infinite server is within the availability range of the model proposed by [5], varying from 99.440% to 99.444%; that means a average downtime of 48.8808 hours/year. The model with pumps and pipes achieves a smaller availability, as expected, due to these two new components, ranging from 99.434% to 99.438%, representing a average downtime of 49.4064 hours/year.

---

[2] How to Build a Data Centre Cooling Budget - https://www.bicsi.org/pdf/presentations/euro_11/How_to_Build_a_Data_Centre_Cooling_Budget.pdf. Last Accessed: Apr 10, 2017.

[3] Calculating Total Cooling Requirements for Data Centers - http://www.apc.com/salestools/NRAN-5TE6HE/NRAN-5TE6HE_R3_EN.pdf. Last Access: Apr 10, 2017.

Table VII shows the availability with and without CRAC rotation. Results of these two approaches are similar with slightly higher availability for rotation; with the basic Tier II presenting a mean downtime of 0.22338 hours/year (99.9974 - 99.9975%), and the Tier II with CRAC rotation 0.20586 hours/year (99.9976 - 99.9977%).

TABLE VII
TIER II EVALUATION

| **Tier II** | **Availability** | **9's** |
|---|---|---|
| Basic Cooling Subsystem | 99.9974 - 99.9975% | 4.593 |
| Cooling Subsystem with CRAC Rotation | 99.9976 - 99.9977% | 4.636 |

### *B. Discussion*

Regarding the Tier I model, the major difference of our proposal in comparison with [5] resides in the subsystem status modeling. The system status facilitates the connection with other data center subsystems (such as power and IT), specifying the state of the cooling subsystem to calculate the total data center availability.

Another difference is that we modeled CRAC with only one building block that contains five tokens to represent each CRAC. Callou, et al. [5] created one building block for each CRAC. In this way, the model proposed by [5] tends to be less scalable due to the number of places and transitions in higher tier models. In our proposal, the scalability is achieved by maintaining the same model structure for all tiers, just adding the tokens. The transitions in the model are modeled as infinite server type, allowing the independence of components. The firing of a timed transition may cause a failure of one or more components simultaneously [16]. Therefore, our proposed model has the same behavior with fewer states in Petri Net.

In comparison with [5], we generated the corresponding Markov Chain from our Petri Net of Basic Tier I model, and it presented only 24 states, while the [5] Markov Chain had 128 (a reduction of 104 states). Our Basic Tier II Markov Chain model was composed of 108 states, and the [5] had 1024 (a reduction of 1024 states).

The Tier II model increases the availability to four nines and the CRAC rotation does not present a considerable impact on availability. This result can be explained because all CRACs have the same MTTF value. In a real scenario, it is assumed that a component has a greater probability of failure if it stays off for a long time. This assumption is not inserted into our model, since we use exponential transitions and the time off does not influence the CRAC MTTF value. In this way, it is necessary to analyze datasheets and technical reports to verify if the assumption is correct.

## VII. CONCLUSIONS AND FUTURE WORKS

Data centers rely on cooling subsystem to keep their services running without unplanned failures. Some data centers implement a cooling subsystem with a rotation of CRACs in an attempt to increase the availability and decrease failures.

We presented scalable models based on Petri Nets regarding Tier I and II. Our Tier I model was validated with other model from literature. In relation to Tier II, our models presented similar results in comparison with CRAC operation. Our models are easily scalable by adding tokens, allowing a straightforward modeling of Tier III and IV and the connection with other data center subsystems, such as power and IT.

As future work, we plan to model tiers III and IV, insert the available uptime metric in our model, and perform sensitivity analysis to understand which components are more crucial regarding data center availability.

## ACKNOWLEDGMENT

This work was supported by the RLAM Innovation Center, Ericsson Telecomunicações S.A., Brazil. We would like to thank the design team of Networking and Telecommunication Research Group (GPRT) for the support.

## REFERENCES

[1] M. Gagnaire, F. Diaz, C. Coti, C. Cerin, K. Shiozaki, Y. Xu, P. Delort, J.-P. Smets, J. Le Lous, S. Lubiarz, *et al.*, "Downtime statistics of current cloud solutions," *International Working Group on Cloud Computing Resiliency, Tech. Rep*, 2014.

[2] W. J. Kosik and H. Geng, "Data center benchmark metrics," *Data Center Handbook*, pp. 593–599, 2015.

[3] T. Evans, "The different technologies for cooling data centers." http://www.apc.com/salestools/VAVR-5UDTU5/VAVR-5UDTU5_R2_EN.pdf, 2012. Last access: February, 2017.

[4] A. Krone, "Tia-942: Data center standards overview," 2008.

[5] G. Callou, P. Maciel, D. Tutsch, and J. Araújo, "Models for dependability and sustainability analysis of data center cooling architectures," in *Dependable Systems and Networks Workshops (DSN-W), 2012 IEEE/IFIP 42nd International Conference on*, pp. 1–6, IEEE, 2012.

[6] G. Callou, P. Maciel, D. Tutsch, J. Ferreira, J. Araújo, and R. Souza, "Estimating sustainability impact of high dependable data centers: a comparative study between brazilian and us energy mixes," *Computing*, vol. 95, no. 12, pp. 1137–1170, 2013.

[7] F. Bause and P. S. Kritzinger, *Stochastic Petri Nets*, vol. 1. Vieweg Wiesbaden, 2002.

[8] "Ansi/bicsi 002, data center design and implementation best practices." https://www.bicsi.org/uploadedFiles/BICSI_Website/Global_Community/Presentations/CALA/Ciordia_002_Colombia_2016.pdf. Last access: November, 2016.

[9] D. A. Patterson, "The data center is the computer," *Communications of the ACM*, vol. 51, no. 1, pp. 105–105, 2008.

[10] H. Zhang, S. Shao, H. Xu, H. Zou, and C. Tian, "Free cooling of data centers: A review," *Renewable and Sustainable Energy Reviews*, vol. 35, pp. 171–182, 2014.

[11] Z. Liu, Y. Chen, C. Bash, A. Wierman, D. Gmach, Z. Wang, M. Marwah, and C. Hyser, "Renewable and cooling aware workload management for sustainable data centers," in *ACM SIGMETRICS Performance Evaluation Review*, vol. 40, pp. 175–186, ACM, 2012.

[12] H. A. Alissa, K. Nemati, B. G. Sammakia, M. J. Seymour, R. Tipton, D. Mendo, D. W. Demetriou, and K. Schneebeli, "Chip to chiller experimental cooling failure analysis of data centers: The interaction between it and facility," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 6, no. 9, pp. 1361–1378, 2016.

[13] B. Silva, R. Matos, G. Callou, J. Figueiredo, D. Oliveira, J. Ferreira, J. Dantas, A. Lobo, V. Alves, and P. Maciel, "Mercury: An integrated environment for performance and dependability evaluation of general systems," in *Proceedings of Industrial Track at 45th Dependable Systems and Networks Conference, DSN*, 2015.

[14] P. Endo, G. Goncalves, D. Gomes, D. Rosendo, G. Leoni, D. Sadok, and M. Mahloo, "Technical report: Data center specifications.," 2017.

[15] C. Heising *et al.*, "Ieee recommended practice for the design of reliable industrial and commercial power systems," *IEEE Inc., New York*, 2007.

[16] R. David and H. Alla, *Discrete, continuous, and hybrid Petri nets*. Springer Science & Business Media, 2010.